Eye (2011) 25, 555–577

www.nature.com/eye

REVIEW

Patient-reported outcomes (PRO's) in glaucoma: a systematic review

S Vandenbroeck[1], S De Geest[1,2], T Zeyen[3], I Stalmans[3] and F Dobbels[1,4]

Abstract

The aim of this review was to summarize literature in view of patient-reported outcome (PRO) instruments for glaucoma and provide guidance on how outcomes are best assessed based on evidence about their content and validity. A systematic literature review was performed on papers describing the developmental process and/or psychometric properties of glaucoma or vision-specific PRO-instruments. Each of them was assessed on their adherence to a framework of quality criteria. Fifty-three articles were identified addressing 27 PRO-instruments. In all, 18 PRO's were developed for glaucoma and 9 for diverse ophthalmologic conditions. Seven instruments addressed functional status, 11 instruments quality of life and 9 instruments disease and treatment-related factors. Most of the instruments demonstrated only partially adherence to predefined quality standards. The tools for assessing functional status were of poor quality, while the Glaucoma Quality of Life Questionnaire and the Vision Quality of Life Index were well-developed QoL measures, yet only validated using classical techniques. The Rasch-scaled QoL-tools, IVI and VCM1 need to improve their item-content for glaucoma patients. The questionnaires to measure adherence should improve their validity and the Treatment Satisfaction Survey for Intra Ocular Pressure pops out as the highest quality tool for measuring topical treatment side effects. This review revealed that most PRO-instruments demonstrated poor developmental quality, more specifically a lack of conceptual framework and item generation strategies not involving the patients' perspective. Psychometric characteristics were mostly tested using classical validation techniques.

Eye (2011) **25,** 555–577; doi:10.1038/eye.2011.45; published online 18 March 2011

Keywords: glaucoma; patient-reported outcomes; development; validation

Introduction

Glaucoma is one of the leading causes of irreversible blindness, with nearly 70 million people suffering from this chronic ophthalmologic condition worldwide.[1] Ninety percent of all cases are primary open-angle glaucoma (POAG).[2] POAG is often called 'the silent thief of sight', as in the early disease phase, typically no symptoms are experienced.[2,3] Patients may experience progressive worsening of their vision, initially peripherally (ie, vision outside the center of gaze), but eventually involving the central vision.[4]

Objective endpoints of vision loss, such as the measurement of visual acuity and visual field, may fall short in capturing the impact of glaucoma on the patient's daily life.[5] The reduced vision is a debilitating condition substantially affecting a patient's ability to perform activities that are dependent on peripheral vision or perception of contrast, such as driving, performing household tasks, reading and may have also a great impact on a person's quality of life (QoL).[6] The patient's perspective is therefore important in order to fully understand the impact of glaucoma and its treatment on their functioning and well-being, and should be more integrated in clinical practice and research evaluations because some treatment effects are only known by the patients and are not detectable or interpretable by the health care provider.

The US Food and Drug Administration (FDA) recently recommended the term 'patient-reported outcomes (PRO's)' as an umbrella term covering a broad range of health data reported by the patient.[7,8] PRO self-report questionnaires

[1]Department of Public Health, Centre for Health Services and Nursing Research, Katholieke Universiteit Leuven, Leuven, Belgium

[2]Institute of Nursing Science, University of Basel, Basel, Switzerland

[3]Department of Ophthalmology, University Hospitals Leuven, Leuven, Belgium

[4]Post-doctoral Researcher, FWO, Flanders, Belgium

Correspondence: F Dobbels, Center for Health Services and Nursing Research, Katholieke Universiteit Leuven, Kapucijnenvoer 35/4, Leuven, Vlaams-Brabant B-3000, Belgium. Tel: +321 633 6981; Fax: +321 633 6970. E-mail: Fabienne.Dobbels@med.kuleuven.be

Received: 30 June 2010
Accepted in revised form: 9 February 2011
Published online: 18 March 2011

have been developed to assess several aspects of the patients' health status, for example, the patients' perception of side effects, the functional impact of illness, the impact of illness on QoL, treatment satisfaction and adherence.[9] Although generic PRO instruments capture a broad range of health status aspects, allowing comparisons among different diseases, they do not capture the patient's perception on specific aspects of a disease or health problem, such as glaucoma. Disease-specific instruments are more sensitive to capture small changes in the condition-specific health status, and may help to interpret and capture clinical outcomes of glaucoma or its treatment comprehensively if well developed and validated. It is also likely that they are more acceptable for the patient than generic instruments, because of their clear relevance to the patient's situation.[10] PRO's are therefore a unique indicator of the disease's impact on a patient's life and are essential for evaluating treatment efficacy or side effects. Hence, instruments measuring PRO's may provide essential disease and treatment information and their results can be considered as a key-element in treatment decision making and research.[9]

However, it may be challenging for clinicians or researchers to evaluate which PRO's are most appropriate for their intended clinical evaluation or research project. Clinicians may benefit from a clear and comprehensive overview of the quality of existing glaucoma-specific PRO's. The aim of this systematic review is therefore to summarize the literature in view of PRO instruments for glaucoma and to provide guidance on how specific outcomes are best assessed based on published evidence about their content and validity.

Materials and methods

Search strategy

The databases PUBMED, CINAHL, Psycinfo and Embase were systematically searched (from 01-01-1980 to 31-12-2010) for relevant articles using the following search string (glaucoma OR ocular hypertens* OR visual impairment OR vision impairment) AND (adheren* OR nonadheren* OR non-adheren* OR complian* OR noncomplian* OR non-complian* OR persistenc* OR impression OR well-being OR mobility OR utility OR preference OR ADL OR symptom* OR activities of daily living OR satisfaction OR pain OR performance status OR disability OR functional status OR quality of life OR health status OR patient based OR self-report OR patient report OR patient related OR patient-reported outcome OR PRO OR score OR questionnaire OR scale OR measure OR instrument) AND (valid* OR reliable OR reliability OR psychomet* OR test–retest OR acceptability OR reproducibility OR sensitivity OR effect size OR responsive*). Next, the reference lists of the selected publications were hand searched for additional relevant articles. Finally, the names of the instruments described in the selected publications as well as the names of the first authors were used as separate search terms.

Study selection criteria

Inclusion- and exclusion criteria

Full-text papers written in English were included if they described the developmental process and/or psychometric properties of a glaucoma-specific self-report instrument capturing a PRO. Vision-specific instruments, developed for a broad range of ophthalmic conditions including glaucoma, and generic instruments adapted for use in glaucoma patients specifically, were also selected for review. Additionally, if glaucoma instruments were further validated in other eye disease populations (eg, cataract), relevant validation data that is of importance for the field of glaucoma (ie, testing of unidimensionality) was integrated in this review.

Papers were excluded if: (1) the instruments were only used in studies, without reporting information on their development or validation; (2) the instruments were developed to assess the need for or the effect of vision-related rehabilitation services; (3) the instruments were specifically developed for children; (4) existing PRO-instruments were translated to another language or adapted for a specific population, (5) the instruments were developed for use in a specific minority population (eg, the population of a developing country), (6) only a subset of items of an instrument was further validated or (7) if a specific scoring algorithm or item response theory was tested in an already existing PRO-instrument without further validating the tool.

Study selection

The researcher scrutinized the titles and abstracts of all identified citations (see Figure 1). The full text was obtained of any article that was deemed potentially eligible by the reviewer. The full text of all retrieved papers was then evaluated on its eligibility based on the previously mentioned inclusion- and exclusion criteria. It should be noticed that for the purpose of this review, we took the most recent paper addressing the development or validity of a selected instrument or its revised version into consideration. Yet, additional information on its development and validity described in previously published papers on the selected tool was integrated.

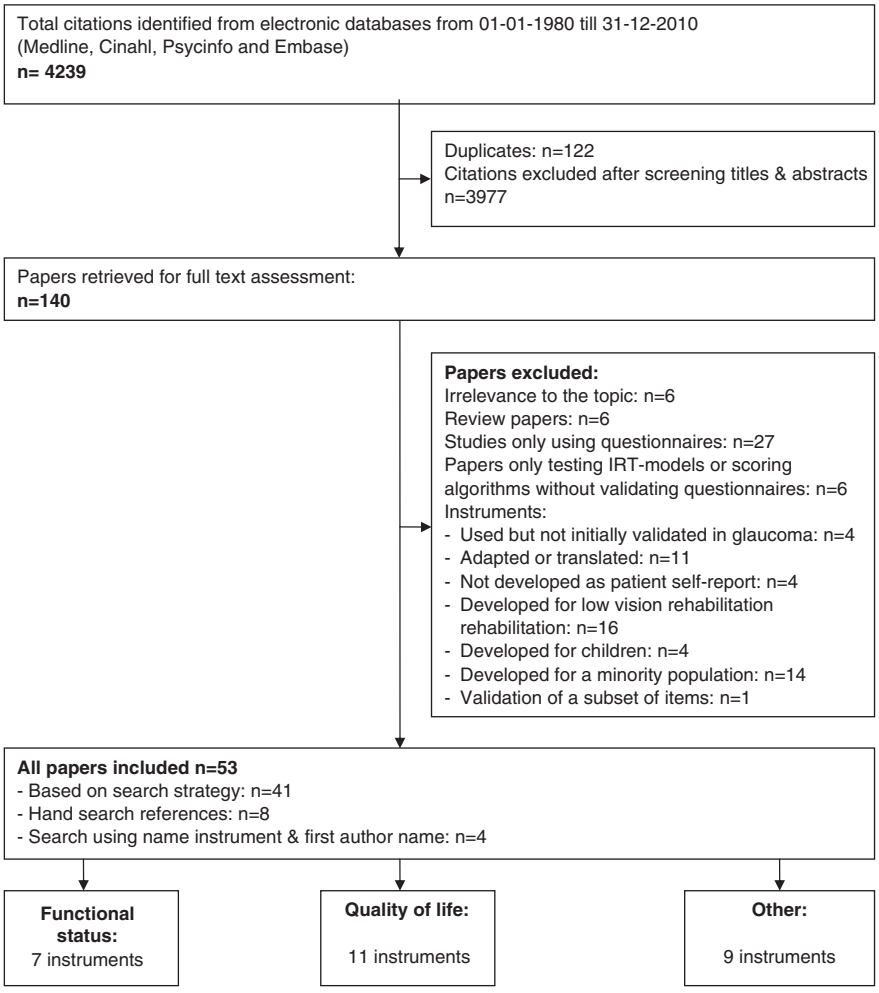

Figure 1 Flow from electronic database searches to final inclusion of eligible studies.

Data extraction, classification and evaluation

The retrieved instruments were classified according to the conceptual framework of Acquadro *et al*,[9] distinguishing PRO's on the perceptions of symptoms, functional impact of illness, utility and preference measures for treatment options, QoL and well-being, patient and treatment satisfaction and adherence.

Next, all the selected and classified PRO's were assessed on their quality using both the FDA-guidelines[11] and the framework of Pesudovs *et al*.[12] The FDA-criteria were applied, as they were specifically developed for PRO-instruments used for supporting medication labeling claims. The latter quality assessment tool was specifically created to determine if existing instruments are adequate for their intended use in the intended target population.[12] These outlined quality criteria emphasize the importance of both the developmental history and the psychometric characteristics of PRO's and put forward the more modern methods of scoring and validation, namely Rasch-analysis.[13] This is a validation technique gaining more and more attention in the ophthalmic literature and transforms ordinal scores into interval scores to strengthen the instruments' content and validity. More specifically, all PRO instruments were evaluated against the following criteria: (1) were the purpose of the instrument and its target population well defined; (2) were adequate steps taken in defining the content of the instrument, the rating scale and the scoring system and (3) is the instrument performing well in view of validity and reliability.[12] There are several existing guidelines and published standards for evaluating and judging these psychometric properties of PRO-instruments,[14–19] but ideally good PRO-instruments require scientific evidence

concerning: construct-, criterion-validity, responsiveness and reliability.[20,21] According to the FDA-guidelines validity, reliability and responsiveness testing should be repeated when a PRO instrument is modified: (a) to measure another concept; (b) to be used in a different population or condition, (c) changing the item content or instrument format, or (d) in terms of mode of administration, culture or language application.[11] A more detailed overview of the definitions of the above mentioned quality criteria including the required psychometric characteristics is given in Table 1.

Results

The search strategy yielded a total of 53 articles addressing 27 PRO's (Figure 1). In all, 18 instruments were specifically developed for patients with glaucoma, and 9 instruments for use in patients with diverse

Table 1 Quality assessment tool for evaluation of PRO-instruments[12]

Property	*Definition*		*Quality criteria*
Instrument development			
Pre-study hypothesis	The pre-study specification of the aim of the instrument and the intended population	√√	A clear description of the aim of the instrument and the intended population
		√	Only one of the above
		X	Neither reported
Intended population	The extent to which the instrument has been studied in the intended population	√√	Intended population studied
		√	Partly studied, or sample size was small (<50 patients)
		X	Not studied in the intended population, only generic
Actual content area	The extent to which the content meets the pre-study hypothesis specifications	√√	Content as intended, and relevant to the intended population
		√	Some of the intended content areas are missing
		X	Content area not relevant to the intended population
Conceptual definition/ framework	The extent to which a conceptual definition/ framework was provided of the concept of interest[11]	√√	A conceptual definition/framework was provided
		X	No conceptual definition/framework was provided
Item identification	Selection of the items relevant to the target population for inclusion in the pilot questionnaire	√√	Comprehensive consulting with patients (focus groups or in-depth interviews) and a literature review
		√	Minimal consultation with patients and expert opinion and literature review
		X	No consultation with patients
Item selection	Determining the items included in the final instrument	√√	A pilot instrument was developed and tested with Rasch- or factor-analysis and statistical justification provided for removing items, plus items with floor and ceiling effects removed and the amount of missing data considered
		√	Only some of the above techniques were used
		X	No pilot instrument OR no statistical justification of items included in the final instrument
Unidimensionality	Demonstration that all items fit with a single underlying construct	√√	Rasch-analysis using fit statistics (0.7–1.3) or item–trait interaction or factor analysis on Rasch scores (first factor loadings >0.4 for all items)
		√	Rasch fit statistics mostly within 0.7 to 1.3 range but some less well fitting items retained, or Chronbach's alpha >0.7, and <0.9 or factor analysis on raw scores (first factor loadings >0.4 for all items)
		X	Rasch-analysis does not support unidimensionality or factor analysis does not support unidimensionality or Chronbach's alpha <0.7 or >0.9

Table 1 (Continued)

Property	*Definition*		*Quality criteria*
Item-person targeting	The extent to which the item-difficulty is targeted to the person-ability[12]	√√	The difference between means of the distribution of items and persons, ≤ 0.5 logits
		X	The difference between means of the distribution of items and persons, >0.5 logits
Response scale	Categories used to rate the items	√√	Statistically justified scale without significant missing data, floor and ceiling effects, and a demonstration of ordered thresholds on Rasch-analysis
		√	Some, but not all of above
		X	Methods for determining response scale not justified statistically
Scoring	A description of how the instrument should be scored	√√	Rasch-scoring of statistically justified response scale
		√	Summary scoring of statistically justified response scale
		X	Scoring system not described or scoring of a statistically unjustified or faulty scale
Psychometric evaluation			
Criterion validity			
Concurrent validity	The extent to which the new instrument correlates with scores of another measure of the same construct or with a highly related construct that is measured concurrently in the same subject[21]	√√	Tested against appropriate measure, correlates between 0.3 and 0.9
		√	Debatable choice of measure, but correlation between 0.3 and 0.9
		X	Tested and correlates <0.3 and >0.9
Predictive validity	The extent to which the test is able to accurately predict the criterion which is evaluated[21]	√√	Tested against appropriate measure, P <0.05
		√	Debatable choice of measure, but P <0.05
		X	Tested and P>0.05
Construct validity			
Convergent	The extent to which the new measure correlates with measures that should be theoretically related to each other[60]	√√	Tested against appropriate measure, correlates between 0.3 and 0.9
		√	Debatable choice of measure, but correlation between 0.3 and 0.9
		X	Tested and correlates <0.3 and >0.9
Discriminant	The extent to which the new measure does not correlate with measures of attributes that are different from the attribute the measure is intended to assess[60]	√√	Tested against appropriate measure, correlates <0.3
		√	Debatable choice of measure, but correlation <0.3
		X	Tested and correlates >0.3
Group differences	The extent to which the new measure demonstrate significant differences between groups who are known to differ on that specific construct[60]	√√	Tested between appropriate groups, and significant differences between groups
		√	Debatable choice of groups, but significant differences between groups
		X	Tested and nonsignificant difference between groups
DIF	The extent to which items have different meanings for different groups[13]	√√	The instrument is free of DIF, with an item estimate difference ≤0.5 logits
		X	The instrument demonstrates DIF, with an item estimate difference > 0.5 logits
Reliability			
Internal consistency	The extent to which all items of the new measure are measuring the same construct[21]	√√	Chronbach's alpha >0.70 and <0.90
		X	Chronbach's alpha <0.70 or >0.90
T–R agreement	The extent to which the results are repeatable when taken by the same observer	√√	LOA appear tight and less than MID, or weighted kappa or ICC >0.8 (T–R) or 0.70 (int)
		√	LOA broader but still close to MID, or weighted kappa or ICC 0.60 to 0.79 (T–R) or 0.50 to 0.69 (int)

Table 1 (Continued)

Property	*Definition*		*Quality criteria*
Interobserver agreement/ intermode (int) agreement	The extent to which the results are repeatable between observers/ between modes of administration	X	LOA ≫MID, weighted kappa or ICC <0.60 (T–R) or 0.50 (int) or incorrect statistical test or inadequate sample (<30 subjects)
Person and item separation reliability	A Rasch-analysis indication of reliability—the proportion of true variance in the observed variance	√√	Reliability of ≥0.8 for both person and item separation or a G-value or separation ratio >2
		√	Only one of person or item separation of ≥ 0.8 or a G-value or separation ratio >2
		X	Person or item separation of <0.8, or a G-value or separation ratio <2
		0	Not reported (not a Rasch-scaled measure)
Responsiveness	The extent to which the instrument van detect clinically important changes over time	√√	Score changes >MID for measures of progression over time or changes with intervention. Effect size or responsiveness statistic given
		√	Changes over time but relationship to MID not reported, small sample, and inadequate time frame
		X	Score changes ≤MID
Interpretation	The extent to which score differences are meaningful	√√	Normative data (ie, mean scores and SD) and MID given for a representative target population and test population demographic reported
		√	MID or normative data or demographic details of study populations, or *ad hoc* population
		X	No normative data and no MID

Abbreviations: DIF, differential item functioning; ICC, intraclass correlation; MID, minimal important difference; LOA, limits of agreement; SD, standard deviation; T–R, test–retest; √√, positive rating; √, minimal acceptable rating; X, negative rating; 0, not reported.
Reproduced with permission from: Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci* 2007; 84: 663–674. (c) The American Academy of Optometry 2007.

ophthalmologic conditions, including glaucoma. Three major categories of PRO's were distinguished, more specifically PRO's addressing functional status related to vision ($n=7$); overall QoL ($n=11$) and other factors related to disease and treatment (eg, symptoms, side effects, adherence, satisfaction, self-efficacy) ($n=9$).

PRO's addressing functional status related to vision

In all, 7 out of the 27 retrieved instruments were developed to assess functional status (Table 2). Functional status refers to the person's ability to undertake activities designed to meet basic needs, fulfill life roles, and maintain health and well-being.[22]

All the selected functional status instruments contain a set of items referring to visual activities, which have to be rated by the patient as being difficult or problematic. The Visual Activities Questionnaire[23] slightly differs from the other instruments as it contains more vision-related items (eg, reading small print), compared with the other six instruments, which are focusing more on important mobility situations in daily life (eg, walking at night). Only two of the selected instruments were validated according to modern validation standards, referred to as Rasch-analysis.

Table 2 shows that the instruments assessing functional status, demonstrated poor quality with regard to its developmental process, as none of the questionnaires used a conceptual framework or comprehensive consultation with patients to guide their item generation process. The selection of items for the final questionnaire were mostly not or only partially supported by adequate statistical techniques, such as factor analysis or Rasch-analysis and only one of the rating scales was statistically justified (ie, Glaucoma Symptom Identifier (GSI)). In view of validation, only the Independent Mobility Questionnaire (IMQ) and GSI were tested using appropriate Rasch-analysis, demonstrating convincing validity evidence of a Rasch-scaled glaucoma measure.[24–26] The GQL-15, validated according to the more classical standards, demonstrated satisfactory validity and reliability evidence. Yet, the GQL-15, as opposed to what it intends to measure, does not assess QoL. QoL is a multi-dimensional concept, yet the GQL-15 only contains items representing visual activities, which is only one dimension of QoL.[27,28] The same is true for the GSI, pretending to measure the impact of glaucoma symptoms on QoL, yet mostly containing items related to visual activities as well.[26]

Table 2 Patient-reported outcomes addressing functional status related to vision (see Table for an explanation of the criteria and their rating)

Instrument description		*Instrument development*	*Psychometric evaluation*		
1. Concept *2. Vision/glaucoma specific* *3. Type of assessment*	*1. Items (subscales)* *2. Rating scale* *3. Interpretation scores*	*Content*	*Validity*	*Reliability*	*Other important indicators*
Visual activity questionnaire (VAQ)[a,23,61]					
1. Performance in daily visual activities 2. Vision specific 3. Written	1. 33 items/visual activities (8 factors): peripheral vision (5); acuity/spatial vision (4); visual search (5); depth perception (3); color discrimination (3); light–dark adaptation (4); glare disability (3); visual processing speed (6) 2. Five-point likert scale 1: never having problems with visual activity 5: always having problems with visual activity 3. Subscale score: mean score for each visual function (range: 1–5)	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition 0 Item identification X Item selection √ Unidimensionality √ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent X Predictive 0 Construct Convergent 0 Discriminant 0 Group differences 0 Differential item functioning 0	Test–retest reliability 0 Internal consistency √√ Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation X
Questionnaire of Ross et al[62]					
1. Perceived visual disability 2. Glaucoma specific 3. Written or interview	1. 16 items/ visual activities (4 factors): navigation out of doors; near vision; navigation at night; vision when cooking 2. Five-point likert scale 1: no disability 5: severe disability 3. Not reported	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition 0 Item identification 0 Item selection √ Unidimensionality √ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent √√ Predictive 0 Construct Convergent 0 Discriminant 0 Group differences 0 Differential item functioning 0	Test–retest reliability √√ Internal consistency 0 Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √
Questionnaire of Mills and Drance[63]					
1. Perceived visual disability 2. Glaucoma specific 3. Written	1. 15 items/questions relating to visual disability 2. 3 answer possibilities: 1: no 2: uncertain 3: yes 3. Not reported	Pre-study hypothesis X Intended population √√ Content area √√ Conceptual framework/ definition 0 Item identification √ Item selection X Unidimensionality 0 Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent √√ Predictive 0 Construct Convergent 0 Discriminant 0 Group differences 0 Differential item functioning 0	Test–retest reliability X Internal consistency 0 Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √
Viswanathan et al[64,65]					
1. Impact on function and activities 2. Glaucoma specific 3. Written	1. 10 items/questions relating to visual disability. 2. 2 answering possibilities yes or no 3. Not reported	Pre-study hypothesis √ Intended population √√ Content area √√ Conceptual framework/ definition 0 Item identification X Item selection X Unidimensionality 0 Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent √√ Predictive 0 Construct Convergent 0 Discriminant 0 Group differences √√ Differential item functioning 0	Test–retest reliability 0 Internal consistency 0 Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation X

Table 2 (Continued)

Instrument description		*Instrument development*	*Psychometric evaluation*		
1. Concept *2. Vision/glaucoma specific* *3. Type of assessment*	*1. Items (subscales)* *2. Rating scale* *3. Interpretation scores*	*Content*	*Validity*	*Reliability*	*Other important indicators*
Glaucoma quality of life questionnaire (GQL-15)[27,28]					
1. Perceived visual disability in daily tasks 2. Glaucoma specific 3. Written	1. 15 items/ visual activities (4 factors): central and near vision (2); peripheral vision (6); glare and dark adaptation (6); outdoor mobility (1) 2. Five-point likert scale 0: do not perform for non-visual reasons 1: no difficulty with visual activity 5: severe difficulty with visual activity 3. Total score: (range: 0–100)	Pre-study hypothesis √√ Intended population √√ Content area √ Conceptual framework/ definition 0 Item identification X Item selection √ Unidimensionality √ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent √√ Predictive 0 Construct Convergent 0 Discriminant 0 Group differences √√ Differential item functioning 0	Test–retest reliability √√ Internal consistency √√ Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √
Independent mobility questionnaire[24,25]					
1. Perceived visual disability for independent mobility 2. Glaucoma specific 3. Written	1. 35 items/mobility situations 2. Five-point likert scale 1: no difficulty with mobility situation 5: extreme difficulty with mobility Situation 3. Rasch-scaled	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition 0 Item identification X Item selection X Unidimensionality √ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent X Predictive 0 Construct Convergent 0 Discriminant 0 Group differences √√ Differential item functioning 0	Test–retest reliability 0 Internal consistency 0 Inter rater reliability 0 Person and item separation reliability √√	Responsiveness 0 Interpretation X
Glaucoma symptom identifier[26]					
1. Impact of glaucoma symptoms on quality of life 2. Glaucoma specific 3. Written	1. 32 items/visual activities (1 factor) 2. Three-point likert scale 1: none or do not do this for nonvisual reasons 3: yes or I no longer do this for visual reasons 3. Rasch scaled	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition √ Item identification √ Item selection √ Unidimensionality √ Item-person targeting 0 Response scale √√ Scoring √√	Criterion Concurrent 0 Predictive 0 Construct Convergent √√ Discriminant 0 Group differences √√ Differential item functioning 0	Test–retest reliability 0 Internal consistency X Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √

[a]Gothwall *et al*[61] proposed a 13-item Rasch-analyzed version of the VAQ validated in a cataract population demonstrating unidimensionality, absence of DIF, good person-separation reliability, yet poor item-person targeting.

PRO's addressing QoL

QoL is a multi-dimensional concept[29] referring to the degree of overall life satisfaction that is positively or negatively influenced by individuals' perception of certain aspects of life important to them (Table 3), including matters both related and unrelated to health.[30,31]

The literature search yielded 11 QoL-instruments developed for patients either with a visual impairment including glaucoma ($n=3$) or for patients suffering from glaucoma specifically ($n=8$). Four QoL-instruments were analyzed or revised using Rasch-analysis and seven instruments according to the classical validation techniques.

The process of instrument development was very extensive in most of the classically tested QoL-PRO's with almost all instruments using the patients' input in view of item generation except for the GHPI[32,33] and the QoL–VFQ.[34] The latter instruments should therefore not

be considered to measure QoL in glaucoma patients because of their poor quality regarding their item content. The majority of remaining questionnaires demonstrated acceptable to good item selection procedures and were tested on unidimensionality except for the original and widely used NEI-VFQ 51 items[35,36] and 25 items[37] questionnaires, the GUI[38] and the NHVQoL.[39,40] Yet, this latter criterion is fundamental to test if all items tap the same underlying construct in order to be able to calculate valid (sub) scale scores. Therefore, Marella *et al*[41] tested the NEI-VFQ-25 on its dimensionality using Rasch-analysis, which resulted in two factors (ie, visual functioning and socioemotional traits) and hence could not confirm the 12 domains from the original form. These analyses resulted in statistically ordered response scale-thresholds and validity evidence. From the NEI-VFQ 51, a selection of 27 items was used to further validate the questionnaire, resulting in a unidimensional 17 item questionnaire with statistically tested rating scales. Yet, only limited validity evidence was provided.[42]

Two types of instruments can be distinguished based on the scoring systems within our selected tools. Seven questionnaires have to be rated on a simple ordinal scale, while two instruments are preference based measures (ie, ViSQoL[43,44] and GUI[38]) where all patients are asked to choose between different health situations (eg, perfect health *vs* worst possible health/death). However, none of them demonstrated statistically justified response scales. As a glaucoma-specific QoL-tool, the Glaucoma QoL Questionnaire (Glau-QoL)[45] demonstrated good developmental characteristics (except statistical evidence for the rating scale) and strong validity evidence, while the VisQoL[43,44] demonstrated high quality scores as a measurement tool across diverse ophthalmic conditions, yet additional Rasch-analysis might be mandatory to strengthen validity evidence of both instruments.

The revised VCM1[46] and IVI[47] based on Rasch-analysis showed both an extensive and high quality development process and a statistically justified rating scale, yet item-person targeting was poor. This means that the selected items were suboptimal for the intended population, possibly requiring adding and/or removal of items.

PRO's addressing other aspects

Nine instruments were developed to assess either topical treatment or disease-related factors, yet only one of them was tested using modern test theories (ie, Rasch-analysis) (Table 4). Five instruments assess frequency of and perceived distress related to side effects, satisfaction with eye drop treatment, adherence to eye drop treatment or a combination of these aspects. Two instruments focus on symptoms of glaucoma, while two other tools were developed to assess self-efficacy and outcome expectation.

Both the Treatment Satisfaction Survey for Intra Ocular Pressure (TSS-IOP) and COMTOL assess side effects and satisfaction with glaucoma treatment, yet the COMTOL-questionnaire was only validated in patients treated with pilocarpine or timolol, meaning that not all the instrument domains could be adequately psychometrically evaluated. Yet, if studies aim to compare different eye drops, the TSS-IOP should be chosen as it shows acceptable reliability and good validity across all eye drop classes. Except for side effects, the content of both instruments differs, as the TSS-IOP addresses satisfaction and bothersomeness with factors related to eye drops (eg, eye drop effectiveness) and the COMTOL questions activity limitations (ie, driving) because of eye drops as well as the impact of side effects and activity limitations on QoL. Compared with the COMTOL, the TSS-IOP demonstrated both a higher quality developmental process in view of identifying and selecting items and showed better validity evidence.[48–50] The Glaucoma Satisfaction Questionnaire (Glausat) was created to primarily assess patient satisfaction with eye drop treatment. Besides containing items addressing side effects and general treatment satisfaction, the Glausat also contains items describing the 'ease of use', 'efficacy', 'expectations and beliefs about treatment', 'impact on HRQoL', 'medical care' and 'general satisfaction'. Its developmental strategy was satisfactory, yet validation evidence is limited requiring further improvement of the instrument.[51]

The adherence instrument of Schwartz *et al* (2009) and the EDSQ questionnaire primarily focused on adherence with eye drop treatment and both their developmental process was theory driven. Only the EDSQ used patient input to generate the items in order to strengthen its content. Yet, both instruments demonstrated significant pitfalls in view of validity. First, the adherence questionnaire of Schwartz *et al* (2009) showed an inadequate item selection process, a non-statistically justified rating scale and provided poor validity evidence.[52,53] Second the EDSQ, however, well developed was not able to significantly discriminate between patients with different adherence-profiles. Further adaptations and validation of both instruments seem necessary, preferably by means of modern psychometric techniques.

Both the Symptom Impact Glaucoma (SIG) and Glaucoma Symptom Scale (GSS) address visual as well as non-visual symptoms, referring to problems related to the disease process (eg, difficulties with seeing in the dark) and problems directly caused by the topical treatment (eg, red eyes), respectively. The SIG was tested using conventional validity tests, while the GSS

Table 3 Patient-reported outcomes addressing quality of life (see Table for an explanation of the criteria and their rating)

Instrument description		*Instrument development*	*Psychometric evaluation*		
1. Concept *2. Vision/glaucoma specific* *3. Type of assessment*	*1. Items (subscales)* *2. Rating scale* *3. Interpretation scores*	*Content*	*Validity*	*Reliability*	*Other important indicators*
Glaucoma quality of life questionnaire (Glau-QoL)[45]					
1. HRQoL effects in patients with glaucoma and ocular hypertension 2. Glaucoma specific 3. Written	1. 36 items (7 components): psychological well-being (6); self-image (5); daily life (9); burden of treatment (5); driving (3); anxiety (4) and confidence in health care (4) 2. Four- or five-point likert scale 3. Subscale score: summing of the results of the items and transformed into a scale from 0 to 100 0: poor HRQoL 100: good HRQoL	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition √√ Item identification √√ Item selection √√ Unidimensionality √ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent 0 Predictive 0 Construct Convergent √√ Discriminant √√ Group differences √√ Differential item functioning 0	Test–retest reliability √ Internal consistency √ Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √
Vision quality of life index[43,44]					
1. Vision and quality of life-related utility measure 2. Vision specific 3. Interview	1. Six items representing six dimensions: physical well-being, independence, social well-being, emotional well-being, self-actualization and planning and organization. 2. Physical well-being: 1 (most unlikely)-5 (certainly) Independence and self-actualization: 1 (no effect)-6 (unable) Social well-being: 1 (easier)-6 (unable) 7 (not applicable) Planning and organization: 1 (no difficulty)-5 (unable) 6 (not applicable) Emotional well-being: 1 (more confident)-6 (not confident at all) 3. Used as summative instrument: summed scores (range: 0–28) 0: enhances QoL 1–4: no or little impact on QoL >4: greater impact on QoL Used as utility instrument: Time trade off procedure (ie, how many years with a particular health condition are patients willing to trade to have perfect health): A. For key parameters: Item worst responses (worst level of response category for each item, with all other items at their best level) All worst responses (all items at their worst level) B. For 28-intermediate item responses: Item levels when items are not at their worst level	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition √√ Item identification √√ Item selection √√ Unidimensionality √√ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent 0 Predictive 0 Construct Convergent 0 Discriminant 0 Group differences √√ Differential item functioning 0	Test–retest reliability 0 Internal consistency √√ Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √
Glaucoma health perception index[32,33]					
1. Perceived impact of glaucoma	1. Six items: Four questions: the extent to which patients perceive that glaucoma and treatment	Pre-study hypothesis √√ Intended population √√	Criterion Concurrent 0 Predictive 0 Construct	Test–retest reliability √√ Internal consistency √ Inter rater reliability 0	Responsiveness 0 Interpretation √

Table 3 (Continued)

Instrument description		*Instrument development*	*Psychometric evaluation*		
1. Concept *2. Vision/glaucoma specific* *3. Type of assessment*	*1. Items (subscales)* *2. Rating scale* *3. Interpretation scores*	*Content*	*Validity*	*Reliability*	*Other important indicators*
2. Glaucoma specific 3. Interview	interfere with physical, emotional, social and cognitive components of their health. Two questions: perception of the amount of stress attributable to glaucoma and treatment and concern about blindness. 2. Five-point likert scale 1: not at all interfering 5: a lot interfering 3. Total score: (range: 1–5) Interpretation not reported	Content area √√ Conceptual framework/ definition √√ Item identification X Item selection X Unidimensionality 0 Item-person targeting 0 Response scale X Scoring X	Convergent √√ Discriminant 0 Group differences 0 Differential item functioning 0	Person and item separation reliability 0	
National eye institute visual function index-51 items[a,35,36,42]					
1. Impact of visual impairment on HRQoL 2. Vision specific 3. Written or interview	1. 51 items (13 domains): general health (2); general vision (2); ocular pain (2); near vision (7); distance vision (7); social functioning (4); mental health (8); expectations (3); role functioning (5); dependency (5), driving (4); peripheral vision (1); color vision (1) 2. General health Five-point likert scale 1: excellent 5: poor + 0-10 health rating General vision Six-point likert scale 1: excellent 6: completely blind + 0-10 vision rating 11 Multi-item scales assessing difficulty with all other domains and a single items scale to assess limitations with peripheral and color vision 3. Subscale score: average of subscale items transformed to a 0 to 100 scale Total score: Cfr subscale score (0: worst possible score; 100: best possible score)	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition 0 Item identification √√ Item selection X Unidimensionality X Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent √√ Predictive 0 Construct Convergent √√ Discriminant 0 Group differences √√ Differential item functioning 0	Test–retest reliability √ Internal consistency √√ Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √
National eye institute visual function index-19 items[b37, 1,56]					
1. Impact of visual impairment on HRQoL 2. Vision specific 3. Written or interview	1. 19 items (2 factors): Visual functioning (10), Socioemotional traits (9) 2. Not reported 3. Rasch-scaled	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/definition 0 Item identification √√ Item selection √√ Unidimensionality √√ Item-person targeting √√ Response scale √√ Scoring √√	Criterion Concurrent √√ Predictive 0 Construct Convergent √√ Discriminant 0 Group differences √√ Differential item functioning √√	Test–retest reliability 0 Internal consistency √√ Inter rater reliability 0 Person and item separation reliability √	Responsiveness 0 Interpretation √
Nursing home vision quality of life questionnaire[39,40 66]					
1. Vision targeted health-related quality of life in older adults residing in nursing	1. 57 items (9 domains): Reading (3); ocular symptoms (9); general vision (6); ADL (6); mobility (7); social activities/hobbies (8); psychological distress (10); adaptation/ coping (2);	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition 0 Item identification √√	Criterion Concurrent √√ Predictive 0 Construct Convergent √√ Discriminant √√	Test–retest reliability √ Internal consistency √√ Inter rater reliability 0 Person and item separation reliability √	Responsiveness 0 Interpretation √

Table 3 (Continued)

Instrument description		*Instrument development*	*Psychometric evaluation*		
1. Concept *2. Vision/glaucoma specific* *3. Type of assessment*	*1. Items (subscales)* *2. Rating scale* *3. Interpretation scores*	*Content*	*Validity*	*Reliability*	*Other important indicators*
homes 2. Vision specific 3. Interview	social interaction (6) 2. General vision Item 1: Scale 0–10: 0: worst possible eyesight 10: best possible eyesight Item 2: Five-point likert scale 1: excellent 5: very poor Ocular symptoms (7) Four-point likert scale 1: not bothered 3: a lot bothered 4: not sure Reading (3)/ADL (6)/mobility (6)/ activities and hobbies (8)/social interaction (4) Five-point likert scale 1: no difficulty at all 4: extreme difficulty 5: stopped doing this because of eyesight General vision (1)/mobility (1)/psychological distress (8)/adaptation/coping(1)/social interaction (1) Five-point likert scale 1: definitely true 5: definitely false Ocular symptoms (2)/general vision (3)/psychological distress (2)/adaptation/coping (1)/social interaction (1) Six-point likert scale 1: none of the time 5: all of the time 6: not sure 3. (a) *Classical validation* Subscale score: average of items in the subscale transformed to a 0 to 100 scale Total score: Cfr subscale score 0: lowest functional level 100: highest functional level (b) *Modern validation* Rasch-scaled	Item selection X Unidimensionality √√ Item-person targeting X Response scale √√ Scoring √√	Group differences 0 Differential item functioning √√		
Glaucoma utility index[38] 1. Preference based glaucoma utility assessment of QoL (functional status, symptoms and side effects) 2. Glaucoma specific 3. Written	1. 32 choice questions: comparing 2 profiles both having a different combination of levels of all the 6 dimensions: Dimensions: central and near vision, lighting and glare, mobility, activities of daily living, eye discomfort Each dimension is rated using four levels: 0: No difficulty with 1: Some difficulty 2: Quite a lot of difficulty 3: Severe difficulty 2. Choice between profile A or B 3. Total score: a quality weight, situated between 0 and 1, is attributed to each possible level of each dimension. The quality weights of all dimensions are summed	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition √√ Item identification √√ Item selection X Unidimensionality 0 Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent 0 Predictive 0 Construct Convergent √√ Discriminant 0 Group differences √√ Differential item functioning 0	Test–retest reliability 0 Internal consistency 0 Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √

Table 3 (Continued)

Instrument description		*Instrument development*	*Psychometric evaluation*		
1. *Concept* 2. *Vision/glaucoma specific* 3. *Type of assessment*	1. *Items (subscales)* 2. *Rating scale* 3. *Interpretation scores*	*Content*	*Validity*	*Reliability*	*Other important indicators*
Low vision quality of life questionnaire[67–70]					
1. Impact of visual impairment on QoL and functional status and the outcome of rehabilitation strategy 2. Vision specific 3. Written	1. 18 items/activities (4 factors): Basic aspects of vision (7); mobility (3); adjustment (3); reading and fine works (5) 2. Six-point likert scale 0: could not be performed because of vision 1: great difficulty because of vision 5: no difficulty because of vision 3. Total score (range: 0–125): 0: low QoL 125: high QoL	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition 0 Item identification √ Item selection √√ Unidimensionality √ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent √√ Predictive 0 Construct Convergent 0 Discriminant 0 Group differences √√ Differential item functioning 0	Test–retest reliability √ Internal consistency √√ Inter rater reliability 0 Person and item separation reliability 0	Responsiveness √ Interpretation √
Quality of life and visual function questionnaire[34]					
1. Visual satisfaction 2. Vision specific 3. Interview	1. 17 items: visual satisfaction (5), visual field (3); distance visual acuity (3); near visual acuity (2); sensory adaptation (3); color vision (1) + questions concerning attitude to health condition (3) 2. 17 items: three-point likert scale 1: not at all 3: very much Additional questions 4 point likert scale Higher score indicating more optimistic views (no details mentioned) 3. Total score: mean of six subscale scores Subscale score: mean of scores of each single question (except general health)	Pre-study hypothesis √√ Intended population √√ Content area √ Conceptual framework/ definition 0 Item identification X Item selection X Unidimensionality √ Item-person targeting X Response scale X Scoring	Criterion Concurrent √√ Predictive 0 Construct Convergent 0 Discriminant 0 Group differences √ Differential item functioning 0	Test–retest reliability 0 Internal consistency √√ Inter rater reliability X Person and item separation reliability 0	Responsiveness 0 Interpretation X
Vision core module 1[c46,71–73]					
1. Impact of visual impairment on QoL 2. Vision specific 3. Interview or written	1. 10 items/perceptions: embarrassment, anger, depression, loneliness, fear of deterioration in vision, safety at home, safety outside the home, coping with everyday life, inability to do preferred activities and life interference. 2. Six-point likert scale 0: not at all 5: all the time 3. (a) *Classical validation*: Total score: summing individual items and dividing by 10 (range: 0–5, with 50 intervals). (b) *Modern validation*: Rasch-scaled	Pre-study hypothesis √√ Intended population √√ Content area √ Conceptual framework/ definition √√ Item identification √√ Item selection √ Unidimensionality √√ Item-person targeting X Response scale √√ Scoring √√	Criterion Concurrent 0 Predictive 0 Construct Convergent 0 Discriminant 0 Group differences 0 Differential item functioning √√	Test–retest reliability 0 Internal consistency √√ Inter rater reliability Person and item separation reliability √	Responsiveness 0 Interpretation √
Impact of vision impairment[47,74–79]					
1. Impact of vision impairment on a person's ability 2. Vision specific	1. 28 items/common daily activities (3 factors): mobility and independence (11), emotional well-being (8), reading and accessing information (9) 2. Four-point likert scale (26 items) 0: not at all	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/	Criterion Concurrent 0 Predictive 0 Construct Convergent 0 Discriminant 0 Group	Test–retest reliability 0 Internal consistency 0 Inter rater reliability 0	Responsiveness 0 Interpretation √

Table 3 (Continued)

Instrument description		*Instrument development*	*Psychometric evaluation*		
1. Concept *2. Vision/glaucoma specific* *3. Type of assessment*	*1. Items (subscales)* *2. Rating scale* *3. Interpretation scores*	*Content*	*Validity*	*Reliability*	*Other important indicators*
3. Written or interview	3: all the time/cannot do it because of eyesight 8: do not do it because of other reasons Three-point likert scale (2 items) 0: not at all 2: all the time/ cannot do it because of eyesight 3. Total and domain score: arithmetic average of the rating of applicable items Rasch-scaled	definition √√ Item identification √√ Item selection √√ Unidimensionality √√ Item-person targeting X Response scale √√ Scoring √√	differences √√ Differential item functioning √√	Person and item separation reliability √	

[a]Massof and Fletcher[42] tested 27 items of the NEI-VFQ-51 yielding 17 items fitting in the Rasch-model with a statistically justified response scale and evidence for convergent validity.
[b]Concurrent, convergent and group differences- validity of the NEI-VFQ was only tested on its original form (ie, NEI-VFQ 25).[37]
[c]DIF was found between two administration modes of the VCM1 (ie, self-report and proxy-report), yet has no substantial impact on the VCM1.
Langelaan *et al*[80] proposed a Rasch-scaled 22-item version of the NEI-VFQ 25 with a four-factor structure (ie, near activities, distance activities and mobility, mental health and dependency, pain and discomfort) with a statistical justified response scale, yet demonstrating inadequate fit to the Rasch-model for three of the four factors. DIF was present for two of the NEI-VFQ-22 items.
Pesudovs *et al*[41] reengineered the NEI-VFQ-25 in a cataract population to a two-factor structure (ie, visual functioning (8 items) and socioemotional traits (10 items)) similar to the instrument presented above all due to lack of unidimensionality in the original instrument. The new instrument shows adequate fit to the Rasch-model, DIF for two items of the visual functioning scale, good person separation reliability and poor item-person targeting.[56]

underwent both classical and modern validity testing. The SIG adheres more to the quality criteria in view of instrument development compared with the GSS, as it is based on a conceptual framework and patients were involved during the item generation process. However, the item-selection and rating scales were not statistically justified for both instruments. The SIG demonstrated poor validity evidence based on classical tests and the Rasch-analysis of the GSS, elucidated poor item-person targeting in a sample of glaucoma patients, requiring further adjustment of the instrument.[32,33,47,54] Both instruments do not seem adequate for assessing the presence and bothersomeness of glaucoma symptoms according to the predefined quality criteria.

Sleath *et al*[55] developed two scales, more specifically one focusing on glaucoma medication self-efficacy and one addressing glaucoma outcome expectations. Self-efficacy refers to the confidence in using the eye drops (eg, overcoming barriers, carrying out specific tasks required to use eye drops correctly). Outcome expectations on the other hand are whether an individual believes that a certain behavior (eg, taking eye drops) will have a positive impact on a health condition (eg, glaucoma). Both tools were developed based on already existing questionnaire with limited involvement of patients. The item selection procedure was based on floor- and ceiling effects and principal component analysis, yet response scales were not statistically tested on disordered thresholds given that the investigators chose the classical approach of validation. Validity evidence is not convincing with only limited evidence on convergent validity for the self-efficacy scale.[55]

Discussion

Objective measures such as visual field defects and visual acuity only provide limited information about the impact of glaucoma and its treatment on patient's daily life. Therefore, integrating the patients' perspective by using PRO-instruments gain more and more attention in clinical studies as well as in clinical practice. In clinical studies for instance, it is no longer sufficient to demonstrate that a new drug is significantly more effective than another drug based on traditional medical endpoints. Other treatment effects coming from the patients such as side effects and tolerability of eye drops, and the impact of a specific eye drop treatment on the QoL are important to capture and should therefore, according to the FDA regulatory agencies, be assessed in a structured and consistent way. Subsequently, PRO's will be increasingly used as relevant endpoint measures as they are: (1) unique indicators of disease impact, (2) essential for evaluating treatment efficacy, (3) useful for interpreting clinical outcomes and (4) a key element in treatment decision making, which should be based on a combination of objective and patient-reported subjective parameters.[9]

Table 4 Patient-reported outcomes addressing side effects, satisfaction and adherence with eye drop treatment and symptoms of glaucoma, self-efficacy and glaucoma outcome expectations (see Table for an explanation of the criteria and their rating)

Instrument description		*Instrument development*	*Psychometric evaluation*		
1. Concept *2. Vision/glaucoma specific* *3. Type of assessment*	*1. Items (subscales)* *2. Rating scale* *3. Interpretation scores*	*Content*	*Validity*	*Reliability*	*Other important indicators*
Treatment satisfaction survey for intraocular pressure[48,49]					
1. Patient satisfaction with topical treatment 2. Glaucoma specific 3. Written	1. 15 items (5 factors): effectiveness (2); hyperemia (3); eye irritation (4); convenience of use (3); ease of use (3) 2. Five-point likert scale 1: extremely satisfied 5: extremely dissatisfied Or Seven-point likert scale 1: extremely bothered 7: not bothered (not clear which scoring for which domain) 3. Total score: equating the scale range of items, adding the scale values of items within a factor and transforming the resulting value into a score between 0 and 100. Higher score = greater satisfaction	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition √√ Item identification √√ Item selection √ Unidimensionality √ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent √√ Predictive 0 Construct Convergent √√ Discriminant 0 Group differences √√ Differential item functioning 0	Test–retest reliability √ Internal consistency √√ Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √
Comparison of ophthalmic medication for tolerability[50]					
1. Frequency and bother of common side effects and its effect on QOL, adherence and satisfaction with the medication 2. Glaucoma specific 3. Interview	1. Frequency of side effects (5 factors): ocular symptoms (7), taste (2), vision difficulties (3), accommodation difficulties (2), brow ache (1) Bothersomeness side effects (5 factors): ocular symptoms (7), taste (2), vision difficulties (3), accommodation difficulties (2), brow ache (1) Limitation living activities (3 factors): driving (2), reading (2), moderate activities (3) Global questions: impact side effects (1) and limitation of activities (1) on QoL, adherence (1), reasons for non-adherence (1), eye drop satisfaction (1) 2. Bothersomeness side effect, limitations of living activities and QoL: Five-point likert scale 1: not at all limited/bothered 5: extremely limited/bothered Frequency side effects and adherence: Six-point likert scale 0: did not experience/did not miss 1: rarely 6: always Reasons for non-adherence: 0: did not take the eye drops because of side effects I experienced 1: I did not take the eye drop	Pre-study hypothesis √√ Intended population √ Content area √√ Conceptual framework/ definition 0 Item identification X Item selection √ Unidimensionality √ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent 0 Predictive 0 Construct Convergent √√ Discriminant 0 Group differences √√ Differential item functioning 0	Test–retest reliability √√ Internal consistency √√ Inter rater reliability 0 Person and item separation reliability 0	Responsiveness √ Interpretation √

Table 4 (Continued)

Instrument description		*Instrument development*	*Psychometric evaluation*		
1. Concept *2. Vision/glaucoma specific* *3. Type of assessment*	*1. Items (subscales)* *2. Rating scale* *3. Interpretation scores*	*Content*	*Validity*	*Reliability*	*Other important indicators*
	for some other reason Satisfaction: six-point likert scale: 0: totally satisfied 5: totally dissatisfied 3. Subscale score: average of the items in the domain (range: 0–5 or 0–6) with higher scores indicating increased discomfort (except for adherence)				
Glausat[51]					
1. Patient satisfaction with glaucoma treatment 2. Glaucoma specific 3. Written	1. 22 items (7 factors): Expectations and beliefs about treatment (3), ease of use (3), efficacy (3), undesired effects (4), impact on HRQoL (3), medical care (3), general satisfaction with treatment (3) 2. Five-point likert scale 1: Strongly agree 5: Strongly disagree Only for undesired effects: 1: Not at all 5: Very much 3. Not reported	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition √√ Item identification √√ Item selection √ Unidimensionality √ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent 0 Predictive 0 Construct Convergent √ Discriminant 0 Group differences 0 Differential item functioning 0	Test–retest reliability 0 Internal consistency √√ Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √
Eye drop satisfaction questionnaire[52]					
1. Patient satisfaction and adherence with eye drop treatment 2. Glaucoma specific 3. Written	1. 21 items (6 components): concerns about treatments (5), concerns about disease (2), satisfaction with patient–clinician relationship (5), positive beliefs (3), treatment convenience (3), self-declared compliance (3) 2. Not reported 3. Subscale score: (range 0–100) Higher score = more of the implied attribute	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition √√ Item identification √√ Item selection √ Unidimensionality √ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent 0 Predictive 0 Construct Convergent 0 Discriminant 0 Group differences X Differential item functioning 0	Test–retest reliability 0 Internal consistency √√ Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √
Adherence questionnaire[53]					
1. Adherence/ readiness for changes 2. Glaucoma specific 3. Written	1. 62 items (6 subscales): frequency and occurrence of thoughts and experiences that can affect the use of eye drops (11), adherence (5), adherence in the context of the transtheoretical model of change (5), side effects (12), reasons for not using eye drops (23), demographics (6). 2. Frequency and occurrence of thoughts and experiences that can affect the use of eye drops and adherence: five-point likert scale 1: never 5: always Adherence Five-point likert scale 1: never 5: always	Pre-study hypothesis √ Intended population √√ Content area √√ Conceptual framework/ definition √√ Item identification X Item selection X Unidimensionality X Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent 0 Predictive 0 Construct Convergent 0 Discriminant 0 Group differences 0 Differential item functioning 0	Test–retest reliability 0 Internal consistency 0 Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √

Table 4 (Continued)

Instrument description		*Instrument development*	*Psychometric evaluation*		
1. Concept *2. Vision/glaucoma specific* *3. Type of assessment*	*1. Items (subscales)* *2. Rating scale* *3. Interpretation scores*	*Content*	*Validity*	*Reliability*	*Other important indicators*
	Transtheoretical model of change (taking and timing adherence): Stages of change: A: no, and I do not plan to start in the next 6 months E: yes, and I have for >6 months Side effects: Check if experienced on a regular basis Reasons for not using eye drops: Check those that explain why you miss a dose of your eye drops or do not use your eye drops 3. Not reported				
Symptom impact of glaucoma scale[32,33]					
1. Impact of symptoms of glaucoma 2. Glaucoma specific 3. Interview	1. 43 items/symptoms (4 domains): visual function symptoms (11); local eye symptoms (7); systemic symptoms (20), psychological symptoms (5) 2. Presence of symptom: Yes or no If yes, due to glaucoma 'Entirely due', 'partially due' or 'not at all due' Bothersomeness 5-point likert scale 1: not at all bothersome 5: very bothersome 3. Total score: adding bothersomeness score only for the symptoms being rated as at least ''partially due to glaucoma'' (range: 0–125) Subscale score: Cfr Total score Visual function (range: 0–55) Local eye (range: 0–35) Systemic (range: 0–100) Psychological (0–25)	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition √√ Item identification √√ Item selection X Unidimensionality 0 Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent √ Predictive 0 Construct Convergent √√ Discriminant 0 Group differences 0 Differential item functioning 0	Test–retest reliability √ Internal consistency √ Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √
Glaucoma symptom scale[47,54]					
1. Symptoms and side effects of glaucoma or eye drop treatment 2. Glaucoma specific 3. Written	1. 10 items/symptoms (2 factors): symptoms of visual nature (4); symptoms of non-visual nature (6) 2. Five-point likert scale (for each eye) 0 = complaint present and very bothersome 4 = complaint absent 3. (a) *Classical validation* Total score: unweighted average of responses of 10 items, averaged between two eyes, transformed to 0 to 100 scale 0: presence of very bothersome problem 100: absence of problem	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition √√ Item identification X Item selection X Unidimensionality √√ Item-person targeting X Response scale √√ Scoring √√	Criterion Concurrent 0 Predictive 0 Construct Convergent 0 Discriminant 0 Group differences X Differential item functioning √√	Test–retest reliability 0 Internal consistency √√ Inter rater reliability 0 Person and item separation reliability √	Responsiveness 0 Interpretation √

Table 4 (Continued)

Instrument description		*Instrument development*	*Psychometric evaluation*		
1. Concept *2. Vision/glaucoma specific* *3. Type of assessment*	*1. Items (subscales)* *2. Rating scale* *3. Interpretation scores*	*Content*	*Validity*	*Reliability*	*Other important indicators*
	Subscale score: Cfr total score (b) *Modern validation* Rasch-scaled				
Glaucoma self-efficacy scale[55]					
1. Medication self-efficacy 2. Glaucoma specific 3. Written	1. 2 scales 21 items (1 factor): self-efficacy in overcoming barriers interfering with the use of glaucoma medications (21) 14 items (2 factors): Confidence in using eye drops in general (8); confidence in the ability to get the eye drops correctly in the eye (6) 2. Three-point likert scale Not at all confident Very confident 3. Not reported	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition √√ Item identification √ Item selection √√ Unidimensionality √ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent 0 Predictive 0 Construct Convergent √√ Discriminant 0 Group differences 0 Differential item functioning 0	Test–retest reliability 0 Internal consistency √√ Inter rater reliability 0 Person and item separation reliability0	Responsiveness 0 Interpretation √
Outcome expectations scale[55]					
1. Outcome expectations 2. Glaucoma specific 3. Written	1. 4 items (1 factor): glaucoma expectations (4) 2. Nine-point likert scale Not at all Extremely 3. Not reported	Pre-study hypothesis √√ Intended population √√ Content area √√ Conceptual framework/ definition √√ Item identification √ Item selection √√ Unidimensionality √ Item-person targeting 0 Response scale X Scoring X	Criterion Concurrent 0 Predictive 0 Construct Convergent X Discriminant 0 Group differences 0 Differential item functioning 0	Test–retest reliability 0 Internal consistency √√ Inter rater reliability 0 Person and item separation reliability 0	Responsiveness 0 Interpretation √

In order to facilitate the choice for PRO instruments with a high quality to be included in future clinical trials, the primary aim of this review was to provide an overview of all existing PRO-instruments developed for glaucoma specifically or for a broad range of visual impairments including glaucoma, as well as to scrutinize their developmental process and psychometric properties by rating these characteristics based on the FDA-guidelines and quality criteria outlined by Pesudovs *et al*.[12] To our knowledge, this is the first literature review addressing all PRO's available for glaucoma patients.

This review demonstrates that PRO instruments exist covering all categories of PRO's as described in the framework of Acquadro *et al*.[9] Yet, most of the PRO-instruments were developed in view of assessing QoL ($n = 11$), followed by seven instruments with a focus on functional status and nine instruments assessing treatment and disease-related factors (ie, side effects, treatment satisfaction, symptoms, adherence). The latter category seems to be less well addressed, given that this category covers a broad set of PRO's.

This review revealed that the vision-related literature and the glaucoma literature in particular, contain PRO-instruments with different levels of quality, which should therefore be selected and used with caution. An evaluation of these instruments based on a comprehensive framework of quality criteria elucidates that not all retrieved PRO-instruments have been developed or validated following one of the available validation guidelines. According to this framework, PRO-instruments should meet following important criteria: (1) a clear description of its aim and intended use; (2) a conceptual framework or definition of the concept of interest; relevant

to the study population; (3) comprehensive consulting with patients and a literature review in view of generating items, and adequate statistical techniques to support item selection; (4) evidence for unidimensionality using appropriate statistics; (5) statistically justified response scales and subsequent scoring system and (6) evidence on validity, reliability and responsiveness.

Of the 27 instruments found, only a few fulfill partially these quality criteria. Overall, the tools for assessing functional status demonstrated poor quality both in view of their development as well as for their validation process. Yet, further adaptation and testing could improve instruments with potential, such as the GQL-15, IMQ and the GSI[24–28] Within the QoL-measures, both the Glau-QoL and VisQoL had an extensive and theory based development process, but were generated and validated according to classical techniques.[43–45] Applying additional Rasch-analysis could strengthen their content and validity. The NEI-VFQ, which is a widely used QoL-instrument, has initially never been tested on its dimensionality,[35–37] which is a major flaw in its development. Other investigators convincingly demonstrated by using modern psychometric techniques that the original tool should be adapted and revalidated.[41,56] The TSS-IOP pops up as the highest quality instrument to assess side effects across different topical treatments,[48,49] yet might be improved as well using Rasch-analysis. If interested in assessing adherence with eye drop treatment, both the adherence questionnaire of Schwartz *et al* and the EDSQ should be improved, given that both intend to predict nonadherence, but that the discrimination between adherent and nonadherent patients remains difficult.[52,53] Both the scales developed by Sleath *et al*[55] promise to measure self-efficacy and outcome expectations respectively, yet more validity evidence should be provided first to strengthen this statement.

Where do most existing PRO instruments show weaknesses and which pitfalls should future instrument developers avoid?

First, using a conceptual framework, derived from patient input in qualitative studies, as a starting point during the instrument developmental process looks like an exception in the glaucoma-related literature, given that $<50\%$ of the instrument developers use it. This confirms previous research of Ferrans[57] on QoL-PRO's in cancer patients showing that most of the instruments in the literature do not use a theoretical approach. Nevertheless, developing and using an appropriate and clearly defined conceptual definition/framework is important in order to know what concept to measure and how to measure it. In a conceptual framework, the interrelationships between items within a domain and of domains within a PRO-concept are depicted in a way that the concept of interest can be operationalized and appropriate psychometric analysis can be performed.[11,58] It should provide the rationale for, and specification of, the PRO-outcomes of interest (eg, side effects) in the population of interest (eg, glaucoma patients undergoing eye drop treatment) for a particular decision (eg, choice of appropriate eye drop treatment). Hence, not using a framework can cause difficulties with (1) grouping and scoring of items into domains, (2) the analysis and (3) the interpretation of PRO-scores if one doesn't know what is assessed.[58]

Second, many investigators only use expert opinion and/or a literature review to generate a preliminary list of items, yet the crucial factor to ensure a good breadth of relevance, which is the perspective of patients, was neglected in 11 instruments. According to the FDA[11] and the applied quality criteria,[12] PRO-instrument item-generation is incomplete without patient involvement (eg, patient interviews or focus groups) and should incorporate the input of a wide range of patients with the condition of interest to represent appropriate variations in severity and in population characteristics (eg, age, gender).[11]

Third, only a limited number of investigators used an appropriate item reduction strategy on a pilot questionnaire using statistical techniques such as factor analysis or Rasch-analysis (60%). Yet, this approach is needed to determine if all items tap the underlying construct being measured. Items discriminating poorly (ie, large floor and ceiling effect), items with large percentages of missing data, unreliable and invalid items need to be discarded. Using these techniques will improve item quality, measurement precision and item-person targeting (ie, targeting of item-difficulty to person-ability).[12]

Fourth, a lot of instruments contain several dimensions covering a set of items. Too often, those dimensions are created based on the opinion of experts or patients, without performing any analyses to test for unidimensionality (ie, factor analysis, Rasch-analysis or chronbach's alpha) within a scale or subscale. This is necessary to demonstrate that all the items included, fit within a single underlying construct in order to be able to calculate valid subscale scores.

Fifth, traditional summary scoring still remains the most popular scoring system in the ophthalmologic literature, yet most of the instruments did not statistically justify their rating scales and scoring systems ($n=20$). Summing items hypothesize that all questions have equal importance. Response categories are often accordingly scaled and have equal values with uniform increments form one category to the other (eg, distance between score 1 and 2 is the same as distance between response option 3 and 4).[12] Rasch-analysis can therefore

be used to detect redundant and disordered thresholds. Differently calibrated response categories can help to provide a more valid scale, compared with a 'one size fits all' scoring approach.

Sixth, further validation of PRO-instruments is of course only meaningful if the item content of the new tool was obtained by following the adequate steps as mentioned above. If not, instrument-developers should first optimize their instrument before obtaining evidence on their performance in view of validity, reliability and responsiveness. It is obvious that most of the instruments selected for this review did not follow the ideal developmental process as described in the framework of Pesudovs *et al*.[12] Without first improving the content of the tool, most of the investigators already started to validate their instrument. In that perspective modern psychometric testing may help to improve the content of the instrument and may help to provide stronger reliability and validity evidence. More specifically, Rasch-analysis that is a modern psychometric statistical technique provides a transformation of the ordinal raw score into a linear interval scale permitting the use of parametric statistical techniques. This approach improves the accuracy of scoring and removes noise from the measure, which in turn improves sensitivity to change and correlation with other variables. Additionally, the instruments' validity can be assessed by analyzing the fit of items to the overall construct and the item-person targeting (ie, targeting of item-difficulty to person-ability).[47] Therefore, instruments that are only tested using the conventional techniques are not necessary invalid, yet could still be improved using Rasch-analysis. However, this review revealed that most authors did not try to improve the quality of their instrument, even if the results from validity and reliability tests show unsatisfactory evidence.

Seventh, most papers reported only a limited amount of information related to the practical use of the instrument, more specifically concerning: (1) instructions for users describing how to complete and to score a questionnaire and how to interpret the results; (2) the burden of questionnaire administration, such as the duration of questionnaire administration (only mentioned in six instruments), the font size, the presence of new instructions for each item and the formatting and; (3) understandability and readability of the questionnaire tested in the patient population of interest. This information would be helpful for researchers and clinicians to allow them selecting the best instrument for its intended goal.

Hence, there is still a lot of room for improvement of the quality of existing instruments and newly developed PRO instruments should learn from the drawbacks of others. The quality criteria outlined in the framework of Pesudovs *et al*[12] can certainly help investigators in these efforts.

Recommendations

Following the conceptual framework of Acquadro *et al*,[9] the glaucoma PRO-literature covers all classes of PRO's, yet most of the instruments only adhere to a limited extent to the predefined quality criteria. Ideally, researchers should start from a conceptual framework and should most importantly use the patients' perspective in view of item generation, for example, by organizing focus groups and in depth interviews. This review clearly shows that this aspect has to improve in future studies and should also be more clearly reported in future papers. The same is true for item-reduction techniques and scoring systems of instruments, which should both be statistically justified. Psychometric testing is limited in some of the PRO's, yet it seems that modern test theories gain more and more attention in the vision-related literature to optimize instruments in terms of item-content and to provide stronger validity evidence.

Other future directions in instrument development could be glaucoma-specific 'item banking', referring to Rasch-analysis on all items extracted from several existing questionnaires measuring the same construct. In this approach, all item are calibrated onto a single scale and can be selected manually or by a computer algorithm to target the ability of the patients under test.[59]

This review adds to the state of the art literature, as it is the first overview of all PRO's available for glaucoma patients, wherein their quality is rated following the FDA-guidelines[11] and the comprehensive framework developed by Pesudovs *et al*.[12] Therefore, this overview could serve as a guidance instrument for ophthalmologists and researchers, who plan to use them in pharmaceutical studies or during clinical practice.

Conflict of interest

The authors declare no conflict of interest.

References

1 Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol* 2006; **90**(3): 253–254.
2 American Academy of Ophthalmology Preferred Practice Patterns Glaucoma Panel. *Primary Open-Angle Glaucoma*. American Academy of Ophthalmology: San Fransisco, 2000.
3 Janz NK, Wren PA, Lichter PR, Musch DC, Gillespie BW, Guire KE *et al*. The Collaborative Initial Glaucoma Treatment Study: interim quality of life findings after initial medical or surgical treatment of glaucoma. *Ophthalmology* 2001; **108**(11): 1954–1965.

4 Lee DA, Higginbotham EJ. Glaucoma and its treatment: a review. *Am J Health Syst Pharm* 2005; **62**(7): 691–699.
5 Hartmann CW, Rhee DJ. The patient's journey: glaucoma. *BMJ* 2006; **333**(7571): 738–739.
6 Gutierrez P, Wilson MR, Johnson C, Gordon M, Cioffi GA, Ritch R *et al*. Influence of glaucomatous visual field loss on health-related quality of life. *Arch Ophthalmol* 1997; **115**(6): 777–784.
7 Shaw MJ, Crawley JA. Improving health-related quality of life in gastro-oesophageal reflux disease. *Drugs* 2003; **63**(21): 2307–2316.
8 Leidy NK, Revicki DA, Geneste B. Recommendations or evaluating the validity of quality of life claims for labeling and promotion. *Value Health* 1999; **2**(2): 113–127.
9 Acquadro C, Berzon R, Dubois D, Leidy NK, Marquis P, Revicki D *et al*. Incorporating the patient's perspective into drug development and communication: an *ad hoc* task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. *Value Health* 2003; **6**(5): 522–531.
10 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–74.
11 US Food and Drug Administration DoHaHS. Guidance for Industry: Patient Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims: Draft Guidance. 2006.
12 Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci* 2007; **84**(8): 663–674.
13 Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 2nd edn. Lawrence Erlbaum Associates: Mahwah, NJ, 2007.
14 American Educational Research Association, American Psychological Association and National Council of Measurements. *Standards for Educational and Psychological Testing*. AERA: Washington, DC, 1999.
15 Revicki DA, Ehreth JL. Health-related quality-of-life assessment and planning for the pharmaceutical industry. *Clin Ther* 1997; **19**(5): 1101–1115.
16 Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 1992; **45**(12): 1341–1345.
17 Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB *et al*. Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clin Ther* 1996; **18**(5): 979–992.
18 Nunally JC, Bernstein IH. *Psychometric Theory*. 3rd edn. McGraw-Hill: New York, 1994.
19 Hays R, Anderson R, Revicki DA. Assessing reliability and validity of measurement in clinical trials. In: Staquet M, Hays R, Fayers P, (eds). *Quality of Life Assessment in Clinical Trials*. Oxford University Press: New York, 1998, pp. 169–182.
20 Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK *et al*. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual Life Res* 2000; **9**(8): 887–900.
21 Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm* 2008; **65**(23): 2276–2284.
22 Leidy NK. Functional status and the forward progress of merry-go-rounds: toward a coherent analytical framework. *Nurs Res* 1994; **43**(4): 196–202.
23 Sloane M, Ball K, Owsley C, Bruni J, Roenker D. The Visual Activities Questionnaire: developing an instrument for assessing problems in everyday visual tasks. *Technical Digest, Noninvasive Assessment of the Visual System, Topical Meeting of the Optical Society of America* 1992; January: 1–14.
24 Turano KA, Geruschat DR, Stahl JW, Massof RW. Perceived visual ability for independent mobility in persons with retinitis pigmentosa. *Invest Ophthalmol Vis Sci* 1999; **40**(5): 865–877.
25 Turano KA, Massof RW, Quigley HA. A self-assessment instrument designed for measuring independent mobility in RP patients: generalizability to glaucoma patients. *Invest Ophthalmol Vis Sci* 2002; **43**(9): 2874–2881.
26 Walt JG, Rendas-Baum R, Kosinski M, Vaishali P. Psychometric evaluation of the glaucoma symptom identifier. *J Glaucoma* 2011; **20**(3): 148–159.
27 Nelson P, Aspinall P, O'Brien C. Patients' perception of visual impairment in glaucoma: a pilot study. *Br J Ophthalmol* 1999; **83**(5): 546–552.
28 Nelson P, Aspinall P, Papasouliotis O, Worton B, O'Brien C. Quality of life in glaucoma and its relationship with visual function. *J Glaucoma* 2003; **12**(2): 139–150.
29 Spilker B, Molinek Jr FR, Johnston KA, Simpson Jr RL, Tilson HH. Quality of life bibliography and indexes. *Med Care* 1990; **28**(12 Suppl): DS1–D77.
30 Moons P, Budts W, De GS. Critique on the conceptualisation of quality of life: a review and evaluation of different conceptual approaches. *Int J Nurs Stud* 2006; **43**(7): 891–901.
31 Moons P, Marquet K, Budts W, De GS. Validity, reliability and responsiveness of the 'Schedule for the Evaluation of Individual Quality of Life-Direct Weighting' (SEIQoL-DW) in congenital heart disease. *Health Qual Life Outcomes* 2004; **2**: 27.
32 Janz NK, Wren PA, Lichter PR, Musch DC, Gillespie BW, Guire KE. Quality of life in newly diagnosed glaucoma patients: the Collaborative Initial Glaucoma Treatment Study. *Ophthalmology* 2001; **108**(5): 887–897.
33 Mills RP, Janz NK, Wren PA, Guire KE. Correlation of visual field with quality-of-life measures at diagnosis in the Collaborative Initial Glaucoma Treatment Study (CIGTS). *J Glaucoma* 2001; **10**(3): 192–198.
34 Carta A, Braccio L, Belpoliti M, Soliani L, Sartore F, Gandolfi SA *et al*. Self-assessment of the quality of vision: association of questionnaire score with objective clinical tests. *Curr Eye Res* 1998; **17**(5): 506–511.
35 Mangione CM, Berry S, Spritzer K, Janz NK, Klein R, Owsley C *et al*. Identifying the content area for the 51-item National Eye Institute Visual Function Questionnaire: results from focus groups with visually impaired persons. *Arch Ophthalmol* 1998; **116**(2): 227–233.
36 Mangione CM, Lee PP, Pitts J, Gutierrez P, Berry S, Hays RD. Psychometric properties of the National Eye Institute Visual Function Questionnaire (NEI-VFQ). NEI-VFQ Field Test Investigators. *Arch Ophthalmol* 1998; **116**(11): 1496–1504.
37 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.
38 Burr JM, Kilonzo M, Vale L, Ryan M. Developing a preference-based glaucoma utility ndex using a discrete choice experiment. *Optom Vis Sci* 2007; **84**: 797–808.

39 Dreer LE, McGwin Jr G, Scilley K, Meek GC, Dyer A, Seker D *et al*. Development of a nursing home vision-targeted health-related quality of life questionnaire for older adults. *Aging Ment Health* 2007; **11**(6): 722–733.
40 Scilley K, Owsley C. Vision-specific health-related quality of life: content areas for nursing home residents. *Qual Life Res* 2002; **11**(5): 449–462.
41 Marella M, Pesudovs K, Keeffe JE, O'Connor PM, Rees G, Lamoureux EL. The psychometric validity of the NEI VFQ-25 for use in a low-vision population. *Invest Ophthalmol Vis Sci* 2010; **51**(6): 2878–2884.
42 Massof RW, Fletcher DC. Evaluation of the NEI visual functioning questionnaire as an interval measure of visual ability in low vision. *Vision Res* 2001; **41**(3): 397–413.
43 Misajon R, Hawthorne G, Richardson J, Barton J, Peacock S, Iezzi A *et al*. Vision and quality of life: the development of a utility measure. *Invest Ophthalmol Vis Sci* 2005; **46**(11): 4007–4015.
44 Peacock S, Misajon R, Iezzi A, Richardson J, Hawthorne G, Keeffe J. Vision and quality of life: development of methods for the VisQoL vision-related utility instrument. *Ophthalmic Epidemiol* 2008; **15**(4): 218–223.
45 Béchetoille A, Arnould B, Bron A, Baudouin C, Renard JP, 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.
46 Lamoureux EL, Pesudovs K, Pallant JF, Rees G, Hassell JB, Caudle LE *et al*. An evaluation of the 10-item vision core measure 1 (VCM1) scale (the Core Module of the Vision-Related Quality of Life Scale) using Rasch analysis. *Ophthalmic Epidemiol* 2008; **15**(4): 224–233.
47 Lamoureux EL, Ferraro JG, Pallant JF, Pesudovs K, Rees G, Keeffe JE. Are standard instruments valid for the assessment of quality of life and symptoms in glaucoma? *Optom Vis Sci* 2007; **84**(8): 789–796.
48 Atkinson MJ, Stewart WC, Fain JM, Stewart JA, Dhawan R, Mozaffari E *et al*. A new measure of patient satisfaction with ocular hypotensive medications: the Treatment Satisfaction Survey for Intraocular Pressure (TSS-IOP). *Health Qual Life Outcomes* 2003; **1**(1): 67.
49 Day DG, Sharpe ED, Atkinson MJ, Stewart JA, Stewart WC. The clinical validity of the treatment satisfaction survey for intraocular pressure in ocular hypertensive and glaucoma patients. *Eye* 2006; **20**(5): 583–590.
50 Barber BL, Strahlman ER, Laibovitz R, Guess HA, Reines SA. Validation of a questionnaire for comparing the tolerability of ophthalmic medications. *Ophthalmology* 1997; **104**(2): 334–342.
51 Ruiz MA, Pardo A, Martinez de la Casa JM, Polo V, Esquiro J, Soto J. Development of a specific questionnaire measuring patient satisfaction with glaucoma treatment: Glausat. *Ophthalmic Epidemiol* 2010; **17**(3): 131–143.
52 Nordmann JP, Denis P, Vigneux M, Trudeau E, Guillemin I, Berdeaux G. Development of the conceptual framework for the Eye-Drop Satisfaction Questionnaire (EDSQ) in glaucoma using a qualitative study. *BMC Health Serv Res* 2007; **7**: 124.
53 Schwartz GF, Plake KS, Mychaskiw MA. An assessment of readiness for behaviour change in patients prescribed ocular hypotensive therapy. *Eye* 2009; **23**(8): 1668–1674.
54 Lee BL, Gutierrez P, Gordon M, Wilson MR, Cioffi GA, Ritch R *et al*. The Glaucoma Symptom Scale. A brief index of glaucoma-specific symptoms. *Arch Ophthalmol* 1998; **116**(7): 861–866.
55 Sleath B, Blalock SJ, Robin A, Hartnett ME, Covert D, Devellis B *et al*. Development of an instrument to measure glaucoma medication self-efficacy and outcome expectations. *Eye (Lond)* 2010; **24**(4): 624–631.
56 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.
57 Ferrans CE. Definitions and conceptual models of quality of life. In: Lipscomb J, Gotay CC, Snyder C, (eds). *Outcomes Assessment in Cancer: Measures, Methods, and Applications*. Cambridge University Press: Cambridge, 2005.
58 Rothman ML, Beltran P, Cappelleri JC, Lipscomb J, Teschendorf B. Patient-reported outcomes: conceptual issues. *Value Health* 2007; **10**(Suppl 2): S66–S75.
59 Pesudovs K. Item banking: a generational change in patient-reported outcome measurement. *Optom Vis Sci* 2010; **87**(4): 285–293.
60 Allen MJ, Yen WM. *Introduction to Measurement Theory*. Waveland Press, Inc.: Long Grove, IL, 2002.
61 Gothwal VK, Wright TA, Lamoureux EL, Pesudovs K. Visual Activities Questionnaire: assessment of subscale validity for cataract surgery outcomes. *J Cataract Refract Surg* 2009; **35**(11): 1961–1969.
62 Ross JE, Bron AJ, Clarke DD. Contrast sensitivity and visual disability in chronic simple glaucoma. *Br J Ophthalmol* 1984; **68**(11): 821–827.
63 Mills RP, Drance SM. Esterman disability rating in severe glaucoma. *Ophthalmology* 1986; **93**(3): 371–378.
64 Viswanathan AC, McNaught AI, Poinoosawmy D, Fontana L, Crabb DP, Fitzke FW *et al*. Severity and stability of glaucoma: patient perception compared with objective measurement. *Arch Ophthalmol* 1999; **117**(4): 450–454.
65 Iester M, Zingirian M. Quality of life in patients with early, moderate and advanced glaucoma. *Eye* 2002; **16**(1): 44–49.
66 Lamoureux EL, Fenwick E, Moore K, Klaic M, Borschmann K, Hill K. Impact of the severity of distance and near-vision impairment on depression and vision-specific quality of life in older people living in residential care. *Invest Ophthalmol Vis Sci* 2009; **50**(9): 4103–4109.
67 Wolffsohn JS, Cochrane AL. Design of the low vision quality-of-life questionnaire (LVQOL) and measuring the outcome of low-vision rehabilitation. *Am J Ophthalmol* 2000; **130**(6): 793–802.
68 Wolffsohn JS, Cochrane AL, Watt NA. Implementation methods for vision related quality of life questionnaires. *Br J Ophthalmol* 2000; **84**(9): 1035–1040.
69 de Boer MR, Terwee CB, de Vet HC, Moll AC, Volker-Dieben HJ, van Rens GH. Evaluation of cross-sectional and longitudinal construct validity of two vision-related quality of life questionnaires: the LVQOL and VCM1. *Qual Life Res* 2006; **15**(2): 233–248.
70 de Boer MR, de Vet HC, Terwee CB, Moll AC, Volker-Dieben HJ, van Rens GH. Changes to the subscales of two vision-related quality of life questionnaires are proposed. *J Clin Epidemiol* 2005; **58**(12): 1260–1268.
71 Frost NA, Sparrow JM, Durant JS, Donovan JL, Peters TJ, Brookes ST. Development of a questionnaire for measurement of vision-related quality of life. *Ophthalmic Epidemiol* 1998; **5**(4): 185–210.

72 Frost NA, Sparrow JM, Hopper CD, Peters TJ. Reliability of the VCM1 Questionnaire when administered by post and by telephone. *Ophthalmic Epidemiol* 2001; **8**(1): 1–11.
73 van Nispen RM, Knol DL, Mokkink LB, Comijs HC, Deeg DJ, van Rens GH. Vision-related quality of life core measure (VCM1) showed low-impact differential item functioning between groups with different administration modes. *J Clin Epidemiol* 2010; **63**(11): 1232–1241.
74 Keeffe JE, Lam D, Cheung A, Dinh T, McCarty CA. Impact of vision impairment on functioning. *Aust N Z J Ophthalmol* 1998; **26**(Suppl 1): S16–S18.
75 Hassell JB, Weih LM, Keeffe JE. A measure of handicap for low vision rehabilitation: the impact of vision impairment profile. *Clin Experiment Ophthalmol* 2000; **28**(3): 156–161.
76 Weih LM, Hassell JB, Keeffe J. Assessment of the impact of vision impairment. *Invest Ophthalmol Vis Sci* 2002; **43**(4): 927–935.
77 Noe G, Ferraro J, Lamoureux E, Rait J, Keeffe JE. Associations between glaucomatous visual field loss and participation in activities of daily living. *Clin Experiment Ophthalmol* 2003; **31**(6): 482–486.
78 Lamoureux EL, Pallant JF, Pesudovs K, Rees G, Hassell JB, Keeffe JE. The impact of vision impairment questionnaire: an assessment of its domain structure using confirmatory factor analysis and Rasch analysis. *Invest Ophthalmol Vis Sci* 2007; **48**(3): 1001–1006.
79 Lamoureux EL, Pallant JF, Pesudovs K, Hassell JB, Keeffe JE. The Impact of Vision Impairment Questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci* 2006; **47**(11): 4732–4741.
80 Langelaan M, van Nispen RM, Knol DL, Moll AC, de Boer MR, Wouters B *et al* Visual Functioning Questionnaire: reevaluation of psychometric properties for a group of working-age adults. *Optom Vis Sci* 2007; **84**(8): 775–784.