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

Tasanee Braithwaite1,2
Melanie Calvert1,3
Alastair Gray4
Konrad Pesudovs5
Alastair K Denniston1,6,8
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. 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. PROM data from trials may be used to inform pharmaceutical labeling claims, clinical guideline development, reimbursement...
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. 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. 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, 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, technical, and cultural factors.

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. In 2001, Massof and Rubin reported that more than 12 PROMs had been developed since 1980. Now there are more than 160 PROM instruments in ophthalmology and optometry. 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. 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 and have informed the US Food and Drug Administration’s guidance. 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.11

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). 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. A further three bolt-on items have been developed for EQ-5D, including a vision bolt on. 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. Valuation sets have since been obtained through various approaches in many other countries, and differences between valuation sets are generally small. The original EQ-5D scale using the UK valuations extends from –0.59 to 1.00, and a more recent UK value set for the EQ-5D-5L extends from –0.28 to 0.95. The mean minimally important difference reported in a review of eight studies in different conditions was 0.074 (range –0.011 to –0.140). 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. 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. 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). 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. Six domains are captured by HUI version 2 (HUI-2) including sensation, mobility, emotion, cognition, self-care, pain, and fertility. 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. 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. 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. 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. 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.

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. 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.

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” group and highlighted those of higher quality, by ophthalmic subspecialty. 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. A shorter version, the (15-item) Brief IVI, has also been validated.

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. 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. 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. 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. 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. 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. 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. 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,
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. 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. 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. Another core outcome set has been proposed for JIA-associated uveitis. 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.

Table 1 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-centered experience of disease and its treatment has been 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. Further work has formed the basis for a hereditary retinal disease item bank.

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. 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. 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. Another core outcome set has been proposed for JIA-associated uveitis. 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. 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. 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. 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.
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.65 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, which includes administration of the CATquest reaffirmed in 2006.19

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, which includes administration of the CATquest reaffirmed in 2006.19

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, which includes administration of the CATquest reaffirmed in 2006.19

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, which includes administration of the CATquest reaffirmed in 2006.19

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, which includes administration of the CATquest reaffirmed in 2006.19

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, which includes administration of the CATquest reaffirmed in 2006.19

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

<table>
<thead>
<tr>
<th>Study name</th>
<th>N</th>
<th>Intervention</th>
<th>Outcome measures</th>
<th>Impact</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARINA and ANCHOR</td>
<td>646 (MARINA) and 379 (ANCHOR) patients with wet AMD</td>
<td>Ranibizumab vs sham or photodynamic therapy</td>
<td>NEI-VFQ-25 at baseline, 12 and 24 months</td>
<td>Improvement in outcomes with intervention</td>
<td>Bressler et al118</td>
</tr>
<tr>
<td>SCORE2 Report 5</td>
<td>362 with CRVO or HRVO and macular edema</td>
<td>Intravitreal bevacizumab vs aflibercept</td>
<td>NEI-VFQ-25 composite and subscale scores</td>
<td>Non-inferiority of bevacizumab</td>
<td>Scott et al119</td>
</tr>
<tr>
<td>OZDRY</td>
<td>100 patients with refractory center involving DMe</td>
<td>5-monthly fixed dosing vs OCT-guided pro-re-nata regimen of Ozurdex</td>
<td>Retinopathy-Dependent QOL, NEI VFQ-25, (RetTSQ) (primary outcome)</td>
<td>No significant difference at month 12</td>
<td>Ramu et al120</td>
</tr>
<tr>
<td>IVAN</td>
<td>610 with active wet AMD</td>
<td>Ranibizumab vs bevacizumab, continuous or discontinuous</td>
<td>EQ-5D</td>
<td>Similar efficacy of drugs in terms of visual acuity. Continuous ranibizumab cost £3.5 million per QALY compared with bevacizumab</td>
<td>Chakravarthy et al121</td>
</tr>
<tr>
<td>RIDE and RISE</td>
<td>382 RIDE and 377 RISE patients with center-involving DME</td>
<td>Ranibizumab vs sham</td>
<td>NEI-VFQ-25 at baseline, 12, and 24 months</td>
<td>Treatment improved vision-related function significantly more than sham</td>
<td>Bressler et al122</td>
</tr>
<tr>
<td>RESTORE Open-label extension</td>
<td>303 with DME</td>
<td>Ranibizumab 0.5 mg vs laser monotherapy</td>
<td>NEI-VFQ-25 (primary outcome)</td>
<td>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</td>
<td>Mitchell et al123</td>
</tr>
<tr>
<td>BEVORDEX</td>
<td>61 patients with center-involving DME</td>
<td>Ozurdex implant every 16 weeks vs bevacizumab every 4 weeks</td>
<td>IVI</td>
<td>Both the groups had significant improvement in IVI scores</td>
<td>Gillies et al124</td>
</tr>
<tr>
<td>MACUGEN</td>
<td>260 with center-involving DME</td>
<td>Pegaptanib sodium vs sham, with focal/grid laser</td>
<td>NEI-VFQ-25, EQ-5D</td>
<td>Clinically and statistically significant differences between groups in composite and subscores, no difference in mean change in EQ-5D utility scores</td>
<td>Loftus et al125</td>
</tr>
<tr>
<td>BRAVO and CRUISE</td>
<td>397 with branch and 392 with central retinal vein occlusion and macular edema</td>
<td>Ranibizumab vs sham</td>
<td>NEI-VFQ-25</td>
<td>Treatment results in significant mean improvement in composite score compared to sham from month 1</td>
<td>Varma et al126</td>
</tr>
</tbody>
</table>

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. Tadić and Rahi further elaborated on issues particular to the development of PROMs for use in children. 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.

**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. 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. In the TEARS trial (Tolera-

**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. 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 eyebrows, and a module for children and young adults with diverse conditions causing facial appearance differences has also been developed. 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.

**Refractive error**

Kandel et al conducted a systematic review for studies using PROMs to assess refractive surgery outcomes. 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. 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.
which identified 34 vision-specific PROMs, and critically appraised the quality of the identified instruments.\textsuperscript{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.\textsuperscript{94} This was assessed for content and quality by Ramey et al and considered to perform reasonably well by classical test criteria.\textsuperscript{90} Generic instruments including NEI-VFQ and SF-36 have been used in an RCT in idiopathic intracranial hypertension (IIH) patients,\textsuperscript{95} and in a study of neurofibromatosis (NF) type 2.\textsuperscript{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.\textsuperscript{76} Disease-specific instruments have been developed for patients with neuromyelitis optica spectrum disorders.\textsuperscript{97,98} The first use of a PROM information system utilizing CAT in patients with neurofibromatosis has also been reported.\textsuperscript{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.\textsuperscript{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).\textsuperscript{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.\textsuperscript{101}

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.\textsuperscript{102} A similar finding was reported when Catquest-9SF was administered to 870 patients in five Dutch hospitals.\textsuperscript{103} 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.\textsuperscript{104} 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.\textsuperscript{105} 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, respondents 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 sub specialties 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. 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. Methods to develop one such “Eye-tem bank” to measure vision and eye disease-related QoL have been outlined. 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. 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. 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.

A systematic review has identified methodological frameworks to measure the health care impact of research. 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**

Tasanee 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


## Supplementary material

### Table S1 PubMed search

<table>
<thead>
<tr>
<th>Search</th>
<th>Keywords in search</th>
<th>PubMed hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Patient reported outcome</td>
<td>79,103</td>
</tr>
<tr>
<td>2</td>
<td>1 AND glaucoma</td>
<td>190</td>
</tr>
<tr>
<td>3</td>
<td>1 AND cataract</td>
<td>394</td>
</tr>
<tr>
<td>4</td>
<td>1 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)</td>
<td>1,131</td>
</tr>
<tr>
<td>5</td>
<td>1 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</td>
<td>653</td>
</tr>
<tr>
<td>6</td>
<td>1 AND strabismus) OR amblyopia) OR squint) OR ocular motility) OR pediatric ophthalmology))</td>
<td>183</td>
</tr>
<tr>
<td>7</td>
<td>1 AND adnexal) OR oculoplastic) OR lid) OR eyelid) OR orbit</td>
<td>514</td>
</tr>
<tr>
<td>8</td>
<td>1 AND vitreous) OR vitreoretina*) OR epiretina*) OR vitreomacula*</td>
<td>150</td>
</tr>
<tr>
<td>9</td>
<td>1 and Cornea or refractive</td>
<td>364</td>
</tr>
<tr>
<td>10</td>
<td>1 and low vision or vision impaired or visually impaired</td>
<td>535</td>
</tr>
<tr>
<td></td>
<td>Total screened</td>
<td>4,114</td>
</tr>
</tbody>
</table>