

The use of patient-reported outcome research in modern ophthalmology: impact on clinical trials and routine clinical practice

Tasane Braithwaite^{1,2}

Melanie Calvert^{1,3}

Alastair Gray⁴

Konrad Pesudovs⁵

Alastair K Denniston^{1,6–8}

¹Centre for Patient Reported Outcomes Research and NIHR Birmingham Biomedical Research Centre, University of Birmingham, Edgbaston, Birmingham, UK;

²Moorfields Eye Hospital, London, UK; ³Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham, UK;

⁴Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK; ⁵Consultant, Adelaide, SA, Australia;

⁶Department of Ophthalmology, Queen Elizabeth Hospital Birmingham, University Hospitals Birmingham NHSFT, Birmingham, UK; ⁷Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK; ⁸NIHR Biomedical Research Centre (Moorfields Eye Hospital/UCL), London, UK

Abstract: This review article considers the rising demand for patient-reported outcome measures (PROMs) in modern ophthalmic research and clinical practice. We review what PROMs are, how they are developed and chosen for use, and how their quality can be critically appraised. We outline the progress made to develop PROMs in each clinical subspecialty. We highlight recent examples of the use of PROMs as secondary outcome measures in randomized controlled clinical trials and consider the impact they have had. With increasing interest in using PROMs as primary outcome measures, particularly where interventions have been found to be of equivalent efficacy by traditional outcome metrics, we highlight the importance of instrument precision in permitting smaller sample sizes to be recruited. Our review finds that while there has been considerable progress in PROM development, particularly in cataract, glaucoma, medical retina, and low vision, there is a paucity of useful tools for less common ophthalmic conditions. Development and validation of item banks, administered using computer adaptive testing, has been proposed as a solution to overcome many of the traditional limitations of PROMs, but further work will be needed to examine their acceptability to patients, clinicians, and investigators.

Keywords: patient-reported outcome measures, Rasch analysis, eye disease, randomized controlled trials

Introduction

Recent years have seen greater awareness of the importance of the patient voice in ophthalmology.⁶ This paradigm shift influences our understanding of the impact of disease, and the efficacy of interventions, with implications for both clinical practice and clinical trials. There has been a move away from the sole use of traditional outcome metrics (eg, visual acuity, intraocular pressure [IOP]) toward inclusion of metrics that matter as much, or possibly more, to patients and providers (eg, symptoms, quality of life [QoL], convenience, and cost of treatment). Patient-reported outcome measures (PROMs) seek to comprehensively capture these important outcomes.

PROMs are increasingly used in clinical trials to assess the impact of treatment from the patient perspective. They offer particular value as the primary outcome measure where two interventions have been established to be equally efficacious in terms of a traditional outcome measure (eg, IOP lowering effect), but where differences are anticipated in terms of side effects, cost, and convenience. Multiple randomized controlled clinical trials (RCTs) have recently completed or are in progress using PROMs as the primary outcome measure.^{7–10} PROM data from trials may be used to inform pharmaceutical labeling claims, clinical guideline development, reimbursement

Correspondence: Tasane Braithwaite
Moorfields Eye Hospital, 162 City Road,
London EC1V 2PD, UK
Email tasanebraithwaite@gmail.com

decisions, and health policy. In addition, PROMs have potential application in clinical governance and quality assurance, performance management of health care providers, and integration into routine clinical practice.^{11,12} The international consortium for health outcome measurement (ICHOM) has proposed standard outcome sets including PROMs for clinical assessment of cataract (using Catquest-9SF) and age-related macular degeneration (AMD) (using the Impact of Vision Impairment, IVI) (<http://www.ichom.org/medical-conditions/>). Reporting of ICHOM outcome sets for ophthalmic conditions was recently mapped to current reporting practices in eight large eye centers internationally.¹⁴ This exercise revealed wide variation in current reporting practice, and no reporting of vision or eye disease-related PROMs by any hospital. Potential barriers to extend the use of PROMs in routine clinical care include logistical, social, legal, technical, and cultural factors.¹²

This review outlines what PROMs are, explores PROMs development, and probes the extent to which they have had meaningful impact on clinical and research practice in modern ophthalmology.

What are PROMs?

PROMs are sets of questions, or “items,” that capture information on health from the patients’ perspective. Some PROMs provide rudimentary summary information, while others provide detailed measurement suited to statistical analysis. Measurement of PROMs began in the 1950s, and there has been rapid expansion in the past two decades in all fields of health care, including ophthalmology. A small survey in 1998 revealed that very few UK ophthalmologists were familiar with QoL outcome measures.¹⁵ In 2001, Massof and Rubin reported that more than 12 PROMs had been developed since 1980.¹⁶ Now there are more than 160 PROM instruments in ophthalmology and optometry.¹⁷ Many have been developed for use in glaucoma, cataract, and low vision, but there are no validated PROMs for a large number of eye diseases and interventions.

There are many generic instruments (eg, EQ-5D, Short Form [SF]-36, Health Utilities Index-3 [HUI-3]), vision-related instruments (eg, IVI, National Eye Institute-Vision Function Questionnaire, NEI-VFQ-25), and ocular disease-specific PROM instruments (eg, Catquest-9SF). The plethora of available instruments presents a challenge – how should one be selected for use in clinical practice or a clinical trial?

Selecting a PROM

There is no “gold standard” PROM. In order to select a PROM, investigators and clinicians must choose the latent

trait that they want to measure, in consultation with patients and their carers. This might be the impact of disease, or treatment, on symptoms, daily activities, emotional well-being, or side effects, measured at one time point or longitudinally. The choice of PROM will depend on the rationale for assessment. For example, if the data will be used to provide in-depth information to clinicians and patients on the impact of disease, then a disease-specific measure may be most appropriate. However, if data will be used for health economic evaluation, then a health utility measure, which seeks to take account of preferences for different health states, such as the EQ-5D, will be required. Impact on QoL is a frequently desired outcome measure, particularly by health policy makers. However, the challenge with measuring QoL is that it is a multidimensional construct. The latent traits encompassed within vision and eye disease-related QoL are proposed to include visual symptoms, ocular surface symptoms, general symptoms, emotional well-being, activity limitation, mobility, convenience, health concerns, social well-being, and economic well-being, but are not necessarily limited to these.¹⁷ Each trait, or domain of interest, requires due consideration and measurement. Having chosen the latent trait or traits to be measured, a PROM can be selected from the pool of available instruments, and piloted for use to establish validity for use in a new patient or population context, or a new PROM may need to be developed. To better understand the multiple factors that should be considered, the next section explores what the ideal PROM might look like.

What does the ideal PROM look like?

Multiple approaches to evaluate PROMs have been proposed^{18–22} and have informed the US Food and Drug Administration’s guidance.²³ In brief, important considerations include 1) content development; 2) the psychometric properties of the instrument (judged using either Classical Test Theory, or Item Response Theory approaches, including Rasch analysis – see Boxes 1 and 2 for more details); 3) responsiveness; and 4) administration burden and resource implications.

The ideal PROM contains a necessary and sufficient set of questions (content) to measure a single underlying construct such as ocular surface symptoms (unidimensionality), or, for a multidimensional construct like QoL, a series of sets of questions, each demonstrating unidimensionality and together targeting each important element of the multidimensional construct. It has a logical order of evenly spaced response categories. It is reliably able to distinguish between patients with different abilities or degrees of severity for each

Box 1 Classical test theory and item response theory

Analysis of patient-reported outcome measures has been centered on two approaches, Classical Test Theory (CTT), and Item Response Theory (IRT).²

In CTT, each item is assumed to have equal difficulty, and each response score is assumed to have equal weight (eg, a score of 4 for “extreme difficulty” is assumed to have twice the value of a score of 2 for “mild difficulty”). Summary scores are assumed to represent measurement of the underlying trait (eg, quality of life). In IRT, both items and responders are scaled according to responses, which are assumed to reflect the different ability of responders and the different difficulty of items. Ordering of category responses is explicitly tested to ensure that “extreme difficulty” scores more highly across responders than “mild difficulty.” Rasch analysis is a special case of IRT, where the data are fit to a simple measurement model. This creates valid measurement to which parametric statistics can be applied. Massof has discussed the theoretical constructs and methodology as applied to ophthalmology in detail.¹³

Box 2 The importance of Rasch analysis

The Rasch model provides interval-level scoring to enable examination of each unidimensional construct. Rasch analysis permits quantitative psychometric assessment of each latent trait and generates measurement data that are readily amenable to statistical analysis, whereas summary scoring does not. The classical test theory approach, which does not use Rasch analysis, is defined by the use of summary scoring (simple adding up of ordinal values assigned to response options) and a high-level reliance on simple reliability statistics like Cronbach’s alpha. The latter statistic is calculated from pairwise correlations between items and provides rudimentary insight into an instrument’s internal consistency. Gothwal et al have argued that Rasch scaling achieves smaller standard errors of the measures and further enhances precision by applying a logistic transformation to expand the range of measurement, thereby reducing ceiling and floor effects.¹ An important implication of this for clinical trials is that Rasch-validated instruments require a smaller sample size to detect significant differences in outcomes.

Rasch analysis has therefore been used to “re-engineer” some of the popular existing instruments, such as the Visual Function Index-14 (VF-14) and NEI-VFQ-25. For example, the VF-14 was used before and after cataract surgery, and re-engineered into a shorter instrument, achieving both reduced respondent burden and administration time, and precision 2.5 times greater than the original instrument.¹ Doubling the precision of the primary outcome measure halves the required sample size, with very important cost implications for clinical trials. Flaws in the psychometric properties of the, widely used, NEI-VFQ-25 have been identified by multiple investigators, and Rasch re-engineered instruments have been proposed.^{3–5}

item, and between patients at each end of the range of the construct (measurement precision), with neither significant floor nor ceiling effects (targeting). The instrument score correlates with important clinical measures such as visual acuity (concurrent validity), and with any existing instruments purporting to measure the same construct (convergent validity), while not correlating highly with instruments purporting to measure a different construct (discriminant validity). It can discriminate between clinically distinct groups and is responsive to detecting clinically important changes over time. The instrument demonstrates test–retest reliability when repeated. Specific quality scoring criteria have been outlined in detail by Prem Senthil et al.²⁴ Further considerations include the cost of the instrument (some are not freely available), the availability of the PROM in different languages, the staffing administration requirements, and the patient response burden. One of the most frequent trade-offs that must be made is between selecting a short PROM that is readily applicable in a busy clinical or research context, and selecting a PROM that provides comprehensive insight,

but takes much longer to administer. It is also important to note that a PROM developed for one target disease or patient population in one cultural context may have poor targeting of item difficulty to respondent ability in another disease or population, which is why validation is necessary.¹⁷

Generic PROMs

Generic, multi-attribute, health-related utility instruments have been used for over three decades, and the most widely used include the EQ-5D, SF-6D, and HUI. In these instruments, answers to a series of questions yield raw health state scores that can be transformed into a utility value, where 1 represents perfect health and 0 is death. Utility values are used to calculate quality-adjusted life years (QALY) lost or gained as a result of a disease state or health care intervention. The health state weights are obtained using cardinal preference measurement approaches, such as the time trade-off or the standard gamble.

The EQ-5D instrument was developed by the EuroQol Group almost 30 years ago.²⁵ It has been translated into

over 100 official languages and is widely used. It includes five questions on mobility, self-care, usual activities, anxiety/depression, and pain/discomfort. The original EQ-5D includes three levels (3L) for each question, resulting in 243 possible health states. A five-level (5L) instrument has been introduced more recently, yielding 3,125 health states.²⁶ A further three bolt-on items have been developed for EQ-5D, including a vision bolt on.²⁷ The preference weights for the EQ-5D-3L were originally obtained from a UK population sample using time-trade off, with regression analysis to estimate a value for each of the health states.²⁸ Valuation sets have since been obtained through various approaches in many other countries, and differences between valuation sets are generally small.²⁹ The original EQ-5D scale using the UK valuations extends from -0.59 to 1.00 ,²⁸ and a more recent UK value set for the EQ-5D-5L extends from -0.28 to 0.95 .³⁰ The mean minimally important difference reported in a review of eight studies in different conditions was 0.074 (range -0.011 to -0.140).³¹ A visual analog scale (VAS) is recommended for use alongside the EQ-5D. This consists of a “thermometer” scale from 0 to 100, on which the respondent is asked to indicate the point that best represents their own health on that day.

The Short Form (SF-6D) includes eleven items in six domains, including physical functioning, role limitations, social functioning, pain, vitality, and mental health.³² This instrument yields 18,000 health states. Items were extracted from the larger, 36-item instrument (SF-36), which was developed for the Medical Outcomes Study.^{33,34} Preference weights are obtained from a UK population-representative sample and models derived to provide utility values for each health state. The SF-6D scale extends from 0.29 to 1.00 , and a review of eight studies in different conditions estimates the mean minimally important difference to be 0.041 (range from 0.011 to 0.097).³¹ The SF-36, and a shorter version – the SF-12 – are also still frequently used in studies to assess aspects of QoL more fully, where obtaining a utility value is not the primary objective.

The Health Utilities Index was developed in the early 1980s in Canada to assess outcomes in low birth weight infants.³⁵ Six domains are captured by HUI version 2 (HUI-2) including sensation, mobility, emotion, cognition, self-care, pain, and fertility.³⁶ Each has between 3 and 5 levels, resulting in 24,000 possible health states. Valuations were originally obtained from Canadian parents using standard gamble and a VAS. Version 3 (HUI-3) expands the sensation domain into vision, hearing, and speech and yields 972,000 health states.³⁷ Valuations are elicited from the general public in Canada and

a utility function estimates for each of the domains, and for the overall instrument.

Up to three decades of experience with these instruments highlights that they yield differing utility values in head-to-head comparisons. In seven health conditions, not including vision disorders, SF-6D is found to have a smaller range and lower variance in values than EQ-5D.³⁸ Differences result in the estimation of different estimates of quality-adjusted survival for the same intervention and thus differing conclusions in relation to cost-effectiveness. As a result, some funding bodies are explicit about which instrument and valuation method they prefer. In England, the National Institute for Health and Clinical Excellence (NICE) prefers EQ-5D, but even among NICE Technology Appraisals, there is considerable variation in the methods used to select and incorporate utility values in economic models.³⁹ Health state valuations obtained from the general public, rather than from patients or clinical experts, are also generally preferred.

The limitation of generic PROMs is that they may lack sensitivity for the impact of eye disease and its treatment. For example, a vision-related QoL instrument, the Vision Function (VF-14), identified significant benefit of cataract surgery at 3 months, but the SF-36 found no significant benefit.⁴⁰ While the very brief preference-based generic QoL instruments such as EQ-5D are unable to capture QoL outcomes comprehensively, their shortness and ease of administration face to face, or by telephone, postal questionnaire, SMS messaging, web or email usually results in higher response, and completion rates than longer questionnaires. Moreover, the ability to transform raw scores into utility values provides wide application across different populations and medical specialties, thereby securing their role as important PROMs in informing resource allocation and reimbursement decisions, which typically have to make comparisons across a wide range of different disease areas. Partly in consequence, they are also increasingly used in medical product development.⁴¹

Some investigators, seeking instruments more sensitive to vision-related preference, have recommended use of the Vision Preference Value Scale, first validated in 2004, in which a score of 0 is equivalent to an outcome as bad as death, and a score of 1.0 is equivalent to perfect vision.⁴² However, caution is needed in interpreting the findings of studies using a “vision-truncated scale,” and scales anchored by vision are not generally used in cost-effectiveness analysis.⁴³

Vision-related PROMs

There are many instruments that focus on the impact of vision impairment and ocular symptoms and signs on different

domains of QoL, such as the NEI-VFQ-25, the IVI, and the VF-14. These are typically referred to as vision-related or ophthalmic PROM instruments, and for consistency we have used the former throughout. Khadka et al conducted a systematic review for vision-related PROM instruments demonstrating interval measurement properties and identified 48 (out of 121 instruments in total). They appraised the quality of each against criteria similar to those proposed by the “Consensus-based Standards for the selection of Measurement Instruments” group⁴⁴ and highlighted those of higher quality, by ophthalmic subspecialty.⁴⁵ Where no disease-specific PROM exists, the IVI has been proposed as being valuable for assessing domains including the ability to read and access information, mobility, and emotional well-being.¹⁷ A shorter version, the (15-item) Brief IVI, has also been validated.⁴⁶

Impact of PROMs by ophthalmic subspecialty

It is beyond the scope of this review to critique all vision-specific and eye disease-specific PROMs. The following sections highlight examples of the more frequently used, or better validated PROMs in ophthalmology, by subspecialty area, and illustrate examples of their impact.

Narrative review search methodology

We performed a PubMed search for “patient reported outcome” and terms relating to each subspecialty, dated January 1990 to September 30, 2018 with no field restrictions. This identified 4,114 hits (Table S1). We screened these to identify systematic reviews of PROMs, RCTs reporting PROMs, and examples of the use of PROMs in clinical practice. In addition, we reviewed the Cochrane Eyes and Vision database (<https://eyes.cochrane.org/>). This revealed that across all subspecialties, relatively few RCTs contained within systematic reviews of interventions have, to date, reported PROMs or economic outcome measures. Greatest progress in terms of developing PROMs and introducing them into RCTs have been made in low vision, medical retina and glaucoma.

Glaucoma

Vandenbroeck et al published a systematic review of PROM instruments in glaucoma in which the search, dated to December 2010, identified 27 instruments, 18 of which were disease specific.⁴⁷ The authors highlighted that the instruments mostly lacked a conceptual framework, had been tested using classical validation techniques, and that item generation strategies had not involved the patients’

perspective adequately. Another systematic review by Che Hamzah et al, in which the search dated to January 2009, cataloged 33 instruments.⁴⁸ They highlighted the NEI-VFQ-25, IVI, and Treatment Satisfaction Survey-Intraocular Pressure (TSS-IOP) as having the highest content validity. Another review of PROM instruments by Khadka et al against quality criteria recommended the Modified Glaucoma Quality of Life questionnaire (GAL-9/10), as a higher quality instrument for assessing activity limitation and mobility.¹⁷ These authors subsequently took a systematic approach to identify 737 unique content items for a Glaucoma-specific item bank and refined these into a minimally representative set containing 342 unique items in ten QoL domains.⁴⁹ The authors highlighted that the majority of items were identified de novo from patient focus groups, rather than existing PRO instruments in glaucoma.

A review of trials and clinical studies registered with Clinicaltrials.gov, assessing the efficacy of minimally invasive glaucoma surgical devices, identified that only one of 51 studies included health-related QoL as a secondary outcome measure.⁵⁰ The recently published RCT protocol for the Treatment of Advanced Glaucoma Study claims to be the first RCT to set patient perspectives as the primary outcome measure.⁵¹ Table 1 summarizes RCTs in glaucoma that have included PROMs as primary outcome measures. This table highlights that the impact of PROMs has been relatively limited to date, with focus on anxiety levels between different treatments, but that RCTs are currently underway using PROMs as the key determinant of comparative efficacy.

Medical retina, uveitis, and vitreoretinal disease

A systematic review of retinal disease PROMs by Prem Senthil et al (search date not specified) identified 217 studies, most frequently on AMD (108 studies), diabetic retinopathy (DR) (31 studies), and hereditary retinal dystrophies (29 studies). In total, 110 different PROM instruments were reported, more than half of which were generic (62 studies, most frequently the SF-36, and the Hospital Anxiety and Depression Scale [HADS]), followed by disease-specific (29 studies) and vision-related (19 studies, most frequently the NEI-VFQ and VF-14) instruments.²⁴ Only three instruments had been rescaled and tested using Rasch analysis. They also critically appraised the psychometric performance of the instruments against criteria and identified numerous limitations. The authors reported that most instruments had limited content coverage, typically measuring only one or a few domains of QoL. In another study by Prem Senthil et al, semi-structured,

Table 1 Impact of PROMs in glaucoma RCTs, highlighting only trials in which PROMs were selected as primary outcome measures

Study name	N	Intervention	PRO outcome measures	Impact	Reference
Tube Versus Trabeculectomy Study	202 patients with previous trabeculectomy and/or cataract surgery	Tube shunt (350 mm ² Baerveldt implant) vs trabeculectomy with MMC	NEI-VFQ composite score and minimally important difference	No significant difference at baseline or annual review for 5 years	Kotecha et al ¹⁰
Glaucoma Australia Educational Impact	101 newly diagnosed glaucoma patients	Glaucoma education vs control	Auckland Glaucoma Knowledge Questionnaire	Significant reduction in anxiety in intervention group	Skalicky et al ¹¹
Glaucoma Intensive Treatment Study	242 glaucoma patients	Topical drug monotherapy vs topical triple therapy plus 360 degree laser trabeculoplasty	Eye-tem Bank Glaucoma module	Study Protocol Published	Lamoureux et al ¹⁷ Bengtsson et al ¹²
Treatment of Advanced Glaucoma Study	440 patients presenting with advanced open angle glaucoma	Medical therapy vs augmented trabeculectomy	NEI-VFQ at 24 m. EQ-5D-5L, HUI-3 and Glaucoma Utility Index	Study protocol published	King et al ⁵¹
Shared Care for Stable Glaucoma Patients	233 patients with stable glaucoma	Primary eye care vs specialist outpatient clinic	Patient satisfaction, cost	Comparable patient satisfaction, clinical care and management, but lower cost with PEC	Goh et al ¹³
Laser in Glaucoma and OHT Trial	718 patients with glaucoma or OHT	Selective laser trabeculoplasty vs topical treatment	EQ-5D-5L, Glaucoma Utility Index, GSS, Glaucoma QoL	Study protocol	Gazzard et al ⁸

Abbreviations: EDSQ, Eye Drop Satisfaction Questionnaire; GSS, Glaucoma Symptom Scale; HADS, Hospital Anxiety and Depression Scale; OHT, ocular hypertension; QoL, quality of life; MMC, mitomycin C; PRO, patient-related outcome; NEI-VFQ, National Eye Institute-Vision Function Questionnaire; RCT, randomized controlled clinical trial; GITS, Glaucoma Intensive Treatment Study.

qualitative interview data from 79 patients with hereditary and acquired retinal diseases identified nine QoL domains relevant to both the groups, which were each explored and reported in detail. This paper provides a scientific basis for splitting vs lumping less common retinal diseases to develop a retina-specific PROM.⁵² Further work has formed the basis for a hereditary retinal disease item bank.⁵³

A systematic review of clinical trial registries to identify uveitis trials reported that none out of 104 registered by October 2013 used a PROM as a primary outcome measure.⁵⁴ The Core Outcome Set for Uveitic Macular Oedema (COSUMO) study aims to develop a core outcome set for trials, using systematic review, qualitative research with focus groups, and a Delphi process to reach consensus.⁵⁵ A core outcome set is also being developed by the Outcome Measures in Rheumatology (OMERACT) Vasculitis Working Group for Behcet's disease, which includes the ocular manifestations.⁵⁶ Another core outcome set has been proposed for JIA-associated uveitis.⁵⁷ The Multicenter Uveitis Steroid Treatment study (MUST) investigators reported that their trial, comparing systemic or implanted corticosteroid therapy in 255 patients, was underpowered to explore secondary outcomes of interest including QoL, highlighting the

importance of considering sample size in future comparative effectiveness trials.⁵⁸ Table 2 provides examples of the inclusion of PROMs in uveitis RCTs. The examples illustrate that PROMs are making an important impact in this specialty, where identification of traditional outcome metrics (eg, cells in the vitreous) that correlate meaningfully with the patient-centered experience of disease and its treatment has been more challenging likely due to the reliance on non-disease-specific instruments.

Krezel et al systematically reviewed the frequency and type of PROMs used in RCTs for AMD published between 2010 and 2013.⁵⁹ They reported 177 RCTs including 858 outcomes, of which 38 outcomes were PROMs (4.4%), and these were included in 25 trials (14.1%). The NEI-VFQ was the most frequently used instrument. A minimum set of standardized outcome measures has been defined for macular degeneration and promoted internationally, recommending IVI be used due to its three measurable traits and valid interval scaling.⁶⁰ However, there are currently no PROMs that are clinically validated and acceptable to regulatory agencies for drug development in intermediate AMD, and development of another novel PROM has been proposed.⁶¹ In a study reviewing health state utility values in AMD and

Table 2 Impact of PROMs in uveitis RCTs, illustrating inclusion of PROMs as secondary outcome measures (no RCTs found including PROMs as primary outcome measure)

Study name	N	Intervention	Outcome measures	Impact	Reference
VISUAL-1 and VISUAL-2	217 with active (VISUAL 1), 226 with inactive (VISUAL-2) uveitis	Subcutaneous adalimumab vs placebo	NEI VFQ-25 composite score	Significant improvement in QoL in both trials in the treatment group comparing baseline to final visit	Sheppard et al ¹⁴
SAKURA	347 posterior noninfectious uveitis	Intravitreal sirolimus, 3 doses	NEI-VFQ-25	The composite score and mental health subscore are relevant visual function response measures	Lescrauwaet et al ¹⁵
RCT on antimetabolites for noninfectious uveitis	80 with noninfectious intermediate, posterior, or panuveitis	Oral methotrexate 25 mg weekly or oral mycophenolate mofetil 1 g bd	Indian VFQ and SF-36 at 6 m	Both the treatments improved vision-related QoL (but not health-related) compared to baseline, but both also worsened mental health	Niemeyer et al ¹⁶
HURON	244 with noninfectious intermediate or posterior uveitis	Ozurdex implant vs sham	NEI-VFQ, SF-36, SF-6D, EuroQol-5D	Significant differences were identified for uveitis participants vs general population, except with SF-36 physical component and EQ-5D	Naik et al ¹⁷

Abbreviations: NEI-VFQ, National Eye Institute Visual Function Questionnaire, QoL, quality of life; RCT, randomized controlled clinical trial; SF, Short Form; VFQ, Vision Function Questionnaire; PROM, patient-reported outcome measures.

their use in health care decision-making, Butt et al highlight that generic health-related QoL instruments may lack sensitivity in AMD and that the choice of a utility value should be explicitly critiqued given the existing variability in utility values derived by different studies.⁶²

PROMs have been used to assess diabetic eye disease for many years. In the landmark Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications, intensive diabetes therapy in this cohort modestly improved NEI-VFQ score at 30 years.⁶³ However, reviews in the past decade have highlighted the importance of capturing the patient perspective in diabetic retinopathy more comprehensively,⁶⁴ including the need to measure its social and emotional impact through further PROMs development.⁶⁵ A systematic and comprehensive approach to identify the content for inclusion in a DR item bank yield 1,165 unique items that were winnowed to a minimally representative set of 314 items across nine domains of QoL.^{66,67} Initial evaluation of DR and DME item banks has been undertaken using computer adaptive testing (CAT).⁶⁸

Table 3 provides examples of landmark RCTs in medical retina in which PROMs have had an impact. Many of these trials have used PROMs not only to demonstrate improvement in patient experience in comparison with sham interventions, but also, importantly, to demonstrate non-inferiority of QoL outcomes for interventions differing

dramatically in cost. Such trial data have very important health policy implications. A very recent example is the NICE guideline (2018) to recommend the more cost-effective anti-VEGF therapy bevacizumab as an effective therapy for the treatment of AMD in the UK's NHS. PROM data (mostly using the NEI-VFQ-25 and SF-36) contributed to this policy decision.⁶⁹ No vitreoretinal PROMs were identified by our search.

Cataract

The first vision-related activity limitation instrument for cataract was introduced in 1992, and a review of questionnaires published since 1992 explored the relative merits and psychometric properties of each.¹⁸ Another review compared 16 Rasch-scaled cataract questionnaires before, and 6 months after, cataract surgery.⁷⁰ This study found the Catquest-9SF to be the most responsive to cataract surgery and, being short, was advocated as the best tool for measuring visual functioning outcomes in trials and routine practice. A minimum standardized outcome set has been proposed internationally for cataract surgery, which includes administration of Catquest-9SF pre- and 3-months postoperatively.⁷¹ This instrument has also been included as a secondary outcome measure in a recent RCT protocol (The FACT trial).⁷² A newer PROM, Cat-PROM5, has been tested head-to-head against Catquest-9SF in 822 typical NHS cataract surgery patients and, as an even shorter instrument, advocated as

Table 3 Impact of PROMs in medical retina RCTs, highlighting two RCTs in which the PROM was the primary outcome measure

Study name	N	Intervention	Outcome measures	Impact	Reference
MARINA and ANCHOR	646 (MARINA) and 379 (ANCHOR) patients with wet AMD	Ranibizumab vs sham or photodynamic therapy	NEI-VFQ-25 at baseline, 12 and 24 months	Improvement in outcomes with intervention	Bressler et al ¹¹⁸
SCORE2 Report 5	362 with CRVO or HRVO and macular edema	Intravitreal bevacizumab vs aflibercept	NEI-VFQ-25 composite and subscale scores	Non-inferiority of bevacizumab	Scott et al ¹¹⁹
OZDRY	100 patients with refractory center involving DME	5-monthly fixed dosing vs OCT-guided pro-re-nata regimen of Ozurdex	Retinopathy-Dependent QOL, NEI VFQ-25, (RetTSQ) (primary outcome)	No significant difference at month 12	Ramu et al ¹²⁰
IVAN	610 with active wet AMD	Ranibizumab vs bevacizumab, continuous or discontinuous	EQ-5D	Similar efficacy of drugs in terms of visual acuity. Continuous ranibizumab cost £3.5 million per QALY compared with bevacizumab	Chakravarthy et al ¹²¹
RIDE and RISE	382 RIDE and 377 RISE patients with center-involving DME	Ranibizumab vs sham	NEI-VFQ 25 at baseline, 12, and 24 months	Treatment improved vision-related function significantly more than sham	Bressler et al ¹²²
RESTORE Open-label extension	303 with DME	Ranibizumab 0.5 mg vs laser monotherapy	NEI-VFQ-25 (primary outcome)	Greater gain in ranibizumab group at 12 months, with similar gain in both the groups treated with open-label extension from 12 to 24 months	Mitchell et al ¹²³
BEVORDEX	61 patients with center-involving DME	Ozurdex implant every 16 weeks vs bevacizumab every 4 weeks	IVI	Both the groups had significant improvement in IVI scores	Gillies et al ¹²⁴
MACUGEN	260 with center-involving DME	Pegaptanib sodium vs sham, with focal/grid laser	NEI-VFQ-25, EQ-5D	Clinically and statistically significant differences between groups in composite and subscores, no difference in mean change in EQ-5D utility scores	Loftus et al ¹²⁵
BRAVO and CRUISE	397 with branch and 392 with central retinal vein occlusion and macular edema	Ranibizumab vs sham	NEI-VFQ-25	Treatment results in significant mean improvement in composite score compared to sham from month 1	Varma et al ¹²⁶

Abbreviations: AMD, age-related macular degeneration; BRVO, branch retinal vein occlusion; CRVO, central retinal vein occlusion; DME, diabetic macular edema; IVI, Impact of Vision Impairment; NEI-VFQ, National Eye Institute Visual Function Questionnaire; RetTSQ, Retinopathy Treatment Satisfaction Questionnaire; PROM, patient-reported outcome measures; RCT, randomized controlled clinical trial.

being preferred by patients, and better suited for use in high volume routine surgical practice.⁷³

Amblyopia, strabismus, and pediatric ophthalmology

Kumaran et al conducted a systematic review of PROMs in amblyopia and strabismus published up to July 2016.⁷⁴ This identified 71 PROMs of which 32 were amblyopia and/or strabismus specific, but only four of these had been subjected to psychometric tests, and only the adult strabismus questionnaire (AS-20) demonstrated good measurement properties. The authors concluded that all instruments had gaps in

their content and failed to assess QoL comprehensively and proposed the development of an item bank to address this. Another review recommended the Modified AS-20 instrument, which measures self-perception, interaction, reading, and general function, as one of the strongest of the existing instruments.^{17,75} Incorporation of AS-20 QoL questionnaires into pre- and postoperative clinical practice has been proposed, on account of finding that many apparent surgical failures report subjective improvement.⁷⁶

Tadić et al conducted a systematic review of PROMs for ophthalmic disorders in children, and identified 17 instruments, of which 11 were condition-specific and six were

for children and young people with visual impairment. The authors recommended the need for the development of new instruments.⁷⁷ Tadić and Rahi further elaborated on issues particular to the development of PROMs for use in children.⁷⁸ These include conflation between theoretically distinct vision-related constructs and outcomes, the importance of developmentally appropriate approaches to design and application, the feasibility of administering self-report standard questionnaire formats to visually impaired children, ethical issues, and statistical issues. More recently, Hatt et al have identified a comprehensive list of child- and parent-derived items from 180 children and 328 parents, which they grouped into 614 unique items identified by children in 36 subgroups, and 589 items identified by parents in 61 bins. The authors report that they intend to develop a formal set of pediatric PROMs from this pool.⁷⁹

Cornea and external disease

A systematic review of PROMs for surgically amenable epiphora identified that 69% of 227 studies included a PROM as the primary outcome measure, although in 48% the PROM was a single-item symptom score.⁸⁰ The authors critically appraised each PROM and concluded that they lacked adequate content validity. In Primary Sjogren's Syndrome, qualitative work and PROM development have been done to identify 484 items covering 86 concepts in 3 dimensions impacting QoL.^{81,82} In the TEARS trial (Tolerability and Efficacy of Rituximab in Primary Sjogren's Syndrome), SF-36 scores were found to be strongly associated with patient-reported symptoms.⁸³ A review of PROMs for use in RCTs of dry eye identified 18 instruments, some of which were generic, many of which focused on symptoms, and concluded that very few available PROMs satisfy FDA guidance on the requirements of a suitable PROM to be acceptable as support for a label claim in support of a drug or medical device.⁸⁴ The first RCT (n=16) to demonstrate the beneficial effects of autologous serum in patients with severe ocular surface disease used a daily subjective symptom scale, the Rasch-scored Faces scale, to demonstrate a significant effect of the intervention.⁸⁵

While limited PROMs have been designed for use in corneal diseases, Catquest-9SF has been validated for use in patients who have had corneal transplant surgery.⁸⁶

Refractive error

Kandel et al conducted a systematic review for studies using PROMs to assess refractive surgery outcomes.⁸⁷ They identified 27 instruments, 12 of which were specific

to refractive error. The authors reported that while the NEI Refractive Error Quality of Life instrument (NEI-RQL) was the most frequently used, it did not provide valid measurement, whereas a number of other instruments, including the Quality of Vision, Near Activity Visual Questionnaire, and Quality of Life questionnaire (QIRC) had been constructed using Rasch analysis and were suited to measurement of visual symptoms, activity limitations, and QoL, respectively. They subsequently developed a pool of refractive error items from patient groups in Australia (n=337 items) and Nepal (n=308 items), spanning 12 QoL domains and are working to develop a CAT system suitable for use in both high- and low-income country settings.⁸⁸ An RCT using PROMs as the primary outcome measure to compare ready-made spectacles and custom-made spectacles for the correction of refractive error in adults in India found that both result in comparable patient satisfaction and large gains in visual function and QoL, with the custom-made spectacles achieving a small but statistically significant higher QoL outcome.⁸⁹

Oculoplastics

A systematic review of PROMs for eyelid, orbit, and lacrimal disorders, conducted in 2013, identified ten generic and 32 disease-specific instruments and assessed their content domains and psychometric quality.⁹⁰ The SF-36 and NEI-VFQ-25 were the most frequently used generic instruments, and thyroid eye disease was the most studied condition. Of the 32 disease-specific instruments, 13 were developed for eyelid-related disease, ten for orbital disease, and nine for lacrimal disease. Physical function and self-image were the most frequently studied domains of QoL. The authors reported that the majority of instruments had very limited psychometric development and poorly defined content domains and concluded that efforts to develop PROMs in oculofacial surgery had been sparse, fragmented, and generally rudimentary, making assimilation into daily clinical practice challenging. More recently, the FACE-Q Eye Module has been developed for use in cosmetic eye treatments and contains four scales measuring appearance of the eyes, upper and lower eyelids and eyelashes,⁹¹ and a module for children and young adults with diverse conditions causing facial appearance differences has also been developed.⁹² While there have been a few further clinical studies reported since 2013, no RCTs utilizing PROMs as key outcome measures were identified by our search.

Neuro-ophthalmology

We identified one systematic review of PROMs for use in patients with vision impairment following stroke,

which identified 34 vision-specific PROMs, and critically appraised the quality of the identified instruments.⁹³ The authors highlighted four high-quality instruments, including NEI-VFQ, Activity Inventory (AI), Daily Living Tasks Dependent on Vision (DLTV), and the Veterans Affairs Low Visual Function Questionnaire (VA LV VFQ), but cautioned that these had each only been assessed in a limited number of patients.

There were no other systematic reviews of PROM instruments for neuro-ophthalmic conditions, and only scattered examples of PROMs that have been developed for specific conditions. A neuro-ophthalmic module was developed for the NEI-VFQ.⁹⁴ This was assessed for content and quality by Ramey et al and considered to perform reasonably well by classical test criteria.⁹⁰ Generic instruments including NEI-VFQ and SF-36 have been used in an RCT in idiopathic intracranial hypertension (IIH) patients,⁹⁵ and in a study of neurofibromatosis (NF) type 2.⁹⁶ The Children's Visual Function Questionnaire has been proposed as a secondary endpoint for clinical trials in children with NF1-associated optic pathway gliomas.⁷⁶ Disease-specific instruments have been developed for patients with neuromyelitis optica spectrum disorders.^{97,98} The first use of a PROM information system utilizing CAT in patients with neurofibromatosis has also been reported.⁹⁹

Low vision

A literature review of RCTs on low vision rehabilitation identified 15 trials, utilizing nine PROMs, and one hybrid PROM and performance-based outcome measure, the Melbourne Low-Vision ADL Index.¹⁰⁰ The other instruments included the AI, Canadian Occupational Performance Measure, Functional Assessment Questionnaire, Groningen Activity Restriction Scale, IVI, Katz' Index of Activities of Daily Living, Low Vision QOL, NEI-VFQ, and the VA LV VFQ. Most of these instruments (seven out of ten) have utilized Rasch or IRT modeling, have been validated for use in low-vision populations, and include items in a number of different domains. The Veterans Affairs Low Vision Intervention Trials (LOVIT I and II) used LV VFQ-48 as the primary outcome measure. Significant benefit on reading ability at 4 months was demonstrated for low-vision rehabilitation (n=126 patients with low vision from macular disease).⁹ The LOVIT II trial randomized 323 patients to receive low vision devices with or without rehabilitation therapy and found that the latter group improved more in all visual function domains except mobility.¹⁰

The Impact of PROMs in routine clinical practice

Our search identified very few examples of the use of PROMs in routine clinical practice. Clinicians may be more likely to report such progress and real-life experience in the gray literature and unpublished sources, and we recognize that this is a limitation of this narrative, rather than systematic, review of the published literature indexed in PubMed.

The Swedish National Cataract Outcome Study (1995–1999) prospectively administered Catquest-9SF, before and after surgery, to 8,595 patient eyes and demonstrated greater impact on satisfaction, and surgical benefit to vision, of second-eye surgery.¹⁰¹ A similar finding was reported when Catquest-9SF was administered to 870 patients in five Dutch hospitals.¹⁰² Data for the Swedish National Outcome Study (2008–2011), on 9,707 patient eyes before and after surgery, further revealed large variation in PROMs, influenced by factors including the degree of anisometropia, indication for surgery, and postoperative problems.¹⁰³ These examples highlight the value of implementing PROMs in real-world clinical practice and illustrate that they may reveal patient preferences unexpected by clinicians and policy makers.

Hee et al recently explored the feasibility of implementing glaucoma PROMs in daily clinical practice in Singapore.¹⁰⁴ They reported that while the majority of health care professionals and patients felt that the four glaucoma PROMs selected for use in this study were relevant to them, there were multiple barriers to their routine use. These included the need for brevity, yet the desire for a more comprehensive instrument able to capture patient concerns more fully, and the challenge for patients with vision impairment to self-administer the instrument on paper. Furthermore, responders highlighted the desire for inclusion of measurement of financial impact. The authors highlighted that participation in completing PROMs was much lower among patients from lower socioeconomic and education backgrounds, who tend to be those most severely affected by eye disease.

A single PROM for all ophthalmic situations?

The previous section outlines considerable achievements in recent years to develop PROMs for the most prevalent eye diseases globally. In some ophthalmic subspecialties, such as low vision, medical retina, and glaucoma, PROMs are frequently included as secondary, and increasingly as primary, outcome measures in clinical trials, and are being

explored for integration into routine clinical practice. Other subspecialties are still at an earlier stage of developing and assessing PROMs that target the impact of the key diseases and treatments. There is a dearth of PROMs for rarer diseases, especially in neuro-ophthalmology. The issues particular to PROM research in rare diseases have been explored by Slade et al.¹⁰⁵ A key challenge is the time it takes to develop a valid and reliable PROM. High-quality PROM development requires extensive qualitative work with patients and focus groups, pilot studies in which a long set of potential items are administered to patients, psychometric data analysis and winnowing of redundant items, before validation of the final instrument in clinical practice and trials, and PROMs are not necessarily directly applicable in differing cultural contexts.

One solution is to develop a very large bank of items and to validate subsets of questions from this bank in many different diseases and patient populations.¹⁰⁶ Methods to develop one such “Eye-tem bank” to measure vision and eye disease-related QoL have been outlined.¹⁷ This bank is being developed across 13 disease groups, namely AMD, cataract, glaucoma, DR, retinal detachment, other vitreo-retinal, cornea, refractive error, uveitis, other inflammation, amblyopia and strabismus, lacrimal and ocular surface, and neuro-ophthalmology.¹⁷ While CAT can be used to target items to the dynamic responses of each individual responder, further work will be needed to ascertain the time response burden and acceptability of such comprehensive tools in both research and clinical practice settings.

Another approach is to routinely include at least one generic PROM such as the EQ-5D alongside the wide range of vision and eye disease-related PROMs currently being used.

Future research priorities

Guidelines for the inclusion of patient-reported outcomes in clinical trial protocols and reporting guidelines have been developed: the SPIRIT-PRO and CONSORT-PRO Extensions, respectively.¹⁰⁷ Standardization of vision-related PROMs is needed, and progress toward achieving this has been made in other fields. For example, the SISAQOL consortium, “Setting International Standards in analyzing PRO and QOL endpoints for cancer clinical trials,” is developing standardized approaches for the analysis of PROM data in cancer.¹⁰⁸

A systematic review has identified methodological frameworks to measure the health care impact of research.¹⁰⁹ Beyond measuring PROMs more precisely, reliably, and comprehensively in the future, further research is needed to

better understand and demonstrate the impact of measuring PROMs in ophthalmic research and clinical practice.

Conclusion

There is much improved awareness of PROMs among both researchers and clinicians over recent decades, but much work needs to be done to standardize the outcomes and the measures. PROMs provide a unique and exciting opportunity to capture what matters to patients and to inform understanding of all stakeholders. Through influencing the decisions of clinicians, regulators, and policy makers involved in the care of patients with ophthalmic diseases, PROMs have the potential to transform medical care.

Author contributions

All authors contributed to data analysis. TB drafted the manuscript in consultation with all co-authors, who provided critical revision. All authors gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Disclosure

Tasane Braithwaite has received funding support from the charity Olivia’s Vision and is a Cochrane Eyes and Vision Contact Editor. Part of this work contributed to her thesis (University of Oxford, Doctor of Medicine), submitted to the University of Oxford for the Doctor of Medicine degree. Melanie Calvert is funded by the NIHR Birmingham Biomedical Research Centre and the NIHR Surgical Reconstruction and Microbiology Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham. Konrad Pesudovs has over 25 years of experience in vision and eye disease-related PROMs, has created numerous instruments including the Visual Disability Assessment, the Quality of Life Impact of Refractive Correction, the Contact Lens Impact on Quality of Life, the Quality of Vision Questionnaire, and the Eye-tem Bank. He has been involved in the revalidation using Rasch analysis of numerous instruments mentioned in this review including the Catquest 9SF, the IVI, and the GAL9. He is a proponent of quality assessment of outcomes using PROMs. Inevitably some of his works are mentioned herein. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. Melanie Calvert reports receipt of personal fees from Astellas, Takeda, and Merck and grants from Macmillan Cancer Support and the NIHR and is a coinvestigator within Health Data Research UK Midlands. The authors report no other conflicts of interest in this work.

References

- Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Measuring outcomes of cataract surgery using the Visual Function Index-14. *J Cataract Refract Surg*. 2010;36(7):1181–1188.
- Pesudovs K. Patient-centred measurement in ophthalmology—a paradigm shift. *BMC Ophthalmol*. 2006;6:25.
- Pesudovs K, Gothwal VK, Wright T, Lamoureux EL. Remediating serious flaws in the National Eye Institute Visual Function Questionnaire. *J Cataract Refract Surg*. 2010;36(5):718–732.
- Petrillo J, Bressler NM, Lamoureux E, Ferreira A, Cano S. Development of a new Rasch-based scoring algorithm for the National Eye Institute Visual Functioning Questionnaire to improve its interpretability. *Health Qual Life Outcomes*. 2017;15(1):157.
- Mcalinden C, Skiadaresi E, Moore J, Pesudovs K. Subscale assessment of the NEI-RQL-42 questionnaire with Rasch analysis. *Invest Ophthalmol Vis Sci*. 2011;52(8):5685–5694.
- Dean S, Mathers JM, Calvert M, et al. “The patient is speaking”: discovering the patient voice in ophthalmology. *Br J Ophthalmol*. 2017;101(6):700–708.
- Lamoureux EL, McIntosh R, Constantinou M, et al. Comparing the effectiveness of selective laser trabeculoplasty with topical medication as initial treatment (the Glaucoma Initial Treatment Study): study protocol for a randomised controlled trial. *Trials*. 2015;16:406.
- Gazzard G, Konstantakopoulou E, Garway-Heath D, et al. Laser in Glaucoma and Ocular Hypertension (LiGHT) trial. A multicentre, randomised controlled trial: design and methodology. *Br J Ophthalmol*. 2018;102(5):593–598.
- Stelmack JA, Tang XC, Reda DJ, et al. Outcomes of the Veterans Affairs Low Vision Intervention Trial (LOVIT). *Arch Ophthalmol*. 2008;126(5):608–617.
- Stelmack JA, Tang XC, Wei Y. Outcomes of the Veterans Affairs Low Vision Intervention Trial II (LOVIT II): a randomized clinical trial. *JAMA Ophthalmol*. Epub 2016 Dec 15.
- Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in health care settings. *BMJ*. 2010;340:c186.
- Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. *BMJ*. 2015;350:g7818.
- Massof RW. The measurement of vision disability. *Optom Vis Sci*. 2002;79(8):516–552.
- Michelotti M, de Korne DF, Weizer JS, et al. Mapping standard ophthalmic outcome sets to metrics currently reported in eight eye hospitals. *BMC Ophthalmol*. 2017;17(1):269.
- Hart PM, Chakravarthy U, Stevenson MR. Questionnaire-based survey on the importance of quality of life measures in ophthalmic practice. *Eye*. 1998;12(Pt 1):124–126.
- Massof RW, Rubin GS. Visual function assessment questionnaires. *Surv Ophthalmol*. 2001;45(6):531–548.
- Khadka J, Fenwick E, Lamoureux E, Pesudovs K. Methods to develop the eye-tem bank to measure ophthalmic quality of life. *Optom Vis Sci*. 2016;93(12):1485–1494.
- Lundström M, Pesudovs K. Questionnaires for measuring cataract surgery outcomes. *J Cataract Refract Surg*. 2011;37(5):945–959.
- Petrillo J, Cano SJ, McLeod LD, Coon CD. Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. *Value Health*. 2015;18(1):25–34.
- Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42.
- de Boer MR, Moll AC, de Vet HC, Terwee CB, Völker-Dieben HJ, van Rens GH. Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt*. 2004;24(4):257–273.
- Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci*. 2007;84(8):663–674.
- US Department of Health and Human Services FaDAF. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labelling Claims; 2009. Available from: <http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf>. Accessed September 25, 2018.
- Prem Senthil M, Khadka J, Pesudovs K. Assessment of patient-reported outcomes in retinal diseases: a systematic review. *Surv Ophthalmol*. 2017;62(4):546–582.
- EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199–208.
- Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736.
- Yang Y, Rowen D, Brazier J, Tsuchiya A, Young T, Longworth L. An exploratory study to test the impact on three “bolt-on” items to the EQ-5D. *Value Health*. 2015;18(1):52–60.
- Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108.
- Bernert S, Fernández A, Haro JM, et al. Comparison of different valuation methods for population health status measured by the EQ-5D in three European countries. *Value Health*. 2009;12(5):750–758.
- Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Econ*. 2018;27(1):7–22.
- Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res*. 2005;14(6):1523–1532.
- Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. *J Clin Epidemiol*. 1998;51(11):1115–1128.
- Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473–483.
- Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ*. 2002;21(2):271–292.
- Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ*. 1986;5(1):1–30.
- Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes*. 2003;1:54.
- Furlong WFD, Torrance GW, Goldsmith C, et al. *Multiplicative Multi-Attribute Utility Function for the Health Utilities Index Mark 3 (HUI3) System: A Technical Report*. Hamilton, Ontario, Canada: McMaster University;1998.
- Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ*. 2004;13(9):873–884.
- Tosh JC, Longworth LJ, George E. Utility values in National Institute for Health and Clinical Excellence (NICE) Technology Appraisals. *Value Health*. 2011;14(1):102–109.
- Bilbao A, Quintana JM, Escobar A, et al. Responsiveness and clinically important differences for the VF-14 index, SF-36, and visual acuity in patients undergoing cataract surgery. *Ophthalmology*. 2009;116(3):418–424.
- Ijzerman MJ, Koffijberg H, Fenwick E, Krahn M. Emerging use of early health technology assessment in medical product development: a scoping review of the literature. *Pharmacoeconomics*. 2017;35(7):727–740.
- Bass EB, Marsh MJ, Mangione CM, et al. Patients’ perceptions of the value of current vision: assessment of preference values among patients with subfoveal choroidal neovascularization—The Submacular Surgery Trials Vision Preference Value Scale: SST Report No. 6. *Arch Ophthalmol*. 2004;122(12):1856–1867.
- Kymes SM, Lee BS. Preference-based quality of life measures in people with visual impairment. *Optom Vis Sci*. 2007;84(8):809–816.
- Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:22.

45. Khadka J, McAlinden C, Pesudovs K. Quality assessment of ophthalmic questionnaires: review and recommendations. *Optom Vis Sci*. 2013;90(8):720–744.
46. Fenwick EK, Man RE, Rees G, Keeffe J, Wong TY, Lamoureux EL. Reducing respondent burden: validation of the Brief Impact of Vision Impairment questionnaire. *Qual Life Res*. 2017;26(2):479–488.
47. Vandenbroeck S, de Geest S, Zeyen T, Stalmans I, Dobbels F. Patient-reported outcomes (PRO's) in glaucoma: a systematic review. *Eye*. 2011;25(5):555–577.
48. Che Hamzah J, Burr JM, Ramsay CR, Azuara-Blanco A, Prior M. Choosing appropriate patient-reported outcomes instrument for glaucoma research: a systematic review of vision instruments. *Qual Life Res*. 2011;20(7):1141–1158.
49. Khadka J, McAlinden C, Craig JE, Fenwick EK, Lamoureux EL, Pesudovs K. Identifying content for the glaucoma-specific item bank to measure quality-of-life parameters. *J Glaucoma*. 2015;24(1):12–19.
50. Le JT, Viswanathan S, Tarver ME, Eydelman M, Li T. Assessment of the incorporation of patient-centric outcomes in studies of minimally invasive glaucoma surgical devices. *JAMA Ophthalmol*. 2016;134(9):1054–1056.
51. King AJ, Fernie G, Azuara-Blanco A, et al. Treatment of advanced glaucoma study: a multicentre randomised controlled trial comparing primary medical treatment with primary trabeculectomy for people with newly diagnosed advanced glaucoma-study protocol. *Br J Ophthalmol*. 2018;102(7):922–928.
52. Prem Senthil M, Khadka J, Gilhotra JS, Simon S, Pesudovs K. Exploring the quality of life issues in people with retinal diseases: a qualitative study. *J Patient Rep Outcomes*. 2017;1(1):15.
53. Prem Senthil M, Khadka J, de Roach J, et al. Developing an item bank to measure the coping strategies of people with hereditary retinal diseases. *Graefes Arch Clin Exp Ophthalmol*. 2018;256(7):1291–1298.
54. Denniston AK, Holland GN, Kidess A, et al. Heterogeneity of primary outcome measures used in clinical trials of treatments for intermediate, posterior, and panuveitis. *Orphanet J Rare Dis*. 2015;10:97.
55. Tallouzi MO, Mathers JM, Moore DJ, et al. COSUMO: study protocol for the development of a core outcome set for efficacy and effectiveness trials in posterior segment-involving uveitis. *Trials*. 2017;18(1):576.
56. Hatemi G, Meara A, Ozguler Y, et al. Developing a Core Set of Outcome Measures for Behçet Disease: Report from OMERACT 2016. *J Rheumatol*. 2017;44(11):1750–1753.
57. Heiligenhaus A, Foeldvari I, Edelsten C, et al. Proposed outcome measures for prospective clinical trials in juvenile idiopathic arthritis-associated uveitis: a consensus effort from the multinational interdisciplinary working group for uveitis in childhood. *Arthritis Care Res*. 2012;64(9):1365–1372.
58. Holbrook JT, Kempen JH, Prusakowski NA, Altaweel MM, Jabs DA, Multicenter Uveitis Steroid Treatment (MUST) Trial Research Group. Challenges in the design and implementation of the Multicenter Uveitis Steroid Treatment (MUST) Trial-lessons for comparative effectiveness trials. *Clin Trials*. 2011;8(6):736–743.
59. Krezel AK, Hogg RE, Azuara-Blanco A. Patient-reported outcomes in randomised controlled trials on age-related macular degeneration. *Br J Ophthalmol*. 2015;99(11):1560–1564.
60. Rodrigues IA, Sprinkhuizen SM, Barthelmes D, et al. Defining a minimum set of standardized patient-centered outcome measures for macular degeneration. *Am J Ophthalmol*. 2016;168:1–12.
61. Finger RP, Schmitz-Valckenberg S, Schmid M, et al. MACUSTAR: Development and clinical validation of functional, structural, and patient-reported endpoints in intermediate age-related macular degeneration. *Ophthalmologica*. Epub 2018 Aug 28.
62. Butt T, Tufail A, Rubin G. Health state utility values for age-related macular degeneration: review and advice. *Appl Health Econ Health Policy*. 2017;15(1):23–32.
63. Writing Team for the DCCT/EDIC Research Group, Gubitosi-Klug RA, Sun W, et al. Effects of prior intensive insulin therapy and risk factors on patient-reported visual function outcomes in the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Cohort. *JAMA Ophthalmol*. 2016;134(2):137–145.
64. Fenwick EK, Pesudovs K, Rees G, et al. The impact of diabetic retinopathy: understanding the patient's perspective. *Br J Ophthalmol*. 2011;95(6):774–782.
65. Fenwick E, Rees G, Pesudovs K, et al. Social and emotional impact of diabetic retinopathy: a review. *Clin Exp Ophthalmol*. 2012;40(1):27–38.
66. Fenwick EK, Pesudovs K, Khadka J, Rees G, Wong TY, Lamoureux EL. Evaluation of item candidates for a diabetic retinopathy quality of life item bank. *Qual Life Res*. 2013;22(7):1851–1858.
67. Fenwick EK, Pesudovs K, Khadka J, et al. The impact of diabetic retinopathy on quality of life: qualitative findings from an item bank development project. *Qual Life Res*. 2012;21(10):1771–1782.
68. Fenwick EK, Khadka J, Pesudovs K, Rees G, Wong TY, Lamoureux EL. Diabetic retinopathy and macular edema quality-of-life item banks: development and initial evaluation using computerized adaptive testing. *Invest Ophthalmol Vis Sci*. 2017;58(14):6379–6387.
69. Macular Degeneration Appendix H: GRADE tables and meta-analysis results; 2018. Available from: <http://www.nice.org.uk/guidance/ng82/evidence/appendix-h-grade-tables-and-metaanalysis-results-pdf-4723229205>.
70. McAlinden C, Gothwal VK, Khadka J, Wright TA, Lamoureux EL, Pesudovs K. A head-to-head comparison of 16 cataract surgery outcome questionnaires. *Ophthalmology*. 2011;118(12):2374–2381.
71. Mahmud I, Kelley T, Stowell C, et al. A proposed minimum standard set of outcome measures for cataract surgery. *JAMA Ophthalmol*. 2015;133(11):1247–1252.
72. Day AC, Burr JM, Bunce C, et al. Randomised, single-masked non-inferiority trial of femtosecond laser-assisted versus manual phacoemulsification cataract surgery for adults with visually significant cataract: the FACT trial protocol. *BMJ Open*. 2015;5(11):e010381.
73. Sparrow JM, Grzeda MT, Frost NA, et al. Cataract surgery patient-reported outcome measures: a head-to-head comparison of the psychometric performance and patient acceptability of the Cat-PROM5 and Catquest-9SF self-report questionnaires. *Eye*. 2018;32(4):788–795.
74. Kumaran SE, Khadka J, Baker R, Pesudovs K. Patient-reported outcome measures in amblyopia and strabismus: a systematic review. *Clin Exp Optom*. 2017.
75. Leske DA, Hatt SR, Liebermann L, Holmes JM. Evaluation of the adult Strabismus-20 (AS-20) questionnaire using Rasch analysis. *Invest Ophthalmol Vis Sci*. 2012;53(6):2630–2639.
76. Hatt SR, Leske DA, Liebermann L, Holmes JM. Incorporating health-related quality of life into the assessment of outcome following strabismus surgery. *Am J Ophthalmol*. 2016;164:1–5.
77. Tadić V, Hogan A, Sobti N, Knowles RL, Rahi JS. Patient-reported outcome measures (PROMs) in paediatric ophthalmology: a systematic review. *Br J Ophthalmol*. 2013;97(11):1369–1381.
78. Tadić V, Rahi JS. One size doesn't fit all: time to revisit patient-reported outcome measures (PROMs) in paediatric ophthalmology? *Eye*. 2017;31(4):511–518.
79. Hatt SR, Leske DA, Castaneda YS, et al. Patient-derived questionnaire items for patient-reported outcome measures in pediatric eye conditions. *J AAPOS*. 2018;22(6):445–448.e22.
80. Schulz CB, Kennedy A, Rogers S. A Systematic Review of Patient-Reported Outcomes for Surgically Amenable Epiphora. *Ophthalmic Plast Reconstr Surg*. 2018;34(3):193–200.
81. Lackner A, Ficjan A, Stradner MH, et al. It's more than dryness and fatigue: the patient perspective on health-related quality of life in primary Sjögren's syndrome – A qualitative study. *PLoS One*. 2017;12(2):e0172056.

82. Lackner A, Stradner MH, Hermann J, et al. Assessing health-related quality of life in primary Sjögren's syndrome-The PSS-QoL. *Semin Arthritis Rheum*. 2018;48(1):105–110.
83. Cornec D, Devauchelle-Pensec V, Mariette X, et al. Severe health-related quality of life impairment in active primary Sjögren's syndrome and patient-reported outcomes: data from a large therapeutic trial. *Arthritis Care Res*. 2017;69(4):528–535.
84. Guillemin I, Begley C, Chalmers R, Baudouin C, Arnould B. Appraisal of patient-reported outcome instruments available for randomized clinical trials in dry eye: revisiting the standards. *Ocul Surf*. 2012;10(2):84–99.
85. Noble BA, Loh RS, MacLennan S, et al. Comparison of autologous serum eye drops with conventional therapy in a randomised controlled crossover trial for ocular surface disease. *Br J Ophthalmol*. 2004;88(5):647–652.
86. Claesson M, Armitage WJ, Byström B, et al. Validation of Catquest-9SF-A visual disability instrument to evaluate patient function after corneal transplantation. *Cornea*. 2017;36(9):1083–1088.
87. Kandel H, Khadka J, Lundström M, Goggin M, Pesudovs K. Questionnaires for measuring refractive surgery outcomes. *J Refract Surg*. 2017;33(6):416–424.
88. Kandel H, Khadka J, Fenwick EK, et al. Constructing item banks for measuring quality of life in refractive error. *Optom Vis Sci*. 2018;95(7):575–587.
89. Brady CJ, Villanti AC, Gandhi M, Friedman DS, Keay L. Visual function after correction of distance refractive error with ready-made and custom spectacles: a randomized clinical trial. *Ophthalmology*. 2012;119(10):2014–2020.
90. Ramey NA, Butt Z, Burkat CN, Rose JG, Lucarelli MJ. Patient-reported outcomes: comprehensive analysis for the oculofacial clinician. *Ophthalmic Plast Reconstr Surg*. 2014;30(4):279–289.
91. Klassen AF, Cano SJ, Grotting JC, et al. FACE-Q Eye module for measuring patient-reported outcomes following cosmetic eye treatments. *JAMA Facial Plast Surg*. 2017;19(1):7–14.
92. Longmire NM, Wong Riff KKY, O'Hara JL, et al. Development of a new module of the FACE-Q for children and young adults with diverse conditions associated with visible and/or functional facial differences. *Facial Plast Surg*. 2017;33(5):499–508.
93. Hepworth LR, Rowe FJ, Harper R, Jarvis K, Shipman T, Rodgers H. Patient reported outcome measures for visual impairment after stroke: a systematic review. *Health Qual Life Outcomes*. 2015;13:146.
94. Raphael BA, Galetta KM, Jacobs DA, et al. Validation and test characteristics of a 10-item neuro-ophthalmic supplement to the NEI-VFQ-25. *Am J Ophthalmol*. 2006;142(6):1026–1035.
95. Piper RJ, Kalyvas AV, Young AM, Hughes MA, Jamjoom AA, Fouyas IP. Interventions for idiopathic intracranial hypertension. *Cochrane Database Syst Rev*. 2015;8:CD003434.
96. Merker VL, Bergner AL, Vranceanu AM, Muzikansky A, Slaterry W, Plotkin SR. Health-related quality of life of individuals with neurofibromatosis type 2: results from the NF2 natural history study. *Otol Neurotol*. 2016;37(5):574–579.
97. Moore P, Jackson C, Mutch K, et al. Patient-reported outcome measure for neuromyelitis optica: pretesting of preliminary instrument and protocol for further development in accordance with international guidelines. *BMJ Open*. 2016;6(9):e011142.
98. Methley AM, Mutch K, Moore P, Jacob A. Development of a patient-centred conceptual framework of health-related quality of life in neuromyelitis optica: a qualitative study. *Health Expect*. 2017;20(1):47–58.
99. Talaei-Khoei M, Riklin E, Merker VL, et al. First use of patient reported outcomes measurement information system (PROMIS) measures in adults with neurofibromatosis. *J Neurooncol*. 2017;131(2):413–419.
100. Ehrlich JR, Spaeth GL, Carlotto NE, Lee PP. Patient-centered outcome measures to assess functioning in randomized controlled trials of low-vision rehabilitation: a review. *Patient*. 2017;10(1):39–49.
101. Lundström M, Stenevi U, Thorburn W. Quality of life after first- and second-eye cataract surgery: five-year data collected by the Swedish national cataract register. *J Cataract Refract Surg*. 2001;27(10):1553–1559.
102. Stolk-Vos AC, Visser MS, Klijn S, et al. Effects of clinical parameters on patient-reported outcome in cataract patients: a multicentre study. *Acta Ophthalmol*. 2018;96(6):586–591.
103. Mollazadegan K, Lundström M. A study of the correlation between patient-reported outcomes and clinical outcomes after cataract surgery in ophthalmic clinics. *Acta Ophthalmol*. 2015;93(3):293–298.
104. Hee OK, Thng ZX, Zhu HY, Lamoureux EL. Usage of glaucoma-specific patient-reported outcome measures (PROMs) in the Singapore context: a qualitative scoping exercise. *BMC Ophthalmol*. 2018;18(1):197.
105. Slade A, Isa F, Kyte D, et al. Patient reported outcome measures in rare diseases: a narrative review. *Orphanet J Rare Dis*. 2018;13(1):61.
106. Pesudovs K. Item banking: a generational change in patient-reported outcome measurement. *Optom Vis Sci*. 2010;87(4):285–293.
107. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols. *JAMA*. 2018;319(5):483–494.
108. Bottomley A, Pe M, Sloan J, et al. Moving forward toward standardizing analysis of quality of life data in randomized cancer clinical trials. *Clin Trials*. 2018;15(6):624–630.
109. Cruz Rivera S, Kyte DG, Aiyegbusi OL, Keeley TJ, Calvert MJ. Assessing the impact of health care research: a systematic review of methodological frameworks. *PLoS Med*. 2017;14(8):e1002370.
110. Kotecha A, Feuer WJ, Barton K, Gedde SJ; Tube Versus Trabeculectomy Study Group. Quality of life in the tube versus Trabeculectomy Study. *Am J Ophthalmol*. 2017;176:228–235.
111. Skaliky SE, D'Mellow G, House P, Fenwick E; Glaucoma Australia Educational Impact Study Contributors. Glaucoma Australia educational impact study: a randomized short-term clinical trial evaluating the association between glaucoma education and patient knowledge, anxiety and treatment satisfaction. *Clin Exp Ophthalmol*. 2018;46(3):222–231.
112. Bengtsson B, Heijl A, Johannesson G, Andersson-Geimer S, Asperberg J, Lindén C. The Glaucoma Intensive Treatment Study (GITS), a randomized clinical trial: design, methodology and baseline data. *Acta Ophthalmologica*. 2018;96(6):557–566.
113. Goh D, de Korne DF, Ho H, et al. Shared care for stable glaucoma patients: economic benefits and patient-centered outcomes of a feasibility trial. *J Glaucoma*. 2018;27(2):170–175.
114. Sheppard J, Joshi A, Betts KA, et al. Effect of Adalimumab on visual functioning in patients with noninfectious intermediate uveitis, posterior uveitis, and panuveitis in the VISUAL-1 and VISUAL-2 trials. *JAMA Ophthalmology*. 2017;135(6):511–518.
115. Lescrauwaet B, Miserocchi E, Thureau S, et al. Association between visual function response and reduction of inflammation in noninfectious uveitis of the posterior segment. *Invest Ophthalmol Vis Sci*. 2017;58(9):3555–3562.
116. Niemeyer KM, Gonzales JA, Rathinam SR, et al. Quality-of-life outcomes from a randomized clinical trial comparing antimetabolites for intermediate, posterior, and panuveitis. *Am J Ophthalmol*. 2017;179:10–17.
117. Naik RK, Rentz AM, Foster CS, et al. Normative comparison of patient-reported outcomes in patients with noninfectious uveitis. *JAMA Ophthalmology*. 2013;131(2):219–225.
118. Bressler NM, Chang TS, Suñer JJ, et al. Vision-related function after ranibizumab treatment by better- or worse-seeing eye: clinical trial results from MARINA and ANCHOR. *Ophthalmology*. 2010;117(4):747–756.
119. Scott IU, Figueroa MJ, Oden NL, et al. SCORE2 Report 5: vision-related function in patients with macular edema secondary to central retinal or hemiretinal vein occlusion. *Am J Ophthalmol*. 2017;184:147–156.
120. Ramu J, Chatziralli I, Yang Y, et al. Health-related quality of life, visual function and treatment satisfaction following intravitreal dexamethasone implant for diabetic macular edema. *Patient Prefer Adherence*. 2017;11:579–586.

121. Chakravarthy U, Harding SP, Rogers CA, et al. A randomised controlled trial to assess the clinical effectiveness and cost-effectiveness of alternative treatments to Inhibit VEGF in age-related choroidal neovascularisation (IVAN). *Health Technol Assess*. 2015;19(78): 1–298.
122. Bressler NM, Varma R, Suner IJ, et al. Vision-related function after ranibizumab treatment for diabetic macular edema: results from RIDE and RISE. *Ophthalmology*. 2014;121(12):2461–2472.
123. Mitchell P, Massin P, Bressler S, et al. Three-year patient-reported visual function outcomes in diabetic macular edema managed with ranibizumab: the RESTORE extension study. *Curr Med Res Opin*. 2015;31(11):1967–1975.
124. Gillies MC, Lim LL, Campain A, et al. A randomized clinical trial of intravitreal bevacizumab versus intravitreal dexamethasone for diabetic macular edema: the BEVORDEX study. *Ophthalmology*. 2014;121(12):2473–2481.
125. Loftus JV, Sultan MB, Pleil AM. Changes in vision- and health-related quality of life in patients with diabetic macular edema treated with pegaptanib sodium or sham. *Invest Ophthalmol Vis Sci*. 2011;52(10):7498–7505.
126. Varma R, Bressler NM, Suñer I, et al. Improved vision-related function after ranibizumab for macular edema after retinal vein occlusion: results from the BRAVO and CRUISE trials. *Ophthalmology*. 2012;119(10):2108–2118.

Supplementary material

Table S1 PubMed search

Search	Keywords in search	PubMed hits
1	Patient reported outcome	79,103
2	I AND glaucoma	190
3	I AND cataract	394
4	I AND (cranial nerve palsy OR diplopia OR myasthenia OR intracranial hypertension OR neuro-ophthalmology OR optic nerve OR optic neurotosis OR optic disc OR extraocular)	1,131
5	I AND (retinopathy) OR macular degeneration) OR macular dystrophy) OR retinal dystrophy) OR retinal degeneration) OR maculopathy) OR retina) OR macula)) OR retinitis) OR uveitis) OR choroiditis) OR chorioretinitis	653
6	I AND strabismus) OR amblyopia) OR squint) OR ocular motility) OR pediatric ophthalmology))	183
7	I AND adnexal) OR oculoplastic) OR lid) OR eyelid) OR orbit	514
8	I AND vitreous) OR vitreoretina*) OR epiretina*) OR vitreomacula*	150
9	I and Cornea or refractive	364
10	I and low vision or vision impaired or visually impaired	535
Total screened		4,114

Patient Related Outcome Measures

Dovepress

Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups.

The journal is included in PubMed. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>