

An introduction to patient-reported outcome measures in ophthalmic research

AK Denniston^{1,2}, D Kyte³, M Calvert^{3,4}
and JM Burr^{5,6}

REVIEW

Abstract

Clinical outcomes, such as quantifying the extent of visual field loss by automated perimetry, are valued highly by health professionals, but such measures do not capture the impact of the condition on a patient's life. Patient-reported outcomes describe any report or measure of health reported by the patient, without external interpretation by a clinician or researcher. In this review, we discuss the value of the measures that capture this information (patient-reported outcome measures; PROMs), and why they are important to both the clinician and the researcher. We also consider issues around developing or selecting a PROM for ophthalmic research, the emerging challenges around conducting and reporting PROMs in clinical trials and highlight best practice for their use. Search terms for this review comprised: (1) (patient-reported outcomes OR patient-reported outcome measures) AND (2) randomised controlled trials AND (3) limited to ophthalmic conditions. These terms were expanded as follows: (((('patients'(MeSH Terms) OR 'patients'(All Fields) OR 'patient'(All Fields)) AND ('research report'(MeSH Terms) OR ('research'(All Fields) AND 'report'(All Fields)) OR 'research report'(All Fields) OR 'reported'(All Fields)) AND outcomes(All Fields)) OR (('patients'(MeSH Terms) OR 'patients'(All Fields) OR 'patient'(All Fields)) AND ('research report'(MeSH Terms) OR ('research'(All Fields) AND 'report'(All Fields)) OR 'research report'(All Fields) OR 'reported'(All Fields) AND ('outcome assessment (health care)'(MeSH Terms) OR ('outcome'(All Fields) AND 'assessment'(All Fields) AND ('health'(All Fields) AND 'care'(All Fields)) OR 'outcome assessment (health care)'(All Fields) OR ('outcome'(All Fields) AND 'measures'(All Fields)) OR 'outcome measures'(All Fields)))) AND

('randomized controlled trial'(Publication Type) OR 'randomized controlled trials as topic'(MeSH Terms) OR 'randomised controlled trials'(All Fields) OR 'randomized controlled trials'(All Fields)) AND (ophth*(All Fields)). The authors also utilised the extensive non-ophthalmic literature and online resources relating to PROs and PROMs to inform this review.

Eye (2014) 28, 637–645; doi:10.1038/eye.2014.41; published online 14 March 2014

Introduction

Who defines whether a clinical intervention is successful? It is increasingly recognised that the clinical outcome measures that health professionals value so highly (such as a visual acuity or a global index of visual field loss in one eye or central macular thickness) fail to capture the full impact of ophthalmic disease on a patient's life. Although ophthalmologists strive to be patient-centred in their approach, dependence on clinical parameters alone is inadequate and may not correlate with the patient's experience of their disease.^{1,2} Patient-reported outcomes (PROs) and the measures that capture them help address this important deficit in the clinician's knowledge both in the clinical and research settings.

Patient-reported outcome measures: what are they?

A PRO describes any report or measure of the patient's health that comes directly from the patient without interpretation by a clinician or a researcher. The measures that capture this information are known as patient reported outcome measures (PROMs).^{3,4} Commonly, PROMs take the form of either paper- or electronic-based patient questionnaires, which

¹Centre for Translational Inflammation Research, Queen Elizabeth Hospital Birmingham, University of Birmingham, Birmingham, UK

²Department of Ophthalmology, Queen Elizabeth Hospital Birmingham, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

³Patient Reported Outcomes Research Group, School of Health and Population Sciences, University of Birmingham, Birmingham, UK

⁴MRC Midland Hub for Trials Methodology Research, University of Birmingham, Birmingham, UK

⁵School of Medicine, University of St Andrews, Fife, UK

⁶NIHR Biomedical Research Centre for Ophthalmology, University College London, London, UK

Correspondence: AK Denniston, Centre for Translational Inflammation Research, Queen Elizabeth Hospital Birmingham, University of Birmingham, West Midlands, Birmingham B15 2WB, UK. Tel: +44 (0) 121 627 8568; Fax: +44 (0) 121 627 8789. E-mail: a.denniston@bham.ac.uk

Received: 13 August 2013
Accepted in revised form: 20 January 2014
Published online: 14 March 2014

measure treatment benefit or risk, and should be validated for use in the target population.⁴

PROMs may measure outcome in absolute terms, for example, symptom severity, or may measure change, for instance the extent to which a symptom has improved/worsened. PROMs may be used to assess health-related quality of life (a multidimensional concept that usually includes self-report of the way in which physical, emotional, and social well-being are affected by a disease or its treatment), or to assess symptoms or perception of health status.

A helpful approach to thinking about PROMs is to consider their component parts. The specific questions contained within a PROM are known as items, each of which elicit an answer from two or more options. Items that are related may be grouped as domains or subscales. Although there are a number of ways of classifying PROs, we recommend categories that reflect the World Health Organisation International Classification of Functioning, Disability and Health (WHO ICF).⁵ Accordingly, we group items according to whether they reflect disease impact on (1) bodily symptoms/functions, (2) activities, or (3) social participation. Although not the focus of this review, it should be noted that some measures explore aspects of patient satisfaction (such as with the clinical care they are receiving); the recently described glaucoma POEM is an example of such a hybrid tool that contains items relating to fear of blindness, acceptability and side effects of treatment, and impact on daily life in addition to satisfaction/experience items.¹

Why do PROMS matter?

PROMs are an increasing part of the clinical effectiveness research⁶ and may be used as the primary outcome (such as in the Effectiveness in Angle Closure of Lens Extraction (EAGLE) study).⁷ More commonly PROMs are used as secondary outcomes, usually complementing a more familiar 'clinical' primary outcome such as visual acuity. Recent examples include the inhibition of VEGF in age-related choroidal neovascularisation trial,⁸ which used the general health utility measure EQ-5D, and the Full-thickness Macular Hole and Internal Limiting Membrane Peeling Study (FILMS),⁹ which used the vision-specific measure, the National Eye Institute Visual Function Questionnaire (VFQ-25). Further examples are provided in Table 1. In addition to providing evidence of clinical effectiveness, PROs are increasingly used to inform licensing and reimbursement decisions.¹⁰ A study in 2004 reviewed the effectiveness end points reported for clinical trials resulting in Food and Drugs Administration (FDA) approval of new drugs (specifically 'new molecular entities (NMEs)') from 1997

to 2002.¹¹ Of 215 new product labels, PROs were reported in 30% and were the only end point cited in 11%. A more recent study of drugs approved between January 2000 and June 2012 (specifically NMEs and biologic licence applications) looked at FDA approval of 'PRO labelling' (ie, where the label of the drug can include claims of treatment benefit with regard to a PRO). Of 308 drugs approved during this period, 70 were specifically approved for 'PRO labelling', and for 57 of these, this was based on studies in which the PRO was the primary outcome.¹²

There are some areas of medicine where the measurement of a PRO is self-evidently essential and therefore should be a primary trial outcome (eg, pain in a trial of analgesic efficacy). Arguably, however, the assessment of PROs should be a required part of all comparative effectiveness trials. Indeed, PROMs are of particular value in helping inform health-care decisions where there is a significant trade-off to be made between therapeutic benefit and undesirable side effects, or to differentiate between two interventions with apparently similar benefits in terms of the primary outcome (eg, equal reduction in anatomical closure of a macular hole), but which might vary in terms of patient-reported ability to read or drive. Given that in these situations the relative cost of treatments may also need to be considered, it is of no surprise that many studies include a particular type of PROM known as a utility measure, which can be used to provide an estimate of incremental effectiveness in terms of quality-adjusted life years for use in cost-effectiveness analyses.¹³ Incorporating an economic evaluation within a trial provides valuable information for clinical decision-making and policy in particular when the clinical effectiveness of the interventions is similar.

For example, in the FILMS study, comparing surgery with or without peeling of the internal-limiting membrane (ILM), the findings suggest that surgery with ILM peeling is likely to be a cost-effective option for the treatment of macular holes.¹⁴

In the United Kingdom, PROMs are also becoming integrated into routine clinical practice.³ Since April 2009, NHS England has used PROMs to audit PROs following four common operations: inguinal hernia, varicose vein surgery, knee replacement, and hip replacement. For example, from 1 April 2011 to 31 March 2012, there were 247 688 PROM-eligible procedures carried out in hospitals for which 184 829 (74.6%) preoperative questionnaires were obtained and 136 899 (55%) have preoperative and postoperative questionnaires.¹⁵ This systematic and continuous analysis of routine surgical practice provides a wealth of important information, in addition to the standard measures of length of stay and postoperative mortality.³ The internal comparator of a preoperative questionnaire improves the ability to

Table 1 Selected recent ophthalmic trials and their usage of PROMs

Name of the study	Subject	PRO instrument	Reference
IVAN	Neovascular AMD	EQ-5D MacDQoL MacTSQ	Chakravarthy <i>et al</i> ⁸ Mitchell <i>et al</i> ⁴⁸ Mitchell and Bradley ⁴⁹
BRAVO	Branch retinal vein occlusion	VFQ-25	Campochiaro <i>et al</i> , ⁵⁰ Varma <i>et al</i> ⁵¹
CRUISE	Central retinal vein occlusion	VFQ-25	Varma <i>et al</i> , ⁵¹ Brown <i>et al</i> ⁵²
COPERNICUS	Central retinal vein occlusion	VFQ-25	Brown <i>et al</i> ⁵³
RESTORE	Diabetic macular oedema	VFQ-25	Mitchell <i>et al</i> ⁵⁴
FILMS	Macular hole	EQ-5D VFQ-25	Lois <i>et al</i> ⁹
ENDURE	Posterior-segment uveitis	EQ-5D VFQ-25 VCM-1	Dick <i>et al</i> ⁵⁵
HURON	Posterior-segment uveitis	EQ-5D SF-36	Lightman <i>et al</i> , ⁵⁶ Lowder ⁵⁷
MUST	Posterior-segment uveitis	EQ-5D SF-36 VFQ-25	Frick <i>et al</i> ⁵⁸
EAGLE	Primary angle-closure Glaucoma	EQ-5D VFQ-25 Glaucoma Utility Index	Azuara-Blanco <i>et al</i> ⁷

measure the real impact of the intervention rather than simply reflecting demographic differences between regions, ethnic distributions, or case mix. Although the focus of this article is the use of PROMs in ophthalmic research, it is important to realise that PROMs are likely to become a part of routine clinical practice in ophthalmology, and familiarity with them is thus of increasing value to clinical staff, as much as researchers. It should also be recognised that the challenges discussed in this article regarding using PROMs in research are magnified when using these tools in the uncontrolled environment of clinical practice.

The place of PROMs in ophthalmic research

The development of most PROMs in ophthalmic research has been driven by the recognition that clinical tests, such as visual acuity, perimetry, and ocular coherence tomography, imperfectly capture the extent to which patients are impacted by sight impairment. Although PROs in ophthalmology do not need to be solely about vision (eg, severity of pain may be a significant outcome in dry eye disease as recognised by the Ocular Surface Disease Index),¹⁶ most PROMs were developed with a view of assessing visual function and its consequent impact on activities of daily living and social participation. The first measures were introduced in the early 1980s and include the Visual Function Index

developed by Bernth-Peterson¹⁷ and the 'Functional Problems of the Visually Impaired' tool commissioned by the National Eye Institute (NEI) and developed by Bikson and Bikson of the Rand corporation.^{18,19} It is beyond the scope of this review to discuss all the vision-related quality of life instruments that have since been developed, but it is worth highlighting the development of the National Eye Institute Visual Function Questionnaire (NEI VFQ), which was first published in 1998 and which is, in its shorter form (NEI VFQ-25), the most widely used vision-related PROM in the world (see Box 1).

Selection of PROMs for use in randomised controlled trials

The most appropriate PROM for a study will depend on the study objectives and the target population. It is important for the investigator to first identify what domain or domains they want to measure, which will depend on the underlying hypothesis. For example, this might be difficult with near-vision activities in a study of presbyopia, difficult with peripheral vision in glaucoma or ocular pain in a study of scleritis, or keratoconjunctivitis sicca. It is recommended that there is patient engagement from the outset, including patient input into the selection of domains to ensure that the study does indeed capture outcomes that matter to the patient. Having selected the

Box 1 The NEI VFQ and the NEI VFQ-25: a case study in PROM development for ophthalmic research.

The NEI Visual Function Questionnaire (VFQ) was one of the earliest instruments used to try to assess the impact of visual problems on patients' emotional and psychological well-being, as well as their activities. Items were generated through focus groups, comprising around 250 patients with a wider range of conditions than had been used in earlier studies, although still mainly from the 'big four' of glaucoma, diabetic retinopathy, cataract, and age-related macular degeneration. The majority of these patients (73%) had a best-corrected visual acuity of 6/12 or greater in their better eye. Fifty-one items were generated in this way, which were then grouped into 13 domains or subscales: general health, general vision, ocular pain, vision expectations, near vision, distance vision, social problems, mental health, role problems, dependency, driving, peripheral vision, and colour vision. Subsequent studies to validate this tool and to assess reliability were performed on around 600 patients comprising a similar spectrum of disease, but with a reference group (~20% of the study population) who had no known ophthalmic disease. The internal consistency (i.e. the extent to which items within a single domain agreed) was classed as high or good for domains measuring near activities, distance activities, role difficulties, dependency, social function and mental health (as measured by Cronbach alpha >0.85). These domains also scored highly on test-retest reliability (as measured by intraclass correlation coefficients >0.8). Other domains scored more poorly on both internal consistency and reliability. General vision, near vision, and distance vision domains were found to correlate strongly with binocular visual acuity.⁵⁹ These parameters are discussed further in the main text.

In 1998, the same group published a short form of the NEI VFQ, which aimed to preserve the original's multidimensions and retain its validity and reliability, while reducing the completion time to 5 min, thus lessening the patient burden. The resulting NEI VFQ-25 was developed through item reduction of the original with a study population of over 800 patients, and this is the version that is now in widespread usage. This reduced version comprises 25 vision-related items plus one general health item, enabling scoring of the following domains or 'subscales': general health, general vision, ocular pain, near activities, distance activities, vision-specific social functioning, vision-specific mental health, vision-specific role difficulties, vision-specific dependency, driving, colour vision, and peripheral vision. Some of these subscales (such as colour vision) are derived from a single question whereas others are derived from multiple questions (such as vision-specific mental health, which is scored as the mean of four separate questions). High scores represent better levels of functioning. Recoding of scores between questions is performed in a standardised way (resulting in a range of 0–100) to permit the averaging of subscales and the generation of a final composite score, in which each subscale has equal weight, and where 100 represents the highest possible score.

The original is described as 'English for the USA', but there are numerous other translations with over 70 interviewer-delivered translations and over 40 self-administered versions available.⁶⁰ Although it remains the most widely used of all ophthalmic PROMs, there are some methodological concerns over the NEI VFQ-25 leading some investigators to suggest further refinements.⁶¹

domains of interest, an appropriate PROM should be chosen based on its (1) content, (2) measurement properties, and (3) practical application encompassing acceptability, feasibility, and interpretability.

PROM content

In assessing the suitability of a PROM with regard to content, available measures can be analysed or 'mapped' in terms of their items and domain(s) to assess the extent to which they are likely to gather the outcome of interest. At this point, it should be considered whether the study is best served by a generic instrument, a vision-specific instrument, or a condition-specific instrument (if available). Generic instruments, such as the SF-36, assess broad aspects of the quality of life and health status. Vision-specific instruments, such as the NEI VFQ-25, are focused on aspects surrounding visual and emotional function, and so are more sensitive to issues experienced by ophthalmic patients. Condition-specific instruments are targeted to the needs of patients with a single condition. Examples include the VF-14 for patients with cataract or the Glau QOL 36 for patients with glaucoma.^{20,21} Although some fields have no condition-specific instruments, others, such as glaucoma, have seen a rapid growth in available instruments. In 2011,

Hamzah *et al*⁵ conducted a systematic review of available instruments for use in a glaucoma population, identifying 16 vision-specific, 16 glaucoma-specific, and one combined tool. One of the key issues that this study illustrates is the variability of PROMs in this field, both in terms of their scope (which is not always accurately reflected by their title) and their quality. Specifically, it was noted that the glaucoma-specific/combined PROMs comprised five measures focused on the disease ('glaucoma status measures') and twelve measures related to glaucoma medication.⁵ Understanding the construct of the measure is the first step in choosing the appropriate PROM for the study.

PROM measurement properties

Selection of a PROM must also take into account its measurement properties. This may include its reliability, validity, responsiveness, precision, interpretability, acceptability, and feasibility, although it should be noted that these properties do not depend solely on the PROM but will vary according to the population in which it is used.²²

Reliability is scored in terms of test-retest reliability (reproducibility in stable patients over time) and inter-rater reliability (agreement between two trained

interviewers assessing the same patient at the same time; not applicable for self-administered interviews). In addition, internal consistency (correlation between items within a domain) is often calculated for PROMs, which contain multiple subscales that each measure a distinct construct. High levels of internal consistency indicate that items grouped within such subscales are indeed evaluating aspects of the same concept.²³ Recommendations of the European Regulatory Issues on Quality of Life Assessment Group advise minimum standards for these parameters: specifically these are intraclass correlation of 0.7 for test-retest reliability, 0.8 for inter-rater reliability, and a Cronbach alpha of 0.7 for internal consistency (all scored out of 1.0).¹⁰

Validity is in theory, the extent to which the instrument measures what it is intended to, and in practice, the extent to which the instrument correlates with other instruments and/or clinical outcomes and can detect differences in disease severity between groups of patients. It should be assessed in terms of content validity (the extent to which individual items adequately assess the domains of interest without redundancy), construct validity (the extent to which predetermined logical statistical and clinical relationships hold true for the tool), and criterion validity (the correlation with a reference or 'gold standard' instrument).

Responsiveness describes the capacity of the tool to detect significant differences, whether change over time, or differences between cohorts at the same time; it is commonly measured by comparison with a reference tool. Precision reflects the gradations of response, that is, are there enough distinctions within the scale of measurement to capture different states.

PROMs should be validated for use in the target population, such that it is clear that the tool performs adequately, both in terms of its psychometric characteristics and responsiveness, in the intended setting.⁴ This process may be undertaken as a part of a dedicated cross-sectional study, or within an intervention-based clinical trial or longitudinal study.²⁴ Readers interested in PROM development and validation are advised to refer to guidance presented by Fayers and Machin²⁵ or Johnson *et al*²⁶ for further information.

PROM acceptability, feasibility, and interpretability

To be useful in either a research or clinical context, measures must also be acceptable, feasible, and interpretable. Acceptability covers features such as the time required to complete the questionnaire, and will reflect the physical and mental capacity of the population being surveyed. Feasibility considers the resources required to use the measures correctly. Acceptability and

feasibility may vary according to patient group and context. It may therefore be appropriate to explore these issues with patient/clinical groups or in a pilot study before using a PROM in a larger-scale clinical trial. Interpretability describes the ease with which we can understand differences in score. The smallest difference that matter to patients, the minimally important difference,²⁷ needs to be determined both for score interpretation and also for ensuring that randomised controlled trials (RCTs) are powered to detect important differences if they exist.²⁵

Resources for selecting PROMs

Recognising the importance of PROMs, but also the rapid expansion of measures of variable quality, methodologists, and trialists, have established a number of organisations to provide guidance for PROM development, evaluation, and selection. The Patient-Reported Outcome and Quality of life Instruments Database (PROQOLID; www.proqolid.org)²⁸ is an online source of information on PROMs, which provides key facts for each PROM including author, purpose (disease, population, objective), characteristics (type of instrument, eg, health-related quality of life, patient satisfaction, physical functioning, psychological functioning, administration mode, number of questions, recall period), and languages available. The database can be searched, both for a specific instrument (eg, 'NEI VFQ-25'), or for a particular indication (eg, 'cataract'). For the evaluation of PROMs, the CONsensus-based Standards for the selection of health Measurement INSTRUMENTS (COSMIN) group undertook a multidisciplinary Delphi-based initiative to develop a critical appraisal tool that can be used as a checklist (available at www.cosmin.nl)²⁹. Similarly, the International Society for Quality of Life Research (ISOQOL) has published recommendations of minimum measurement standards for PROMs.³⁰⁻³² Finally, the European Medicines Authority and the US FDA provide guidance regarding the use of PROMs in trials that may be submitted as evidence in applications for regulatory approval.^{4,10}

Designing RCTs with a PROM

PROMs are an integral part of comparative effectiveness studies. Until recently PROMs were the 'poor relation' of clinical outcomes, but their importance is increasingly recognised with greater attention being given to their place in study design, conduct, and reporting.³³ Best practice needs to start at the design phase to ensure that the study does indeed meet its objectives, with good compliance, maximal data capture, and interpretable valid results.

Key issues and challenges should be determined at the outset and specified in the study protocol;²⁵ this may be assisted by checklists such as the one by King (University of Sydney, Psycho-Oncology Cooperative Research Group)³⁴ There should be a justification for the use of a PROM within the design, a hypothesis as to the direction and preferably size of any change expected and the evidence on which this estimate is based (including power calculation). It should be specified whether the PRO is a primary or secondary outcome, the statistical analysis that will be used, and which domains this will include, to minimise the problems associated with multiple statistical testing. The timing of PRO measurement needs to be prespecified, both with regard to the stage in the trial (eg, baseline and intervals), but also in terms of when measurement will occur during a visit (commonly before any procedure or treatment). Given that missing data are a particular challenge in PROM research, it is also advisable to prespecify the procedures that will be adopted to maximise compliance and what methods will be used to handle missing data. The mode of administration should also be specified; this is particularly important where patients (such as with sight loss) may not be able to complete forms designed to be self-administered. The use of 'proxies' to complete the forms on behalf of the patient should again be prespecified and documented.²⁵

Using PROMs in RCTs

During the study there are a number of important considerations to ensure that PROM assessment is carried out in a standardised and valid way. It is clearly important that the schedule specified in the protocol is adhered to, and in particular, that the baseline assessment is carried out before randomisation, particularly in open trials where knowledge of the intervention may cause anxiety and impact on the PROM.³⁶

A recurring challenge with PROMs centres around compliance with data collection, that is, how many individual questionnaire items, and how many forms, are missing within a trial. There are two problems. First, missing data points lead to a loss of statistical power, potentially resulting in a type II error (ie, failure to detect a significant difference). Second, the attrition may not be random, but rather due to death, depression, or some other factor resulting in a biased assessment.^{25,36}

In order to improve data capture, the importance of the PROM data to the study and study question should be emphasised.³⁷ PROM data can be collected face to face in the clinical setting or via postal questionnaires coordinated by the study office. The former can be problematic as the clinical setting may influence participant response. The latter may have a greater risk

of missing data. Data capture may be improved by having a centrally managed PRO data monitoring system in place for evaluating compliance across study sites, with data collection reminders to participants where needed and chasing-up of missing items.³⁸ Electronic versions (e-PROMs) that may be filled in online with email prompts are an alternative.

Analysing PROMs in RCTs

As PROMs typically comprise many subdomains and are often assessed at a number of time points within a study, there is a danger of a type I error (ie, a false-positive result) from multiple statistical testing. One option is to use a single test of the aggregate PROM score, but this may result in 'dilution' of differences within a key domain of interest. Alternatively, domains of interest can be prespecified as end points linked closely to address the study hypotheses. In addition, the time point for the outcome data needs to be prespecified or aggregated as an area under the curve analysis. In order to reduce bias, analysis should be undertaken on an intention-to-treat basis (ie, all patients randomised are included in the final analysis whether or not they actually received treatment). This is important as participants lost to or withdrawing from the study may show significant differences in terms of PROM scores, which might be missed.^{10,39,40}

Perhaps the biggest challenges in PROM analysis arise around the handling of missing data. First, the extent of missing data must be reported, including whether this appears to be at random (both between treatment groups, and between different items within the PROM). Second, the extent to which this may undermine the validity of the study needs to be assessed. Third, the missing data may be estimated or 'imputed', although it should be recognised that none of the techniques to do this are perfect. Indeed, the FDA recommend using two prespecified imputation techniques and that any significant difference between them, as determined during sensitivity analysis, is regarded as a cause for concern.⁴ The analytic plan, including details of handling multiplicity and missing data, should be prespecified in the study protocol.⁴¹

Reporting PROMs in RCTs

A major problem with the reporting of PROMs in clinical trials is that they are underreported, or only reported years after the main study.^{4,10,22,27,38,42} Although one could argue that this is of little consequence when the PROM outcome moves in the same direction as the clinical outcome, the PRO data may help inform future patients about the likely impact of treatment on outcomes that matter to them. In some instances, PROs may moderate the interpretation of a beneficial clinical

outcome. Selective reporting of outcomes (unless prespecified) should be avoided, whether these are PROMs or clinical outcomes.⁴³ The Consolidated Standards of Reporting Trials Patient Reported Outcomes (CONSORT PRO) is a recent extension to the CONSORT guidelines specifically around PROs. They provide five checklist items recommended for RCTs to be reported in all RCTs, in which PROs are a primary or important secondary outcome. These are (1) that the PRO is identified as a primary or secondary outcome in the abstract; (2) that the hypothesis regarding the PRO is provided; (3) that evidence of the instrument's validity and reliability is provided; (4) that statistical approaches for dealing with missing data are explicitly stated; and (5) that limitations and generalisability with specific regard to the PRO data are discussed.^{31,33}

Disadvantages of PROMs in RCTs

PRO assessment in trials may be costly, both in terms of the use of licensed instruments, costs for administration, data management, trial analyses, and patient time. Thus, PROMs should only be included when there is a clear rationale and hypothesis.

Missing PRO data can be a particular problem in trials. Retrospective data capture may not be possible²⁵ and often PRO data are not missing at random: commonly, those participants with the poorest outcomes fail to complete items or questionnaires.⁴⁴ Rates of missing PRO data within a trial may be higher in the following instances:

- Where trial participants are overly burdened with lengthy (or numerous PROMs).⁴
- Where participants are given PROMs containing questions they consider intrusive, or of questionable relevance.⁴⁵
- Where data collection staff are unaware of the importance of the PROM to the trial outcome and are not given specific instructions regarding the prevention/management of missing data.

Thus, when designing a trial, PRO data collection should be limited so that the average participant can complete the process in a reasonable time. The PROM should be carefully selected, with input from the target population, to ensure acceptability to the participants in the study.²⁴ Finally, the protocol should contain information on the methods that should be in place to prevent avoidable missing PRO data, for example, education of data collection staff and centralised monitoring of data compliance with back-up data retrieval where possible.⁴⁶

Trial management groups should also be aware that PROMs occasionally detect concerning levels of participant distress (eg, severe depression or suicidal thoughts) or physical symptoms (eg, pain) that may require an immediate response.⁴⁷ Accordingly, there should be an *a priori* plan in place to monitor and manage alerts should they arise, in line with the risk profile of the trial; this may also require additional resources in some instances.

The future of PROMs in ophthalmic research

PROs are a vital part of modern ophthalmology, and are now finding their place in both clinical trials and in routine clinical practice. It should however be recognised that the measures that evaluate these outcomes (PROMs) are complex, and need to be utilised and interpreted with care. Guidance from international consortia is directed towards improving the quality of PROMs, assisting researchers in the selection of suitable instruments and helping them collect, analyse, and publish the results in an accurate and meaningful way.^{29,32} The increasing focus on PROs in high-quality ophthalmic trials is an important step, which has the potential to inform care, encourage shared decision-making, and help deliver a patient-centred approach to clinical practice.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

DK is funded by the National Institute for Health Research (NIHR) School for Primary Care Research (SPCR). MC is funded by the Medical Research Council Midland Hub for Trials Methodology Research (Medical Research Council grant G0800808). JB is supported in part by the Department of Health's NIHR Biomedical Research Centre for Ophthalmology at Moorfields Eye Hospital and UCL Institute of Ophthalmology.

Disclaimer

The views expressed in the publication are those of the authors and not necessarily those of the Department of Health.

References

- 1 Somner JEA, Sii F, Bourne RR, Cross V, Burr JM, Shah P. Moving from PROMs to POEMs for glaucoma care: a qualitative scoping exercise. *Invest Ophthalmol Vis Sci* 2012; **53**(9): 5940–5947.
- 2 Rahi JS, Cumberland PM, Peckham CS. Visual impairment and vision-related quality of life in working-age adults. *Ophthalmology* 2009; **116**(2): 270–274.

- 3 Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013; **346**: f167.
- 4 FDA Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. Available <http://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm193282.pdf>. Accessed 2013.
- 5 Hamzah JJC, Burr JM, Ramsay CRC, Azuara-Blanco AA, Prior MM. Choosing appropriate patient-reported outcomes instrument for glaucoma research: a systematic review of vision instruments. *Qual Life Res* 2011; **20**(7): 1141–1158.
- 6 Basch EM, Abernethy A, Mullins CD, Tiglaio MR, Tunis SR. Ev1 development of a guidance for including patient-reported outcomes (Pros) in post-approval clinical trials of oncology drugs for comparative effectiveness research (Cer). *Value Health* 2011; **14**(3): A10–A10.
- 7 Azuara-Blanco A, Burr JM, Cochran C, Ramsay C, Vale L, Foster P *et al*. The effectiveness of early lens extraction with intraocular lens implantation for the treatment of primary angle-closure glaucoma (EAGLE): study protocol for a randomized controlled trial. *Trials* 2011; **12**: 133.
- 8 Chakravarthy U, Harding SP, Rogers CA, Downes SM, Lotery AJ, Culliford LA *et al*. Alternative treatments to inhibit VEGF in age-related choroidal neovascularisation: 2-year findings of the IVAN randomised controlled trial. *Lancet* 2013; **382**(9900): 1258–1267.
- 9 Lois N, Burr J, Norrie J, Vale L, Cook J, McDonald A *et al*. Internal limiting membrane peeling versus no peeling for idiopathic full-thickness macular hole: a pragmatic randomized controlled trial. *Invest Ophthalmol Vis Sci* 2011; **52**: 1586–1592.
- 10 Chassany O, Sagnier P, Marquis P, Fullerton S, Aaronson N. For the European Regulatory Issues on Quality of Life Assessment Group. Patient-reported outcomes: the example of health-related quality of life—a European guidance document for the improved integration of health-related quality of life assessment in the drug regulatory process. *Drug Inf J* 2002; **36**(1): 209–238.
- 11 Willke RJ, Burke LB, Erickson P. Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels. *Control Clin Trials* 2004; **25**(6): 535–552.
- 12 Gnanasakthy A, Lewis S, Clark M, Mordin M, DeMuro C. Potential of patient-reported outcomes as nonprimary endpoints in clinical trials. *Health Qual Life Outcomes* 2013; **11**: 83.
- 13 NICE. The Guidelines Manual. National Institute of Health and Clinical Excellence; 2009. Available at http://www.nice.org.uk/media/68D/29/The_guidelines_manual_2009_-_Chapter_7_Assessing_cost_effectiveness.pdf. Accessed 1 August 2013.
- 14 Ternent L, Vale L, Boachie C, Burr JM, Lois N. Cost-effectiveness of internal limiting membrane peeling versus no peeling for patients with an idiopathic full-thickness macular hole: results from a randomised controlled trial. *Br J Ophthalmol* 2012; **96**(3): 438–443.
- 15 Health & Social Care Information Centre. *Patient Reported Outcome Measures*. Available at <http://www.hscic.gov.uk>. Accessed 31 July 2013.
- 16 Schiffman RM, Christianson MD, Jacobsen G, Hirsch JD, Reis BL. Reliability and validity of the ocular surface disease index. *Arch Ophthalmol* 2000; **118**(5): 615–621.
- 17 Bernth-Petersen PP. Visual functioning in cataract patients. Methods of measuring and results. *Acta Ophthalmol* 1981; **59**(2): 198–205.
- 18 Massof RW, Rubin GS. Visual function assessment questionnaires. *Surv Ophthalmol* 2001; **45**(6): 531–548.
- 19 Bikson TH, Bikson TK. *Functional problems of the visually impaired: a research approach*. Rand Corporation, Rand Paper Series: Santa Monica, CA, 1981, P-6648, pp 1–24.
- 20 Steinberg EP. The VF-14An index of functional impairment in patients with cataract. *Arch Ophthalmol* 1994; **112**(5): 630–638.
- 21 Béchetoille A, Arnould B, Bron A, Baudouin C, Renard J-P, Sellem E *et al*. Measurement of health-related quality of life with glaucoma: validation of the Glau-QoL© 36-item questionnaire. *Acta Ophthalmol* 2008; **86**(1): 71–80.
- 22 Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998; **2**(14):i–iv 1–74.
- 23 Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; **63**: 737–745.
- 24 Coyne K, Tubaro A, Brubaker L, Bavendam T. Development and validation of patient-reported outcomes measures for overactive bladder: a review of concepts. *Urology* 2006; **68**(Suppl 2a): 9–16.
- 25 Fayers P, Machin D. *Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes*, 2nd edn. John Wiley & Sons, Ltd: Chichester, UK, 2007.
- 26 Johnson J, Aaronson N, Blazeby J, Bottomley A, Fayers P, Koller M *et al*. 2011, EORTC QUALITY OF LIFE GROUP Guidelines for Developing Questionnaire Modules (Fourth Edition). European Organization for Research and Treatment of Cancer. Available http://groups.eortc.be/qol/sites/default/files/archives/guidelines_for_developing_questionnaire_final.pdf. [Last accessed November 2013].
- 27 Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002; **77**(4): 371–383.
- 28 Patient-Reported Outcome and Quality of life Instruments Database (PROQOLID). Available at www.proqolid.org. [Last accessed November 2013].
- 29 Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) Available at <http://www.cosmin.nl/>. (last accessed August 2013).
- 30 Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF *et al*. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013; **22**(8): 1889–1905.
- 31 Brundage M, Blazeby J, Revicki D, Bass B, de Vet H, Duffy H *et al*. Patient-reported outcomes in randomized clinical trials: development of ISOQOL reporting standards. *Qual Life Res* 2012; **22**(6): 1161–1175.
- 32 ISOQOL-International Society for Quality of Life Research. [isoqol.org](http://www.isoqol.org/). Available at <http://www.isoqol.org/>. (last accessed August 2013).
- 33 Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD *et al*. Reporting of patient-reported outcomes

- in randomized trials The CONSORT PRO extension. *JAMA* 2013; **309**(8): 814–822.
- 34 Psycho-Oncology Co-operative Research Group (University of Sydney) Resources. Available at <http://www.pocog.org.au/>. (last accessed November 2013).
- 35 Brooks MM, Jenkins L, Schron EB, Steinberg JS, Cross J, Paeth D. Quality of life at baseline: is assessment after randomization valid? *Med Care* 2010; **36**(10): 1515–1519
- 36 Bernhard JJ, Cella DFD, Coates ASA, Fallowfield LL, Ganz PAP, Moinpour CMC *et al*. Missing quality of life data in cancer clinical trials: serious problems and challenges. *Statist Med* 1998; **17**(5-7): 517–532.
- 37 Fayers PM, Hopwood P, Harvey A, Girling DJ, Machin D, Stephens R. Quality of life assessment in clinical trials—guidelines and a checklist for protocol writers: the U.K. Medical Research Council experience. *MRC cancer trials office. Eur J Cancer* 1997; **33**(1): 20–28.
- 38 Kyte DG, Draper H, Ives J, Liles C, Gheorghe A, Calvert M. Patient reported outcomes (PROs) in clinical trials: is ‘in-trial’ guidance lacking? a systematic review. *PLoS One* 2013; **8**(4): e60684.
- 39 Testa MAM, Hollenberg NKN, Anderson RBR, Williams GHG. Assessment of quality of life by patient and spouse during antihypertensive therapy with atenolol and nifedipine gastrointestinal therapeutic system. *Am J Hypertens* 1991; **4**(4 Pt 1): 363–373.
- 40 Sloan J, Dueck A, Erickson P, Guess H, Revicki D, Santanello N *et al*. Analysis and interpretation of results based on patient-reported outcomes. *Value Health* 2010; **10**(S2): S106–S115.
- 41 Fairclough DL. *Design and Analysis of Quality of Life Studies in Clinical Trials*, 2nd edn. Chapman & Hall/CRC: Boca Raton, Florida, 2010.
- 42 Brundage M, Bass B, Davidson J, Queenan J, Bezjak A, Ringash J *et al*. Patterns of reporting health-related quality of life outcomes in randomized clinical trials: implications for clinicians and quality of life researchers. *Qual Life Res* 2011; **20**(5): 653–664.
- 43 Revicki D, Erickson P, Sloan J, Dueck A, Guess H, Santanello N *et al*. Interpreting and reporting results based on patient-reported outcomes. *Value Health* 2010; **10**(S2): S116–S124.
- 44 Fairclough D, Peterson HF, Chang V. Why are missing quality of life data a problem in clinical trials of cancer therapy? *Stat Med* 1998; **17**: 667–677
- 45 Kyte D, Ives J, Draper H, Keeley T, Calvert M. Inconsistencies in quality of life data collection in clinical trials: a potential source of bias? Interviews with research nurses and trialists. *PLoS One* 2013; **8**(10): e76625.
- 46 Basch EM, Abernethy A, Mullins CD, Tiglaio MR, Tunis SR. Development of a guidance for including patient-reported outcomes (PROs) in post-approval clinical trials of oncology drugs for comparative effectiveness research (CER). *Value Health* 2011; **14**(3): A10.
- 47 Kyte D, Draper H, Calvert M. Patient-reported outcome alerts: ethical and logistical considerations in clinical trials. *JAMA* 2013; **310**(12): 1229–1230.
- 48 Mitchell J, Wolffsohn J, Woodcock A, Anderson SJ, Ffytche T, Rubinstein M *et al*. The MacDQoL individualized measure of the impact of macular degeneration on quality of life: reliability and responsiveness. *Am J Ophthalmol* 2008; **146**: 447–454.
- 49 Mitchell J, Bradley C. Measuring quality of life in macular degeneration. In: Preedy VR, Watson RR (eds). *Handbook of Disease Burdens and Quality of Life Measures*. Springer: New York, 2009, pp 2633–2648.
- 50 Campochiaro PA, Heier JS, Feiner L, Gray S, Saroj N, Rundle AC *et al*. Ranibizumab for macular edema following branch retinal vein occlusion. *Ophthalmology* 2010; **117**(6): 1102–1112.
- 51 Varma R, Bressler NM, Suñer I, Lee P, Dolan CM, Ward J *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.
- 52 Brown DM, Campochiaro PA, Singh RP, Li Z, Gray S, Saroj N *et al*. Ranibizumab for macular edema following central retinal vein occlusion: six-month primary end point results of a phase III study. *Ophthalmology* 2010; **117**(6): 1124–1133.e1.
- 53 Brown DM, Heier JS, Clark WL, Boyer DS, Vittori R, Berliner AJ *et al*. Intravitreal aflibercept injection for macular edema secondary to central retinal vein occlusion: 1-year results from the phase 3 COPERNICUS study. *Am J Ophthalmol* 2013; **155**(3): 429–437.e7.
- 54 Mitchell P, Bandello F, Schmidt-Erfurth U, Lang GE, Massin P, Schlingemann RO *et al*. The RESTORE study: ranibizumab monotherapy or combined with laser versus laser monotherapy for diabetic macular edema. *Ophthalmology* 2011; **118**(4): 615–625.
- 55 Dick AD, Tugal-Tutkun I, Foster S, Zierhut M, Melissa Liew SH, Bezlyak V *et al*. Secukinumab in the treatment of noninfectious uveitis: results of three randomized, controlled clinical trials. *Ophthalmology* 2013; **120**(4): 777–787.
- 56 Lightman S, Belfort R, Naik RK, Lowder C, Foster CS, Rentz AM *et al*. Vision-related functioning outcomes of dexamethasone intravitreal implant in noninfectious intermediate or posterior uveitis. *Invest Ophthalmol Vis Sci* 2013; **54**(7): 4864–4870.
- 57 Lowder C. Dexamethasone intravitreal implant for noninfectious intermediate or posterior uveitis. *Arch Ophthalmol* 2011; **129**(5): 545.
- 58 Frick KD, Drye LT, Kempen JH, Dunn JP, Holland GN, Latkany P *et al*. Associations among visual acuity and vision- and health-related quality of life among patients in the multicenter uveitis steroid treatment trial. *Invest Ophthalmol Vis Sci* 2012; **53**(3): 1169–1176.
- 59 Mangione CM, Lee PP, Pitts J, Gutierrez P, Berry S, Hays RD. Psychometric properties of the national eye institute visual function questionnaire (NEI-VFQ). *Arch Ophthalmol* 1998; **116**(11): 1496.
- 60 Mangione CM, Lee PP, Gutierrez PR, Spritzer K, Berry S, Hays RD. Development of the 25-item National Eye Institute Visual Function Questionnaire. *Arch Ophthalmol* 2001; **119**(7): 1050–1058.
- 61 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.