

REVIEW ARTICLE

OPEN



Statistics

Systematic review of eyecare Patient-Reported Experience Measures (PREMs) following COSMIN guidelines

Pongnugoon Kongjaidee^{1,2}, Timothy Pickles³, Barbara Ryan¹, Francis W. B. Sanders⁴, Colm McAlinden^{1,5} and Jennifer H. Acton¹✉

© The Author(s) 2026

Patient-Reported Experience Measures (PREMs) are essential for assessing healthcare quality from the patient's perspective. However, their use in eyecare is limited. This systematic review aimed to assess the existing evidence on the measurement properties of PREMs used to capture patient experiences of eyecare services. A comprehensive literature search was conducted across Web of Science Core Collection, Embase, PubMed, and Scopus to identify studies reporting the development and/or psychometric evaluation of PREMs in eyecare. The extracted measurement properties encompassed information on item development, content validity (the extent to which items are relevant, comprehensive, and understandable), structural validity (whether the items form the right scales), internal consistency (how well items measure underlying eyecare experience domains), cross-cultural validity, reliability, hypothesis testing, and responsiveness. Quality assessment, evaluation of measurement properties, evidence synthesis, and grading followed COSMIN methodology for systematic reviews. Thirteen studies were included, three of which evaluated 10 distinct PREMs, across settings, including low vision and primary/secondary eyecare. All ten PREMs met criteria for COSMIN Category B (potentially recommended, further validation needed). However, most lacked sufficient evidence for content validity and lacked supporting evidence for structural validity. None demonstrated adequate reliability, and none reported using Item Response Theory or Rasch Measurement Theory. Consequently, there was insufficient evidence to classify any of the identified PREMs into COSMIN recommendation Category A (recommended for use). Future studies on PREMs in eyecare should aim to strengthen the evidence for content and structural validity and provide robust proof of their reliability.

Eye; <https://doi.org/10.1038/s41433-026-04907-1>

INTRODUCTION

In recent years, the concept of the patient experience has become a central focus in evaluating healthcare system performance and driving quality improvement. It is now recognised as a key component in defining high-quality care, which aims to deliver safe, effective, and patient-responsive services; value-based care, which seeks to maximise health outcomes relative to resources used; and patient-centred care [1]. Patient-centred care prioritises the individual healthcare needs of each person, with the primary objective of empowering patients to actively engage in managing and participating in their own care [2]. Understanding the patient's perspective can be achieved through patient-reported indicators, which serve as valuable tools for assessing and enhancing patient-centred care.

Patient-Reported Experience Measures (PREMs) are instruments designed to capture patients' evaluations of their personal experiences with healthcare services. PREMs, typically administered through structured questionnaires, enable patients to

convey the extent to which predetermined facets of care were encountered during healthcare interactions. These instruments have been widely developed and used in the evaluation of care quality from the patient perspective across diverse healthcare settings, although the extent of psychometric validation varies between instruments. They are particularly valuable in palliative care, where they illuminate domains of care that hold critical significance for patients with serious or life-limiting illnesses [1, 3]. Both generic and condition-specific PREMs are in use across a range of clinical settings, including hospitals and outpatient clinics [4]. Generic PREMs are designed to be broadly applicable across patient groups, irrespective of diagnosis or treatment. Condition-specific PREMs are tailored to capture experiences relevant to particular clinical contexts or disease groups, for example, questionnaires developed to assess patient and carer experiences in experimental cancer medicine trials (PREM-ECM) [5] or instruments focused on cardiovascular disease care [6], thereby providing more precise insights into the care experiences

¹Centre for Vision Services Research, College of Biomedical and Life Sciences, Cardiff University, Cardiff, UK. ²Interdisciplinary Health and Data Sciences Research Unit, Department of Optometry, Faculty of Allied Health Sciences, Naresuan University, Phitsanulok 65000, Thailand. ³Psychology, School of Humanities and Social Sciences, Leeds Beckett University, Leeds, UK. ⁴Health and Care Research Wales, Cardiff, UK. ⁵Department of Ophthalmology, Princess of Wales Hospital, Wales, UK. ✉email: actonj@cardiff.ac.uk

Received: 19 March 2026 Revised: 19 August 2026 Accepted: 11 September 2026

Published online: 05 October 2026

of specific patient populations. Given their increasing use in evaluating patient-centred care, informing quality improvement initiatives, and supporting healthcare decision-making, the availability of PREMs with strong evidence of validity and reliability is essential to ensure that patients' experiences are measured accurately and meaningfully.

PREMs capture patients' perceptions of the care they receive, including aspects such as communication, coordination, and responsiveness of services, whereas patient-reported outcome measures (PROMs) focus on patients' self-assessed health status, symptoms, and health-related quality of life throughout the care pathway, including pre-, peri-, and post-treatment phases [7]. Although PROMs provide important information regarding patients' health outcomes, they are not designed to capture patients' experiences of healthcare delivery. PREMs complement these measures by capturing patients' experiences of care processes, including communication, shared decision-making, and respect for patient preferences, thereby offering a more comprehensive assessment of patient-centred care and informing quality improvement initiatives. In eyecare, patient experience has been conceptualised using the five-domain framework developed by Somner et al. [8] comprising administration, environment, patient-centred care, knowledge and understanding, and communication. This framework provides a useful basis for evaluating the content and comprehensiveness of PREMs in eyecare services.

The Consensus-based Standards for the Selection of Health Measurement INstruments (COSMIN) guidelines, updated in 2024 [9], provide a comprehensive framework for evaluating the measurement properties of patient-reported instruments, including content validity, structural validity, internal consistency, cross-cultural validity, reliability, measurement error, criterion validity, construct validity, and responsiveness. These measurement properties provide evidence of whether an instrument adequately captures the intended construct, produces reliable and reproducible measurements, demonstrates measurement equivalence across populations and cultural settings, and is responsive to meaningful changes over time. Establishing these properties is essential to ensure that PREMs yield valid, reliable, and interpretable data that can be confidently used to evaluate healthcare quality, guide service improvement, and support clinical and health services research. Although these standards were originally developed for PROMs, the underlying measurement principles are equally applicable to PREMs, as both are self-reported instruments that require rigorous evaluation. Consequently, COSMIN has increasingly been adopted to evaluate PREMs in recent research [10, 11]. Despite the increasing use of PREM data as performance indicators, research in this area remains limited [1]. PREMs specifically developed for the eyecare context are relatively scarce.

The aim of this systematic review was to evaluate the measurement properties of PREMs used to assess the experiences of individuals receiving eyecare services. We also aimed to provide evidence-based recommendations to guide the selection of the most appropriate PREMs for use in both research and clinical practice.

METHODS

Search strategy

A comprehensive literature search using the following databases: Web of Science Core Collection, Embase, PubMed and Scopus and databases were searched from inception to 1st March 2026. The search filter followed the COSMIN guidelines and employed a combination of subject headings and keywords [12]. In addition to database searching, citation tracking was undertaken by examining the reference lists of key articles identified in the initial search. The strategy incorporated four main conceptual

components aligned with the review objectives: 1. the construct of interest, 2. the target population, 3. the type of instrument, and 4. the measurement properties. The complete search strategies for all databases included in this review (see Supplementary File 1). All searches were conducted by a single reviewer (P.K.).

Literature screening

Following execution of the search strategy, titles, abstracts, and full-text articles were independently screened by the author (P.K.) and the supervisor (J.A.) using Rayyan software, (Rayyan Systems, Inc.; web application, Cambridge, MA, www.rayyan.ai, accessed 1 November 2024) [13]. Any disagreements between the two reviewers were resolved through consultation with the co-supervisor (B.R.). Screening was conducted in accordance with predefined inclusion and exclusion criteria, as outlined below:

Inclusion criteria:

1. Any articles relevant to the use of PREMs, patient reported experience, patient reported experience measure, patient reported outcome and experience, patient experience, patient opinion, patient preference, patient thoughts, patient satisfaction, patient centred, patient perspective, patient responsiveness, patient views, patient perception, patient assessment, patient feedback, and with details that can be reviewed against the COSMIN guidelines.
2. Questionnaires which incorporated both PROM and PREM questions.
3. Questionnaires defined by the authors as POEMs (Patient Outcome and Experience Measures).
4. Any articles relevant to participants who had experience of eyecare services.
5. The instrument should consist of a questionnaire designed to be completed either through self-reporting or interviews by individuals who had experience about eyecare services or by their caregivers.

Exclusion criteria:

1. Questionnaires defined as PROMs, designed to evaluate a patient's health status, including symptoms like pain, fatigue, and mobility or questionnaires are utilised to assess a patient's health condition at various stages, including before, during, and after treatment, providing valuable insights into the treatment's impact on their well-being.
2. Questionnaires involving questions which do not relate to the patient experience.
3. Studies published in languages other than English were excluded due to limited resources, including the unavailability of reliable translation services, which could have hindered the accurate evaluation of their methodological quality and findings. However, studies describing PREMs originally developed in other languages were included if the results and methodological details were reported in English.

Extract the PREM data

The PREMs details of these included studies were extracted following [11], specific aspects related to various PREM versions were also captured, such as the theoretical framework employed, the number of items, domain descriptions, scoring methods and ranges, delivery modes, response rates, completion times, timing of assessment, and recall periods. Any missing or ambiguous data were resolved through consensus between two reviewers (P.K. and J.A.), with consultation from an optometry expert (B.R.) when necessary. When complete data were not available, this was reported.

Evaluate the measurement properties of the PREMs

The quality of PREMs was independently assessed by P.K. and J.A. using a systematic review tool designed in accordance with the COSMIN guidelines [9]. Although the COSMIN framework was originally developed for PROMs, it evaluates fundamental

psychometric properties, that are applicable to all patient-reported measurement instruments. As PREMs are also patient-reported instruments requiring rigorous psychometric evaluation, the COSMIN framework provides an appropriate and comprehensive approach for assessing their methodological quality. Any discrepancies between the two researchers were resolved through consultation with T.P., ensuring consensus and maintaining the reliability and precision of the evaluation process.

Data extraction: study population characteristics. Data were extracted on key characteristics of the study population, including the first author, details of the PREMs used, country of study, instrument language, study setting, research design, patient demographics, type of eyecare service, and sample size. To ensure accuracy and reduce the likelihood of missing relevant information, data extraction was performed by P.K. and subsequently verified by J.A.

Methodological quality of the study assessment. The methodological quality of the included studies was assessed using the COSMIN Risk of Bias Checklist (version 3.0) [9]. This comprehensive tool contains 10 assessment domains and 113 items, covering nine key measurement properties: PREM development; content validity (including relevance, comprehensiveness, and comprehensibility); structural validity; internal consistency; cross-cultural validity; reliability; measurement error; criterion validity; hypothesis testing; and responsiveness. Each aspect was rated on a four-point scale—Very good, Adequate, Doubtful, or Inadequate—to determine the methodological quality. The “worst score counts” approach was applied, whereby the lowest rating in any category determined the overall score for each measurement property.

Criteria for good measurement properties. To determine the adequacy of each result, the study findings for each measurement property were evaluated against the COSMIN criteria for good measurement properties, which classify outcomes as indeterminate (?) rating, insufficient (–) rating, or sufficient (+) rating. This

review followed the COSMIN-recommended table of criteria for good measurement properties (Supplementary File 2). Criterion validity was not assessed, as no established gold standard currently exists for measuring patient-reported experiences in eyecare.

Quality of evidence (certainty assessment). The quality of evidence for the measurement properties of PREMs was assessed using the modified Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach, as outlined in the COSMIN guideline version 2.0 (2024). Evidence was categorised into four levels: **high, moderate, low, or very low**. Downgrading by one or two levels was applied when limitations such as risk of bias, unexplained inconsistencies, imprecision due to small sample size, or indirectness were identified. In cases where the evidence was derived from a single study of inadequate quality, indicating an extremely high risk of bias, the rating could be reduced by up to three levels.

Recommendations

Recommendations were formulated based on the evaluation of measurement properties and the strength of evidence for each included PREM. Instruments were classified as **Category A**—indicating a strong recommendation for use—if they demonstrated sufficient (+) content validity (irrespective of evidence quality) and at least low-quality evidence for sufficient (+) internal consistency. **Category C** was assigned to PREMs for which high-quality evidence indicated that at least one measurement property was insufficient (–), signifying that their use is not recommended. PREMs that did not meet the criteria for either Category A or Category C were assigned to **Category B**, reflecting potential applicability but the need for further research to confirm their suitability.

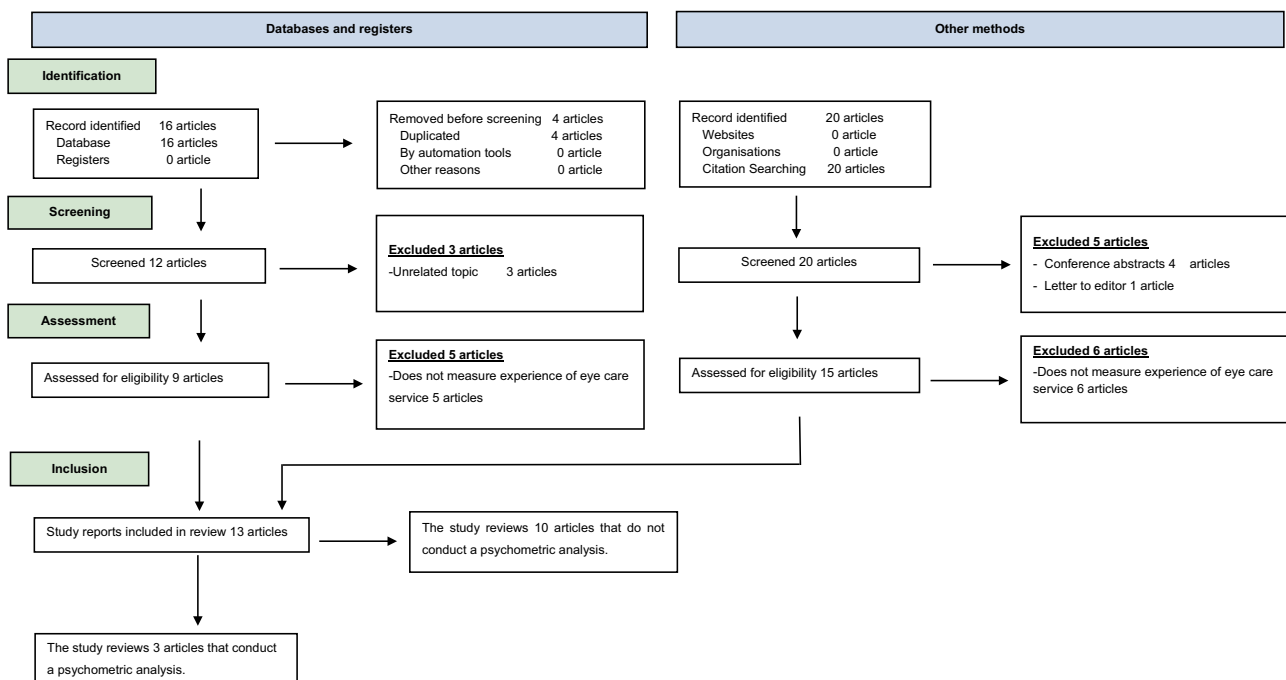


Fig. 1 PRISMA-COSMIN flow diagram illustrating the study selection process for the systematic review of measurement properties, adapted from [38]. The diagram details the identification, screening, eligibility assessment, and final inclusion of studies evaluating PREMs.

RESULTS

Article identification

Figure 1 summarises the literature search process. Searches across four databases identified 16 studies, with a further 20 retrieved via citation tracking. After removing duplicates and screening titles, abstracts, and full texts, 13 studies met the inclusion criteria. Of these, three conducted analyses of measurement properties: PREMs were developed and psychometrically evaluated for low vision services [14]; PREMs were designed and evaluated for inherited retinal disorder services [15]; and PREMs were evaluated for the reliability and dimensional structure of a cataract surgery patient experience questionnaire across multiple hospitals [16].

The remaining ten studies, covering retinal disease services [17, 18], glaucoma services [8, 19–22], paediatric eyecare [23, 24], and general eyecare [25], did not include evaluation of measurement properties and were therefore excluded from the detailed analysis.

Study characteristics

Country or location. Table 1 summarises the main characteristics of the included studies. Research was conducted across varied geographical settings, with most studies from the United Kingdom [8, 14, 15, 18, 22], followed by Australia [17, 20] and the United States [19, 24]. Single studies were undertaken in Sweden [23], India [21], the Netherlands [16], and Nigeria [25]. Instrument languages included Dutch [16], and Swedish [23], with the remaining ten instruments available only in English. One study employed semi-structured interviews but did not specify the language used [25].

Study designs. Ten PREMs were identified across the included studies. Most studies employed a cross-sectional design [14, 16–18, 21, 25]. Three studies used prospective observational designs [15, 20, 22], and another used a retrospective review [24]. A qualitative approach was adopted in one study [8], while two did not report their design [19, 23]. Across all studies, 9465 responses were reported, with the number of responses per study ranging from 11 to 4,635, ages from 7.6 to 88 years, and male representation ranging from 19.9% to 84%.

Scope of PREM questions in three studies that included analysis of measurement properties. All studies were evaluated with respect to the inclusion of key domains, adhering to the framework developed by Somner et al. [8], who distilled five core domains—administration, environment, patient-centred care, knowledge and understanding, and communication—from patient experience. Table 2 presents a summary of the scope of the PREM questions across three studies [14–16]. Among the five domains, communication emerged as the most frequently addressed area, featuring in seven of the ten PREMs: LVSW-specific PREMs, General NHS Wales PREMs, Communication with Ophthalmologist, Communication with Nurses, Communication about Medication, Communication and Information, and Motivation. This was followed by patient-centred care, which was explored in six PREMs: LVSW-specific PREMs, General NHS Wales PREMs, Communication with Ophthalmologist, Communication with Nurses, Communication and Information, and Patient Care. The domain of administration was examined in five PREMs: LVSW-specific PREMs, General NHS Wales PREMs, Communication with Ophthalmologist, Patient Care, and Research Visit versus Regular Clinic Visit. In contrast, the domains of environment and knowledge and understanding were each represented in only one PREM (LVSW-specific PREMs). Notably, the LVSW-specific PREMs provided the most comprehensive coverage, addressing all five key domains. Overall, most PREMs demonstrated substantial coverage of three principal domains: communication, patient-centred care, and administration.

PREM characteristics

As shown in Table 3, only three studies [14–16] reported measurement properties, covering ten PREMs used in settings such as low vision services, inherited retinal disorder care, and cataract services. These PREMs addressed domains including LVSW-specific experiences, NHS Wales general PREMs, communication with clinicians, medication information, efficiency, patient care, motivation, and comparisons between research visits and regular clinic visits. Scoring was based on item-level responses using 4- or 5-point Likert scales.

The remaining ten studies [8, 17–25] did not evaluate measurement properties. These questionnaires were analysed using item-level scores and percentage responses. Reported questionnaire response rates ranged from 39% to 96%, with recall periods from one day to five months.

Risk of bias rating, content validity rating and quality of measurement property rating

Only six PREMs described content validity, both under the heading of Cognitive interview study or other pilot test for the COSMIN comprehensibility domain. (Table 4, Content validity – PREM Development columns). These were LVSW-specific PREMs, Communication and information, Efficiency, Patient care, Motivation, and Research visit versus regular clinic visit. For LVSW-specific PREMs, little detail about the process was available other than the number of participants interviewed, which was $n = 14$ and there was no mention of whether the process involved collaboration with professionals, so “doubtful” risk of bias rating was given under Cognitive interview study or other pilot test Comprehensibility study. For the content validity comprehensibility, people with low vision were asked about the comprehensibility of item instructions, so the overall content validity rating was “+”. The rating of reviewers was “+” because it was clear that all items and response options were appropriately worded and response options matched the items.

For five PREMs (Communication and information, Efficiency, Patient care, Motivation, and Research visit versus regular clinic visit), little detail about the process was available other than the number of participants interviewed, which was $n = 5$ people, so a “doubtful” risk of bias rating was given under Cognitive interview study or other pilot test Comprehensibility study. There was no detail about people with inherited retinal disorders being asked about the comprehensibility of item instructions and of the items and their response options, so the overall content validity rating was “?”. The rating of reviewers was “+” because it was clear that all items and response options were appropriately worded and response options matched the items.

The columns to the right of content validity in Table 4 present summaries of the quality of measurement properties along with the associated risk of bias ratings. All 10 PREMs reported internal consistency. The risk of bias was predominantly rated as Very Good (V, $n = 10$), with no ratings classified as Adequate (A), Doubtful (D), or Inadequate (I). Most of the measurement property ratings were deemed indeterminate (? , $n = 7$), while three were rated as sufficient (+), and none were classified as insufficient (-). Structural validity was reported in three instances ($n = 3$). The risk of bias was mainly assessed as Adequate (A, $n = 3$), and all corresponding measurement property ratings were considered sufficient (+, $n = 3$). In contrast, there was no available evidence for the COSMIN domains of cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypothesis testing for construct validity, and responsiveness.

Overall rating and quality of the evidence for measurement properties of the PREMs

Following the COSMIN guidelines [9], it is essential to first determine whether the results from all available studies on a given measurement property of a PREM can be synthesised, with the primary criterion being the consistency of findings across studies.

Table 1. Characteristics of included studies evaluating PREMs, including study design, setting, population demographics, type of service, and sample size.

Author (s) (year)	PREMs	Country	Instrument language	Setting	Study design	Demographic characteristics patients		Type of eye service	Sample size
						Mean (SD) age	Percentage of male		
Au et al. [17]	Patient Experiences in Retinal Trials (PERT) Questionnaire	Australia	English	Sydney Eye Hospital	Cross-sectional study	Range (61-70 years)	53.0% male	Retinal disease service	246 responses
Charlesworth et al. [22]	Australian Glaucoma POEM	UK	English	Vision and Eye Research Institute in Cambridge.	Observational study	70 ± 12.0 years	49% male	Glaucoma service	96 responses
Fraenkel et al. [20]	Australian Glaucoma POEM	Australia	English	City Eye Centre in Brisbane.	Prospective consecutive study	69 ± 13.0 years	45% male	Glaucoma service	202 responses
Gyllén et al. [23]	PREM-questionnaire in Paediatric Cataract	Sweden	Swedish	The Paediatric Cataract Register (PECARE)	-	1. Parents age: 40.39 ± 6.41 years 2. Children age: 7.6 ± 3.89 years	1. Parents: 19.9% fathers and 81.1% mothers. 2. Children: 48.1% girls and 51.9% boys	Paediatric cataract service	1. Parents: 231 responses 2. Children: 231 responses
Somner et al. [8]	Glaucoma POEM	UK	English	1. International Glaucoma Association (IGA) 2. The NIHR Biomedical Research Centre	Qualitative study	Range (11-82 years)	39.4% male	Glaucoma service	71 responses
Boxell et al. [18]	The MDSQ 1999 and The Macular Society Questionnaire 2013.	UK	English	UK Postal Questionnaire	Cross-sectional survey	1. In 1990, (78.34 ± 7.10 years) 2. In 2013, (80.15 ± 7.99 years)	1. In 1990, 31% male 2. In 2013, 30.6% male	AMD service	1. In 1999, n = 1,187 responses 2. In 2013, n = 1,169 responses
Ayanniyi et al. [25]	Questionnaire in general eye services	Nigeria	-	Federal Medical Centre	Cross-sectional survey	62.6 ± 15.9 years	47.5% male	General eye service	99 responses
Bartlett et al. [14] *	Low Vision Service Wales (LVSW) PREM version 4.0	UK	English	Low Vision Service Wales; Wales Council of the Blind (WCB)	Cross-sectional study	Range: (40 - 75 years)	50% male	Low vision service	128 responses
Dreer et al. [19]	1. Satisfaction with Glaucoma Treatment questionnaire (Glaustat) 2. Client Satisfaction Questionnaire (CSQ)	USA	English	Department of Ophthalmology, University of Alabama at Birmingham (UAB).	-	61.0 ± 7.02 years	27% male	Glaucoma service	11 responses

Table 1. continued

Author (s) (year)	PREMs	Country	Instrument language	Setting	Study design	Demographic characteristics patients		Type of eye service	Sample size
						Mean (SD) age	Percentage of male		
Roy et al. [21]	Questionnaire in glaucoma services	India	English and translation from another unspecified language	1. Glaucoma clinic of L. V. Prasad Eye Institute. 2. Glaucoma clinic of MTC Campus. 3. Glaucoma clinic of Bhubaneswar.	Cross-sectional study	54.6 ± 16.1 years	Male / female (2:1)	Glaucoma service	103 responses
Stubbe et al. [16] *	The Consumer Quality Index Cataract Questionnaire (CQI Cataract)	Netherlands	Dutch	Dutch hospitals and through the administration of four Dutch insurance companies.	Cross-sectional study	(1) 18 – 47 years 49.3% (2) > 75 years 50.7%	37.8% Male	Cataract services	4,635 responses
Summers et al. [24]	Patient satisfaction survey question for in-person and telehealth visits.	USA	English	The Oregon Health and Science University (OHSU) Casey Eye Institute.	Retrospective review	-	-	Paediatric eye care services	1,006 responses
Varela et al. [15] *	PREM questionnaire in Inherited Retinal Disorders	UK	English	Moorfields Eye Hospital (MEH, London, UK)	Observational research	31.1 ± 11.0 years	84% male	Inherited retinal disorders service	50 responses

[Comments: “*” = The studies conduct measurement properties and “–” not report in the study].

Table 2. Comparative overview of the domains covered by identified PREMs, illustrating the distribution of items across key experience domains (administration, environment, patient-centred care, knowledge and understanding, and communication).

PREM	Scope of PREM questions				
	Administration	Environment	Patient-centred care	Knowledge and understanding	Communication
LVSU-specific PREMs (Bartlett et al. [14])	+	+	+	+	+
General NHS Wales PREMs (Bartlett et al. [14])	+		+		+
CQI Cataract: Communication with ophthalmologist (Stubbe et al. [16])	+		+		+
CQI Cataract: Communication with nurses (Stubbe et al. [16])			+		+
CQI Cataract: Communication about medication (Stubbe et al. [16])					+
IRD PREMs : Communication and information (Varela et al. [15])			+		+
IRD PREMs : Efficiency (Varela et al. [15])	+				
IRD PREMs : Patient care (Varela et al. [15])	+		+		
IRD PREMs : Motivation (Varela et al. [15])					+
IRD PREMs: Research visit versus regular clinic visit (Varela et al. [15])	+				

However, in this review, measurement properties for each PREM were only available from a single study. As a result, the summary rating for each measurement property was based on the established criteria for good measurement properties. The quality of evidence was evaluated according to four key factors: risk of bias, inconsistency, imprecision, and indirectness. In this review, no concerns were identified regarding inconsistency or indirectness.

1. LVSU-specific PREMs [14]

- **Relevance:** Overall rating was sufficient (+), but of low-quality rating. It was downgraded by two levels from high due to the risk of bias resulting from the doubtful quality of concept elicitation.
- **Comprehensiveness:** Overall rating was sufficient (+), but of low-quality evidence. It was downgraded by two levels from high due to the risk of bias resulting from the doubtful quality of concept elicitation.
- **Comprehensibility:** Overall rating was sufficient (+), but of low-quality evidence. It was downgraded by two levels from high due to the risk of bias resulting from the doubtful quality of pilot study.
- **Internal Consistency:** Overall rating was indeterminate (?), with high-quality evidence, because structural validity (unidimensionality) was not established prior to the assessment of internal consistency.

2. General NHS Wales PREMs [14]

- **Internal Consistency:** Overall rating was indeterminate (?), with high-quality evidence, because structural validity (unidimensionality) was not established prior to the assessment of internal consistency.

3. Communication with Ophthalmologist [16]

- **Structural Validity:** Overall rating was sufficient (+), with high-quality evidence.
- **Internal Consistency:** Overall rating was sufficient (+), with high-quality evidence.

4. Communication with Nurses [16]

- **Structural Validity:** Overall rating was sufficient (+), with high-quality evidence.
- **Internal Consistency:** Overall rating was sufficient (+), with high-quality evidence.

5. Communication about Medication [16]

- **Structural Validity:** Overall rating was sufficient (+), with high-quality evidence.
- **Internal Consistency:** Overall rating was sufficient (+), with high-quality evidence.

6. Communication and Information [15]

- **Relevance:** Overall rating was sufficient (+), but of very low-quality. It was downgraded by three levels from high due to

Table 3. Detailed characteristics of identified PREMs, including instrument source, theoretical framework, domain structure, scoring system, administration method, response rate, assessment timing, and recall period.

Instruments and reference	Theoretical framework	Item number and domain descriptions	Scoring type	Scoring range	Delivery method	Response rate	Completion time (Min)	Time of assessment	Recall period
Patient Experiences in Retinal Trials (PERT) Questionnaire, Au et al. [17]	None	37 items with two domains: experiences about decision-making and taking part in clinical trial	Score per item. Percentages reported against each response to each question. For 3 questions allowing for multiple responses, scoring was unclear "the % of each response was calculated by dividing the total number of each response by 80.	0-100% for each item	Self-report, an accompanying friend, family member or interpreter.	96.0%	10 min	12 months post- complete clinic.	1 day
Australian Glaucoma POEM, Charlesworth et al. [22]	None	patient staff interactions	Score per item. Percentages reported against each response to each question.	0-100% for each item	Self-report / researcher	-	-	-	3 months
Australian Glaucoma POEM, Fraenkel et al. [20]	None	8 items about patient's understanding of diagnosis and management, acceptability of glaucoma treatment.	Score per item. Patients responded on the VAS for each question.	A score between 0 and 100 of VAS scores.	Self-report	-	-	-	1 day
PREM-questionnaire in Paediatric Cataract, Gyllén et al. [23]	None	4 items about Parents' satisfaction with their child's cataract treatment and their own efforts, as well as their experience of receiving enough support from the clinic and feeling that the clinic takes their concerns seriously.	Score per item.	A score between 1 and 100 for each item	Self-report	46.0%	-	-	If the questionnaires were not returned, two reminders were sent.

Table 3. continued

Instruments and reference	Theoretical framework	Item number and domain descriptions	Scoring type	Scoring range	Delivery method	Response rate	Completion time (Min)	Time of assessment	Recall period
Glaucoma POEM, Somner et al. [8]	None	6 items with eight domains: Peace of mind, daily living, and social relations, Administration, communication, environment, patient centered care, and knowledge and understanding.	-	-	-	-	-	-	-
The MDSQ 1999 and The Macular Society Questionnaire 2013, Boxell et al. [18]	None	3 items with one domain: experiences in the diagnostic consultation.	Score per item. Percentages reported against each response to each question.	0-100% for each item	Postal self-report and Telephone interviews	71.0%	-	-	1 day
		10 items with one domain: factors associated with satisfactory healthcare experiences.	Score per item. Percentages reported against each response to each question.	0-100% for each item		39.0%			
Questionnaire in general eye services, Ayanniyi et al. [25]	None	8 items about perspectives of patients about qualitative, cost-free-to-patient eye care programs.	Score per item. Percentages reported against each response to each question.	0-100% for each item	Interview	-	-	Postoperative follow-up visits	3 days
Low Vision Service Wales (LVSU) PREM version 4.0, Bartlett et al. [14] *	None	23 items with two domains: Part 1 about LVSU-specific PREMs 14 items	Score per item. For 3 questions allowing for yes/no responses, 1 question allowing for free text box response, and 1 question for selected answer response. Scoring was unclear.	9 – 45 from nine Items (Likert 5 points scale).	Self-report / Telephone interviews	51.0%	-	-	3 weeks
		Part 2 about the general NHS Wales PREMs 9 items	Score per item. For 1 question allowing for ten scales responses, and 1 question for selected answer response. Scoring was unclear.	7 – 28 from seven Items (Likert 4 points scale).					

Table 3. continued

Instruments and reference	Theoretical framework	Item number and domain descriptions	Scoring type	Scoring range	Delivery method	Response rate	Completion time (Min)	Time of assessment	Recall period
1.Satisfaction with Glaucoma Treatment questionnaire (Glaustat) 2.Client Satisfaction Questionnaire (CSQ), Dreer et al. [19]	None	22 items with seven domains: expectations and beliefs about treatment (Ex), ease of use (EU), efficacy (Ef), undesired effects (UE), impact on health-related quality of life (IH), medical care (MC), and general satisfaction with treatment (GS).	Score per item.	A score between 1 and 100 for each item.	Self-report	86.0%	-	-	1 month
Questionnaire in glaucoma services, Roy et al. [21]	None	10 items about patient acceptability of the health promotion program	Score per item. Percentages reported against each response to each question. For 1 question for selected answer response.	0–100% for each item	Telephone interviews	-	-	-	1 day
The Consumer Quality Index Cataract Questionnaire (CQI Cataract), Stubbe et al. [16]*	None	14 items with 3 domains: eight items from communication with ophthalmologist Three items from communication with nurses Three items from communication about medication	Score per item. Percentages reported against each response to each question. For 3 questions allowing for open-ended questions response. Score per item. Score per item. Score per item.	8–32 (Likert 4 points scale). 3–12 (Likert 4 points scale). 3–12 (Likert 4 points scale).	Self-report	72.0%	-	-	1 day

Table 3. continued

Instruments and reference	Theoretical framework	Item number and domain descriptions	Scoring type	Scoring range	Delivery method	Response rate	Completion time (Min)	Time of assessment	Recall period
Patient satisfaction survey question for in-person and telehealth visits, Summers et al. [24]	None	4 items about patient satisfaction survey in paediatric ophthalmology	Score per item. Percentages reported against each response to each question.	0–100% for each item	Completed by families who had visits during March to July 2020	-	-	-	5 months
PREM questionnaire in Inherited Retinal Disorders, Varela et al. [15] *	None	27 items with five domains: nine items from Communication and information	Score per item.	9–45 (Likert 5 points scale).	Self-report	-	-	-	1 day
		Six items from Efficiency	Score per item.	6–30 (Likert 5 points scale).					
		Six items from patient care	Score per item.	6–30 (Likert 5 points scale).					
		Three items from Motivation	Score per item.	3–15 (Likert 5 points scale).					
		Three items from Research visit versus regular clinic visit.	Score per item.	3–15 (Likert 5 points scale).					

*Conducted measurement property; – not reported.

Table 4. COSMIN-based evaluation of methodological quality, risk of bias, content validity, and measurement properties of included PREM studies.

PREM	Content validity – PREM Development										Structural validity		Internal consistency	Cross-cultural validity\ measurement invariance			Reliability	Measurement error			Criterion validity	Hypotheses testing for construct validity			Responsiveness																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
	CPC					CPN					CPS			RR	R	Q		R	Q	R		Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R

CPC, Cognitive interview study or other pilot test Comprehensibility study.

Re, Relevance.

CPN, Comprehensiveness.

CPS, Comprehensibility.

RR, Rating of reviewers.

9a, Convergent validity.

9b, Known groups validity.

10a, Comparison with gold standard.

10b, Comparison with other instruments.

10c, Comparison between subgroups.

10d, Comparison before and after intervention.

R, Risk of bias rating.

C, Content validity rating.

Q, Quality of measurement property rating.

V, very good. A, adequate. D, doubtful. I, inadequate. +, sufficient. -, insufficient. ±, inconsistent.?, indeterminate.

the risk of bias resulting from the inadequate quality of concept elicitation.

- Comprehensiveness: Overall rating was indeterminate (?), with very low-quality evidence. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of concept elicitation.
- Comprehensibility: Overall rating was indeterminate (?), with very low-quality evidence. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of pilot study.
- Internal Consistency: Overall rating was indeterminate (?), with high-quality evidence, because structural validity (unidimensionality) was not established prior to the assessment of internal consistency.

7. Efficiency [15]

- Relevance: Overall rating was sufficient (+), but of very low-quality. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of concept elicitation.
- Comprehensiveness: Overall rating was indeterminate (?), with very low-quality evidence. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of concept elicitation.
- Comprehensibility: Overall rating was indeterminate (?), with very low-quality evidence. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of pilot study.
- Internal Consistency: Overall rating was indeterminate (?), with high-quality evidence, because structural validity (unidimensionality) was not established prior to the assessment of internal consistency.

8. Patient Care [15]

- Relevance: Overall rating was sufficient (+), but of very low-quality. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of concept elicitation.
- Comprehensiveness: Overall rating was indeterminate (?), with very low-quality evidence. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of concept elicitation.
- Comprehensibility: Overall rating was indeterminate (?), with very low-quality evidence. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of pilot study.
- Internal Consistency: Overall rating was indeterminate (?), with high-quality evidence, because structural validity (unidimensionality) was not established prior to the assessment of internal consistency.

9. Motivation [15]

- Relevance: Overall rating was sufficient (+), but of very low-quality. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of concept elicitation.
- Comprehensiveness: Overall rating was indeterminate (?), with very low-quality evidence. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of concept elicitation.

- Comprehensibility: Overall rating was indeterminate (?), with very low-quality evidence. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of pilot study.
- Internal Consistency: Overall rating was indeterminate (?), with high-quality evidence, because structural validity (unidimensionality) was not established prior to the assessment of internal consistency.

10. Research Visit vs. Regular Clinic Visit [15]

- Relevance: Overall rating was sufficient (+), but of very low-quality. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of concept elicitation.
- Comprehensiveness: Overall rating was indeterminate (?), with very low-quality evidence. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of concept elicitation.
- Comprehensibility: Overall rating was indeterminate (?), with very low-quality evidence. It was downgraded by three levels from high due to the risk of bias resulting from the inadequate quality of pilot study.
- Internal Consistency: Overall rating was indeterminate (?), with high-quality evidence, because structural validity (unidimensionality) was not established prior to the assessment of internal consistency.

Based on the overall ratings and quality of evidence across COSMIN domains for each PREM, recommendations for their use in research and clinical practice were made (as shown in Table 5, final row). The recommendations are as follows:

- Category B: LVSW-specific PREMs [14]. In addition, communication-related domains—including communication with ophthalmologists [16], communication with nurses [16], and communication about medication [16], General NHS Wales PREMs [14], Communication and information [15], Efficiency [15], Patient care [15], Motivation [15], and Research visit versus regular clinic visit [15].

There were no PREMs met the criteria for Category A, primarily because the available evidence was insufficient to demonstrate both sufficient content validity and at least low-quality evidence for sufficient internal consistency, as stipulated by the COSMIN guideline. Similarly, no PREMs fulfilled the requirements for Category C, as high-quality evidence supporting an insufficient measurement property was not identified for any instrument. Consequently, all included PREMs were assigned to Category B, indicating their potential applicability while underscoring the need for further psychometric validation to establish the robustness of their measurement properties.

DISCUSSION

This systematic review critically evaluated the measurement properties of patient-reported experience measures (PREMs) used in eyecare services in accordance with the COSMIN guidelines. Although several instruments demonstrated promising psychometric characteristics, none fulfilled the stringent COSMIN criteria for recommendation as Category A instruments. This was primarily attributable to insufficient evidence supporting content validity and/or the absence of evidence establishing structural validity, which precluded a sufficient rating for internal consistency in several instruments. Consequently, all identified PREMs require further psychometric evaluation before they can be recommended for routine use in clinical practice. These findings highlight the need for more rigorous instrument development and validation, particularly

Table 5. Overall ratings, quality of evidence (GRADE approach), and recommendations for use of identified PREMs.

COSMIN Domains	PREM																				
	LVSU-specific PREMs		General NHS Wales PREMs		Communication with ophthalmologist		Communication with nurses		Communication about medication		Communication and information		Efficiency		Patient care		Motivation		Research visit versus regular clinic visit		
	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	OR	QoE	
Content Validity																					
- Relevance																					
	+	L										+	VL	+	VL	+	VL	+	VL	+	VL
- Comprehensiveness																					
	+	L										?	VL	?	VL	?	VL	?	VL	?	VL
- Comprehensibility																					
	+	L										?	VL	?	VL	?	VL	?	VL	?	VL
Structural validity																					
Internal consistency																					
	?	H	+	H	+	H	+	H	?	H	?	H	?	H	?	H	?	H	?	H	H
Cross-cultural validity\ measurement invariance																					
Reliability																					
Measurement error																					
Criterion validity																					
Hypotheses testing for construct validity																					
Responsiveness																					
Recommendation																					
	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B

OR Overall rating, QoE Quality of evidence. + sufficient, - insufficient, ? indeterminate. H High, M Moderate, L Low, VL Very low. Recommendation level: A indicated strongly recommendation, B indicated weak recommendation, C indicated not recommendation.

with respect to content validity, structural validity, and internal consistency, to ensure that PREMs accurately capture patients' experiences across diverse eyecare settings. This review is, to the best of our knowledge, the first to systematically compile and assess the psychometric properties of PREMs in eyecare.

The results highlight a notable lack of PREMs with good quality evidence of measurement properties in eyecare. Importantly, there remains a marked scarcity of empirical evidence concerning PREMs specifically tailored to eye care, thereby constraining the transferability of existing findings to eye care contexts. Accordingly, there is a compelling need for the systematic development and robust psychometric validation of eye care-specific PREMs that can more precisely reflect the distinctive features of patient experience in ophthalmic services. The lack of Category A recommendations indicates critical shortcomings in essential psychometric qualities, particularly in terms of content and structural validity. Validity pertains to the extent to which a tool accurately measures what it is intended to assess [26]. Assessing these properties is crucial for identifying possible biases and determining whether the tools are suitable for gathering patient-reported experience data [12]. This brings into question the capacity of these tools to truly capture patients' experiences in eyecare settings, which may compromise their effectiveness in informing clinical practice and shaping healthcare policies. Additionally, most studies failed to demonstrate robust content validity, often due to a lack of direct patient involvement during the development process. This undermines the credibility of these instruments and limits their acceptance in patient-centred care models. Future tool development must include qualitative methods such as interviews or focus groups with diverse patient populations, as well as eliciting staff perspectives, to ensure relevance, comprehensibility and comprehensiveness. Structural validity assesses whether the scores of a measure are an adequate reflection of the dimensionality of the construct being measured. The absence of supporting evidence for structural validity in most tools reviewed further questions their reliability and interpretability.

Although Rasch analysis has been applied to PREMs in other specialties, including oncology [5] and patients with chronic illness [27], notably, none of the studies reviewed herein employed modern psychometric approaches such as Item Response Theory (IRT) or Rasch Measurement Theory (RMT). These models allow for more nuanced assessment of item performance and scale unidimensionality, providing a higher standard of measurement precision. The reliance on classical test theory across all studies is a missed opportunity for enhancing the psychometric robustness of PREMs in eyecare.

The methodological limitations are particularly relevant when considering the broader international context of PREM development and usage. The widespread adoption of PREM data in health care began in the United States and the United Kingdom, as these two countries were among the first to develop PREMs [28]. Later, in a systematic review [29], which sought to identify general healthcare PREMs, validity and reliability was assessed, and potential biases in the study designs were investigated. Their research found that the majority of PREMs originated in the UK (18 PREMs), followed closely by the US with 17 PREMs [29]. Additionally, in Australia, the importance of developing PROM and PREM programs has been highlighted, through establishing nationally standardised measures and survey frameworks [30]. The absence of consistent national PROMs and PREMs has been reported to hinder hospital data collection practices, restrict opportunities for national benchmarking, and limit quality improvement efforts [31]. Interestingly, the prevalence of PREMs developed in the UK can be attributed to the government's emphasis on patient-centred care and the importance placed on PREMs in healthcare evaluation. In the U.K., national standardised patient experience surveys became mandatory by 2010 under NHS policy [32, 33]. In fact, current UK health policy explicitly supports the systematic collection and use of PREMs as part of broader efforts to enhance person-centred and

coordinated care, embedding the patient voice into service evaluation and improvement and aligning with national frameworks that prioritise partnership with people and communities in health service design and delivery [34, 35]. Conversely, the results presented in this study revealed no study from countries in Asia, however, the search was limited to studies published in English language only. It is possible that inherent cultural differences between Asian and Western healthcare systems may influence differences in practitioner-patient communication and how patients openly share feedback about their care experiences and outcomes [36]. Future development of PREM tools may consider ethnocultural tailoring, aligned with the community's living experiences [37].

Ideally, a PREM should provide good quality evidence that it captures experiences that are meaningful to patients, that it is relevant and produces reliable and interpretable scores. The limitations identified in existing instruments highlight the need for more rigorous development, including item generation with patient and public involvement and clinical expertise, cognitive interviewing to assess content validity. Subsequent evaluation should then assess the performance of response options, explore the dimensional structure of the instrument in which multiple domains are proposed, and evaluate key measurement properties including structural validity, internal consistency, construct validity, measurement invariance, measurement error, responsiveness, and reliability. A multi-phased approach, with repeated assessments and comparison against related measures where appropriate, would provide strong evidence to support a PREM.

Strengths and limitations

A key strength of this review is the systematic approach guided by COSMIN standards, which ensured a comprehensive and methodologically sound assessment of each PREM's measurement properties. Despite the strengths and comprehensive approach of this study, several limitations must be acknowledged. The grey literature sources were not included in the identification of PREMs, which may have led to the omission of relevant PREMs that met the inclusion criteria. Additionally, this review utilised search filters in only two databases, PubMed and Web of Science, as the COSMIN guideline-supported search filters were not applicable across all databases. Only English languages studies were included in the review, therefore studies published in other languages were not represented. Importantly, there remains a notable paucity of data on PREMs specific to eye care, which limits the applicability of findings to eye care settings. This gap highlights a critical area for future research and underscores the need for the development and validation of eye care-specific PREMs to better capture patient experiences within this specialty, which is highly relevant to both patients and service-providers in the context of eye care.

CONCLUSION

In conclusion, this systematic review identified 10 PREMs in the field of eyecare, with the most promising being the LVSW PREMs. Furthermore, this review highlights the urgent need for robust PREMs with good quality evidence of measurement properties in eyecare services. Although some tools show promise, none currently meet the standards required for recommendation under COSMIN guidelines. Enhancing the methodological rigor of PREM development and validation is critical to ensuring that patient voices are accurately and effectively integrated into eyecare quality assessment and improvement initiatives. To move the field forward, future studies should prioritise the rigorous development of eyecare-specific PREMs and the generation of good quality evidence for their measurement properties. This includes involving patients during the development process to strengthen content validity, using both exploratory and confirmatory factor analyses to confirm structural validity, and applying IRT or RMT to evaluate measurement accuracy and item

performance. Additionally, it is important to ensure the tools are culturally and contextually appropriate for diverse patient groups, and to carry out longitudinal research to assess their responsiveness and reliability over time.

Supplemental material is available at Eye's website.

REFERENCES

- Jamieson Gilmore K, Corazza I, Coletta L, Allin S. The uses of Patient Reported Experience Measures in health systems: A systematic narrative review. *Health Policy*. 2023;128:1–10.
- Reynolds A. Patient-centered care. *Radiologic technology*. 2009;81.
- Virdun C, Garcia M, Phillips JL, Luckett T. Description of patient reported experience measures (PREMs) for hospitalised patients with palliative care needs and their families, and how these map to noted areas of importance for quality care: A systematic review. *Palliat Med*. 2023;37:898–914.
- 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;29:314–26.
- Sawyer CS, Taylor S, Carter L, Stanworth M, Davies M, Thistlethwaite F, et al. Development and validation of a patient reported experience measure for experimental cancer medicines (PREM-ECM) and their carers (PREM-ECM-Carer). *BMC Cancer*. 2024;24:500.
- Steiner B, Neumann A, Schwertfeger M, Verket M, Romero I, Hurtado C, et al. Patient-reported outcome and experience measures in cardiovascular disease: a scoping review as part of iCARE4CVD. *J Patient-reported Outcomes*. 2025;9:141.
- Coulter A. Measuring what matters to patients. *Br Med J Publ Group*. 2017;2017:356:j816.
- Somner JE, Sii F, Bourne RR, Cross V, Burr JM, Shah P. Moving from PROMs to POEMs for glaucoma care: a qualitative scoping exercise. *Invest Ophthalmol Vis Sci*. 2012;53:5940–7.
- Mokkink LB, Elsmann EB, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. *Qual Life Res*. 2024;33:2929–39.
- Mihaljevic AL, Doerr-Harim C, Kalkum E, Strunk G. Measuring patient centeredness with German language Patient-Reported Experience Measures (PREM)—A systematic review and qualitative analysis according to COSMIN. *PLoS ONE*. 2022;17:e0264045.
- Huang J, Yang J, Han M, Xue Z, Xu M, Qi H, et al. Psychometric evaluation of patient-reported experience measures for peri-anesthesia care: a systematic review based on COSMIN guidelines. *Int J Nurs Stud*. 2025;161:104930.
- Terwee CB, Prinsen C, Ricci Garotti M, Suman A, De Vet H, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res*. 2016;25:767–79.
- Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5:210.
- Bartlett R, Absalom R, Williams O, Withers K. Development and pilot of a community low-vision, service-specific Patient-Reported Experience Measure. *Ophthalmic Physiol Opt*. 2023;43:702–9.
- Varela MD, Hashem S, Sumodhee D, Michaelides M. Patient-Reported Experience Measurements From Individuals With Inherited Retinal Disorders Involved in Observational Research. *Transl Vis Sci Technol*. 2024;13:9.
- Stubbe JH, Brouwer W, Delnoij DM. Patients' experiences with quality of hospital care: the Consumer Quality Index Cataract Questionnaire. *BMC Ophthalmol*. 2007;7:14.
- Au CP-Y, Fardell N, Williams M, Fraser-Bell S, Campaign A, Gillies M. Patient experiences in retinal trials: a cross-sectional study. *BMC Ophthalmol*. 2015;15:80.
- Boxell EM, Amoaku WM, Bradley C. Healthcare experiences of patients with age-related macular degeneration: have things improved? Cross-sectional survey responses of Macular Society members in 2013 compared with 1999. *BMJ Open*. 2017;7:e012790.
- Dreer LE, Owsley C, Campbell L, Gao L, Wood A, Girkin CA. Feasibility, patient acceptability, and preliminary efficacy of a culturally informed, health promotion program to improve Glaucoma medication adherence among African Americans: "G laucoma management O ptimism for the frican Americans L iving with Glaucoma" (GOAL). *Curr Eye Res*. 2016;41:50–8.
- Fraenkel A, Lee GA, Vincent SJ, Vincent RA, Bourne RR, Shah P. Lessons learned from the development and implementation of a patient-reported outcome and experience measure (POEM) in an Australian glaucoma practice. *BMC Ophthalmol*. 2019;19:192.
- Roy AK, Rao A, Pal A, Mohapatra S, Swain R. Evaluating perspective and quality of life of glaucoma patients during the COVID-19 pandemic in India: Results of a telephone survey. *Indian J Ophthalmol*. 2023;71:576–9.
- Charlesworth E, Jolly JK, Farrell S, Bourne R, Pardhan S. Patient-reported outcome and experience measures (POEM) of a community-based glaucoma clinic in Cambridge, UK: an observational study. *BMJ Open*. 2024;14:e077906.
- Gyllén J, Magnusson G, Forsberg A. Parents' reported experiences when having a child with cataract—important aspects of self-management obtained from the Pediatric Cataract Register (PECARE). *Int J Environ Res Public Health*. 2020;17:6329.
- Summers AI, Kuo A, Zaback T, Loh AR, Brinks MV, Hribar MR. Pediatric ophthalmology provider and staff attitudes and patient satisfaction in telehealth implementation during COVID-19. *Telemed e-Health*. 2022;28:675–81.
- Ayanniyi AA, Bob-Egbe S, Olatunji FO, Omolase CO, Omolade EO, Ojehomon FO, et al. Social marketing potential of qualitative cost-free-to-patient eye care program in a Nigerian community. *Ann African Med*. 2009;8.
- Scholle SH, Roski J, Adams JL, Dunn DL, Kerr EA, Dugan DP, et al. Benchmarking physician performance: reliability of individual and composite measures. *Am J Managed Care*. 2008;14:833.
- Statton R, Rosenlund L, Close J, Dencker A, Lloyd H. Patient-reported experience measures for person-centred care: a survey for the development and psychometric testing of a targeted measure, the EPCC-14. *BMJ Open*. 2025;15:e098050.
- Garratt A, Solheim E, Danielsen K. National and cross-national surveys of patient experiences: a structured review. 2008.
- 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.
- Jones CH, Woods J, Brusco NK, Sullivan N, Morris ME. Implementation of the Australian Hospital Patient Experience Question Set (AHPEQS): a consumer-driven patient survey. *Aust Health Rev*. 2021;45:562–9.
- Bull C, Callander EJ. Current PROM and PREM use in health system performance measurement: still a way to go. *Pat Exper J*. 2022;9:12–18.
- Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. *BMJ Qual Saf*. 2012;21:600–5.
- Wong E, Mavondo F, Fisher J. Patient feedback to improve quality of patient-centred care in public hospitals: a systematic review of the evidence. *BMC Health Serv Res*. 2020;20:530.
- NHS England. Experience of Care Improvement Framework 2025 [Available from: <https://www.england.nhs.uk/long-read/experience-of-care-improvement-framework/>].
- NHS England. Working in Partnership with People and Communities: Policy and Guidance 2025 [Available from: <https://www.england.nhs.uk/publication/working-in-partnership-with-people-and-communities/>].
- Pun JK, Chan EA, Wang S, Slade D. Health professional-patient communication practices in East Asia: An integrative review of an emerging field of research and practice in Hong Kong, South Korea, Japan, Taiwan, and Mainland China. *Patient Educ Counsel*. 2018;101:1193–206.
- Takeshita J, Wang S, Loren AW, Mitra N, Shults J, Shin DB, et al. Association of racial/ethnic and gender concordance between patients and physicians with patient experience ratings. *JAMA Netw Open*. 2020;3:e2024583.
- Elsmann EBM, Mokkink LB, Terwee CB, Beaton D, Gagnier JJ, Tricco AC, et al. Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *Qual Life Res*. 2024;33:2029–46. <https://doi.org/10.1007/s11136-024-03634-y>.

AUTHOR CONTRIBUTIONS

All authors were responsible for the review protocol. PK and JHA conducted the search, screening, extracting and analysing data, assessment against COSMIN criteria and initial drafting of the paper. TP and BR were involved in arbitrating potentially eligible studies and analysing the data. All authors contributed oversight of data analysis, interpreting results and reviewing the final manuscript.

FUNDING

The work in this paper was in part funded by a PhD studentship from The Royal Thai Government Scholarship (PK). JHA, BR & TP are funded by Catalytic Infrastructure funding from Health and Care Research Wales for the Centre for Vision Services Research. FS was funded by the Emerging Researcher Grant as part of Health and Care Research Wales. Outside the submitted work, Dr McAlinden has received consultancy fees/honorarium/travel support (past 36 months) from: AkkoLens International B.V. (Rijswijk, the Netherlands), Bausch and Lomb (Bridgewater, New Jersey, USA), Bayer (Leverkusen, Germany), British Society of Refractive Surgery (Oxford, UK), BVI (Liège, Belgium), Cutting Edge (Labège, France), Hoya (Frankfurt, Germany), Johnson & Johnson Surgical Vision (Santa Ana, California, USA), Medevise Consulting SAS (Strasbourg, France), Rayner (Worthing, UK), ROHTO (Tokyo, Japan), Scope (Crawley, UK), SightGlass vision (Menlo Park, California, USA), SpyGlass (Aliso Viejo, California, USA), Thea pharmaceuticals (Clemont-Ferrand, France), Vold Vision (Arkansas, USA). Dr McAlinden developed the Quality of Vision (QoV) questionnaire

and the Orthokeratology and Contact Lens Quality of Life Questionnaire (OCL-QoL), and has a financial interest in these tools. He also consults on topics including Rasch analysis, questionnaires, statistical analyses, and clinical/surgical ophthalmology topics. Dr McAlinden has been co-awarded research grants from the Welsh Government and the Royal College of Ophthalmologists (unpaid roles) and Catalytic funding from Health and Care Research Wales to develop a Centre for Vision Services Research (unpaid role). Dr McAlinden is a council member and treasurer of the British Society for Refractive Surgery (unpaid role) and a PROM advisor to the Royal College of Ophthalmologists (unpaid role). Dr McAlinden has undertaken paid peer reviews for Research Square (Durham, North Carolina, USA). Dr McAlinden is an editorial board member for *Eye and Vision*, *Graefes Archive for Clinical and Experimental Ophthalmology*, *Archives of Medical Science*, *Journal of Clinical Medicine*, *Journal of Ophthalmology*, and *Journal of Clinical and Experimental Ophthalmology*. Dr McAlinden is an Associate Editor for *Frontiers in Medicine – Ophthalmology*.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41433-026-04907-1>.

Correspondence and requests for materials should be addressed to Jennifer H. Acton.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026