

Title: Evaluation of item candidates for a diabetic retinopathy quality of life item bank.

Authors: Eva K Fenwick¹ MA; Konrad Pesudovs² PhD; Jyoti Khadka² PhD; Gwyn Rees¹ PhD; Tien Y Wong^{1,3} PhD; Ecosse L Lamoureux^{1,3} PhD.

¹ Centre for Eye Research, Department of Ophthalmology, University of Melbourne, Royal Victorian Eye and Ear Hospital, Melbourne, Victoria, Australia.

² NH&MRC Centre for Clinical Eye Research, Department of Optometry and Vision Science, Flinders Medical Centre and Flinders University of South Australia, Adelaide, South Australia, Australia.

³ Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore.

Address of Corresponding Author

A/Prof Ecosse Lamoureux

Centre for Eye Research Australia

University of Melbourne

32 Gisborne St

East Melbourne Victoria, 3002, Australia

Telephone: 61 3 9929 8371

Fax: 61 3 9962 3859

Email: ecosse@unimelb.edu.au

Word count: 3890

Abstract

Purpose: We are developing an item bank assessing the impact of diabetic retinopathy (DR) on quality of life (QoL) using a rigorous multi-staged process combining qualitative and quantitative methods. We describe here the first two, qualitative phases: content development and item evaluation.

Methods: After a comprehensive literature review, items were generated from four sources: 1) 34 previously validated patient-reported outcome measures; 2) five published qualitative articles; 3) eight focus groups and 18 semi-structured interviews with 57 DR patients; and 4) seven semi-structured interviews with diabetes or ophthalmic experts. Items were then evaluated during 3 stages, namely binning (grouping) and winnowing (reduction) based on key criteria and panel consensus; development of item stems and response options; and pre-testing of items via cognitive interviews with patients.

Results: The content development phase yielded 1165 unique items across 7 QoL domains. After 3 sessions of binning and winnowing, items were reduced to a minimally representative set (n=312) across 9 domains of QoL: Visual symptoms; Ocular surface symptoms; Activity limitation; Mobility; Emotional; Health concerns; Social; Convenience; and Economic. After 8 cognitive interviews, 42 items were amended, resulting in a final set of 314 items.

Conclusions: We have employed a systematic approach to develop items for a DR-specific QoL item bank. The psychometric properties of the nine QoL subscales will be assessed using Rasch analysis. The resulting validated item bank will allow clinicians and researchers to better understand the QoL impact of DR and DR therapies from the patients' perspective.

Key words: item bank, diabetic retinopathy, quality of life, qualitative, eye disease

INTRODUCTION

Understanding the impact of disease and health-related interventions from the patients' perspective is now recognised as important and the use of a patient-reported outcome measure in clinical trials is a requirement of many funding bodies.[1] Moreover, data compiled from patient reported outcome measures are increasingly utilised in policy planning and health-related budget expenditure.[1] As such, high quality patient reported outcome measures are crucial to ensure accurate assessment of subjective patient experience.

The impact of diabetic retinopathy (DR), a vision-threatening microvascular complication of diabetes, on patients' quality of life (QoL) is substantial, especially at the severe stages of proliferative DR and clinically significant macular oedema.[2] However, our understanding of the impact of DR on QoL has been limited to date by the lack of an appropriate patient reported outcome measure.[2, 3] While a range of visual functioning questionnaires are available,[4-6] these provide information mainly on a single aspect of QoL, namely vision-related activity limitation. Similarly, vision-related QoL instruments,[7-9] while being more holistic, may miss specific aspects of DR-incurred impact. Only one DR-specific patient reported outcome measure currently exists – The Retinopathy Dependent Quality of Life Questionnaire (RetDQoL);[10, 11] however, there are several limitations associated with this instrument. For example, the RetDQoL was developed and validated using Classical Test Theory which harbours several erroneous assumptions, namely that all items are assumed to be of equal difficulty, and also that the change between response categories is uniform and their allocated values can be summed to provide an overall score representing valid measurement of the underlying latent trait. However, modern psychometric theory has shown both assumptions to be false.[12] The RetDQoL also uses a scoring method that involves multiplying two sets of patient-reports together, an approach that has been demonstrated to be fundamentally flawed.[13, 14] Lack of an appropriate DR-specific patient reported outcome measure restricts our understanding of the full impact of DR on all QoL parameters as well as the QoL impact of treatment therapies for DR. This is critical as new DR

treatment modalities, such as anti-vascular endothelial growth factor (VEGF) intravitreal injections, are fast emerging.[15, 16]

Therefore, we are developing a DR-specific QoL item bank to assess the specific impact of DR and diabetic macular oedema (DME) on all relevant domains of QoL, using a rigorous developmental process. Item banking has its roots in educational testing but has recently become popular in health-related research.[17-19] An item bank is a pool of calibrated items (questions) that measure a defined latent trait[20] with the items represent differing amounts of that latent trait along a continuum.[21] Targeted items from the item bank are presented to each participant according to their ability level by the use of a computer adaptive testing (CAT) system, which selects items based on the examinee's responses to previous questions.[22]

Item banking overcomes many of the limitations of traditional paper-pencil questionnaires, which are static, inflexible and often burdensome to complete. Similarly, because relatively few items are available in paper-pencil questionnaires, they often are poorly targeted to person ability and are thus of limited use across a broad disease spectrum. In contrast, item banks with a CAT system can provide more accurate, precise, valid and reliable measurements of the desired latent trait whilst requiring fewer items during computer driven administration where items are targeted to the person.[18, 22]

An item bank is the culmination of several equally important qualitative and quantitative developmental phases, namely: 1) Content development, in which items are generated based on a literature review of previous patient reported outcome measures, and original qualitative interviews with patients and field experts; 2) Item evaluation, in which items are reduced to a minimally representative set that defines the latent trait through a systematic set of criteria based on expert panel consensus; and 3) Pilot testing and Validation through modern psychometric methods such as Rasch analysis. While the majority of published studies focus on reporting the quantitative results of the pilot testing or validation phases, few outline the specific processes involved in the development and evaluation of items and domains for an item bank.[23-25] This is surprising as

these initial developmental phases represent the foundation of an item bank and their importance should not be underestimated. Reporting these data allows the scientific quality of the processes employed during these phases to be evaluated. Moreover, given that use of item banking in health research is increasing, a better understanding of the multi-phased effort required to develop an item bank will benefit researchers and may ultimately lead to improvements in the quality of item banks produced.

In this paper we therefore report on the first two qualitative phases of this multi-staged effort to produce a DR-specific QoL item bank - Content development and Item evaluation (**Electronic Supplemental Figure S1**). The content development phase commenced with a comprehensive literature review on the subject including an inventory of items from currently existing, relevant questionnaires and development of a working theoretical framework of the latent trait under study. Focus groups and semi-structured interviews were then conducted with patients with the condition of interest and experts in the field to ensure all relevant themes were explored, and suitable items were then generated from the qualitative data. Next, items compiled in the preceding phase were grouped into different QoL item pools representing hypothesised latent traits (e.g. Mobility). A systematic set of criteria was then employed by an expert panel to reduce the item pools to a minimally representative set of items defining each latent trait. The items were then pre-tested with a small number of participants to ensure optimal content and clarity. While several measures were taken to ensure that both phases were conducted as objectively as possible, they remain inherently subjective, qualitative processes.

Our methodology will be informative to researchers developing item banks in other health-related fields. The study had ethical approval from the Royal Victorian Eye and Ear Hospital (RVEEH) Human Research Ethics Committee (# 09/888H) and was conducted in accordance with the Declaration of Helsinki.

PHASE 1: Content development

A comprehensive literature review of the impact of DR and DME on patients' visual functioning and QoL was initially conducted. The literature search included the Pubmed, ISI Web of Science and Embase databases and no date restriction was applied. Search terms included *diabetic retinopathy* OR *diabetic macular (o)edema* OR *diabetic vision loss* OR *vision impairment* AND *quality of life* OR *functioning* OR *impact* OR *depression* OR *anxiety* OR *psychosocial* OR *emotional*. Names of specific questionnaires were also used in conjunction with the other search terms. Additional papers were obtained from the reference lists of relevant papers.[26] Papers were included if they a) were qualitative or quantitative; b) provided specific data for DR or DME; c) described any aspect of QoL impact. Conference abstracts, studies using a non-validated patient reported outcome measure, papers discussing other eye diseases, and papers referring to general visual impairment from a variety of ocular conditions without specific data relating to DR were excluded by reviewing the title and abstract. The following data were extracted from each relevant article: authors, year, patient characteristics (diabetes type, number of participants with DR), study design and outcome measure(s) used (if applicable), and summary of key findings. The study findings were categorized and synthesized into relevant QoL domains.

'Quality of life' is a complex concept that encompasses functional ability, symptoms, emotional well-being, social relationships, concerns and convenience.[27] Existing patient reported outcome measures assessing vision-specific QoL, for example the NEI-VFQ[28] and the IND-VFQ,[29] tend to be multidimensional when subjected to Rasch analysis and operate best as a set of separate subscales. Therefore, we hypothesised that our DR-specific QoL item bank would actually comprise several unidimensional item pools each representing a specific latent trait or QoL domain. Our hypothesised set of QoL domains was guided by both a "top-down" and "bottom-up" approach. By top-down, we refer to a comprehensive review of the literature and consideration of QoL domains proposed in existing vision-related patient reported outcome measures, which provided us with a working conceptual framework. However, we also utilised a bottom-up approach where we were guided by the data and the psychometric experience of the expert panel, giving us flexibility to

divide, remove or add domains based on item content. As such, we started with a solid organisational structure but had the freedom to ensure that our unique data was faithfully represented.

Following the literature review, content development proceeded from four sources: 1) previously validated vision- and health-related functioning and QoL patient reported outcome measures; 2) published qualitative articles on the impact of DR on QoL; 3) focus groups and semi-structured interviews with patients with DR and DME and; 4) semi-structured interviews with experts in the field of diabetes and ophthalmology.

Validated patient reported outcome measures

All relevant items were extracted from 34 health-, vision- and condition-specific QoL and vision functioning patient reported outcome measures which were identified from key review papers on vision-related patient reported outcome measures[30-32] (**Electronic Supplemental Table S1**) and compiled into item pools. Items which were clearly not relevant or applicable to DR were excluded. For example, items rating general health or assessing bodily pain in the SF-36 were not included. Most instruments (n=22) focussed primarily on vision-related activity limitation or functioning. A total of 968 items were extracted from these 34 instruments. After accounting for identical items, 586 unique items remained (**Table 1**).

Published qualitative papers

Five qualitative papers that explored the impact of DR and DME on patients' QoL using focus groups and semi-structured interviews were analysed for relevant QoL domains and themes.[10, 33-36] A total of 170 items were extracted covering a variety of aspects of patients' QoL (**Table 1**).

Qualitative interviews: patients and experts

A total of eight focus groups and 18 semi-structured interviews were conducted with 57 patients with DR and/or DME. The methodology of the qualitative interviews has been described previously.[37] In brief, participants were recruited primarily from eye clinics at the RVEEH and

were aged 18 years or older, had type 1 or 2 diabetes, DR and/or DME and had no significant hearing or cognitive impairment. Just under half of participants (n=27, 47.4%) had proliferative DR in the better eye and a quarter (n=14, 24.6%) had clinically significant macular oedema.

Approximately half of participants had at least mild distance visual acuity impairment (>0.3 LogMAR) in at least one eye (**Electronic Supplemental Table S2**). Seven semi-structured interviews were conducted with diabetes and ophthalmic experts. Participants included three retinal specialists; a diabetes educator; a Medical Retina clinic orthoptist; an endocrinologist; and a low vision specialist.

The discussions were prompted by an open-ended set of questions developed from a comprehensive literature review[2, 3] and input from ophthalmic specialists. Groups and interviews were conducted until theoretical saturation was reached.[38] Data were analysed using an iterative process based on the constant comparative method[38, 39] using the qualitative software NVIVO (QSR International Pty Ltd, 2007). Analysis of the qualitative data from the patient and expert sessions generated 661 and 191 items, respectively, covering multiple areas of QoL (**Table 1**). Thus, more than 50% of the data for our item bank originated from original qualitative research. These data were particularly fertile regarding emotional and social well-being, concerns, convenience and work, areas which were poorly represented in most of the 34 previously validated patient reported outcome measures. At the conclusion of the content development phase, the total number of items generated from the four separate sources was 1608. After removal of identical items, 1165 remained across seven domains of QoL (**Table 2, row 1 ‘initial pools’**), namely **Visual symptoms; treatment; activity limitation; mobility; emotional; social; and economic**. Note that these seven domains represent the initial, temporary structure of the item pools and not the final instrument configuration. **Table 2** depicts the systematic process of moving from the initial item pools to the final instrument structure that contains 314 items across nine QoL domains. This is the process of item evaluation, which is described in further detail in the following sections.

PHASE 2: Item evaluation

The item evaluation process comprised three broad stages, namely (1) binning (grouping) and winnowing (reduction) of items; (2) development of item stems, a preceding statement, and item response options; and (3) pre-testing of the final items via cognitive interviews with DR patients.

(1) The process of binning and winnowing

Following the generation of the initial item set (n=1165), the task of reducing the items to a representative set began using a process of binning and winnowing based on the systematic protocol outlined by the NIH funded Patient Reported Outcomes Measurement Information System (PROMIS) group.[23, 25] Binning describes the process of grouping together items that are similar in meaning thus revealing redundancy and allowing elimination of unnecessary items.[23] An example of a working ‘bin’ in our item bank is Luminance under the Activity Limitation QoL domain. Thus, all items referring to functioning under difficult lighting conditions were grouped together. This process was guided by consensus of an expert panel who have substantial experience in patient reported outcomes research (co-authors KP, EL, EF).

The binning process also revealed groups of items that were initially incorrectly classified into particular QoL domains. For example, the Social and Mobility item pools decreased by n=55 and n=58 items, respectively, from the initial pools phase to the first stage of binning and winnowing (**Table 2**). This was largely due to reclassification of certain item bins, such as ‘Driving’ from Mobility and ‘Hobbies and Leisure’ from Social to the Activity Limitation domain, which increased by n=103 items. Similarly, the binning process also exposed a poorly conceived QoL domain, Treatment, which was subsequently dissolved and its items dispersed across existing pools, and also into three newly developed pools, namely Ocular surface symptoms, Convenience and Health Concerns. These three new domains also absorbed items from other existing domains. For example, many Symptom items were transferred into Ocular surface symptoms, Emotional items into Health Concerns and Activity Limitation items into Convenience.

Once items had been ‘binned’, the winnowing process began whereby potential items for deletion were identified and each bin was reduced to a minimally representative item set. A systematic set of criteria was employed by the panel to determine candidates for deletion (**Table 3**), namely:

- a) Item redundancy – item worded identically or very similar in content to another item;
- b) Item clarity – item confusing, poorly worded, double- or multi-barrelled;
- c) Item applicability – item too specialised to have a broad enough application;
- d) Item frequency – item did not occur often, or was not well-represented across the four sources of content development.
- e) Item relevance – particular precedence was given to items from qualitative patient interviews, as these were considered most likely to accurately reflect patient experiences.

The expert panel met three times to proceed with the process of binning and winnowing. In these sessions, each item in each domain was discussed and evaluated, and either retained and allocated to the correct domain or discarded with the reason for removal recorded for future reference. Given the large number of items to consider (n=1165) and the labour-intensive nature of the work, three full sessions of binning and winnowing were required. Following the three sessions, the total number of items was reduced from 1165 to 312 (**Table 2, row 5 ‘cognitive interviews’**) across nine distinct domains of QoL, namely *visual symptoms; ocular surface symptoms; activity limitation; mobility; emotional; health concerns; social; convenience; and economic*.

(2) Development of item stems, preceding statement & response options

Following binning and winnowing, item stems and response options for each QoL domain were chosen by the panel. The item stem is simply the question that precedes each specific item, such as “How much difficulty do you have...” Response options refer to the multiple choice categories provided to the participant to answer the question, such as “None, a little, quite a bit, a lot”. For commonly occurring domains such as *Visual symptoms, Ocular surface symptoms, Activity*

Limitation, and Mobility, the panel utilised empirical evidence from a comprehensive literature search regarding optimally functional item stems and response options.[40, 41] For Visual symptoms and Ocular surface symptoms, evidence suggested that three aspects were important: frequency, severity and bother and thus three different question formats and rating scales were employed for these domains. For more novel QoL domains such as Health Concerns, Convenience and Economic, little guidance was available from prior research and the item stems and response options were generated by the panel. Current evidence from modern psychometric methodology suggests that four to five response options are optimal[14] and thus our item pools predominantly utilised 5-point response categories plus a non-applicable option when necessary (**Electronic Supplemental Table S3**).

As each QoL domain was intended to function as a single subscale, item stems and response options remained consistent within each domain but were able to vary across domains. For QoL domains similar in content and which traditionally load together in Principal Components Analysis (i.e. Activity Limitation and Mobility), item stems and response options were worded identically so as to provide flexibility in the final instrument to combine such item pools if appropriate. The wording of each individual item was then determined so that it fit coherently with its item stem.

Finally, a preceding statement was determined for the overall item bank, namely: *Because of your diabetic retinopathy...* The scientific term ‘diabetic retinopathy’ was chosen instead of a more general phrase like ‘diabetic eye problems’ to ensure that participants’ responses were directly attributable to their DR rather than other ocular complications. Similarly, it assisted in focusing patients on the impact of their DR rather than any other systemic complications of diabetes, such as neuropathy, nephropathy, peripheral vascular disease and cardiovascular disease. To ensure that the term ‘diabetic retinopathy’ was fully understood, a brief description of the condition was included as part of the initial participant instructions.

(3) Pre-testing of the final items via cognitive interviews

Once the final items had been determined, eight cognitive interviews were conducted with patients with DR. Cognitive interviews are a vital step in pre-testing an instrument as they allow the clarity of instructions, items and response options to be assessed and any potential issues to be unearthed prior to pilot testing the instrument. They ensure that content coverage is adequate and that items are relevant, and therefore help improve instrument design and data quality.[42] The 312 questions were administered to each participant and any problematic items and particular comments were noted. Participants were also asked open-ended questions regarding the clarity, appropriateness and content coverage of items. Feedback from the cognitive interviews was iteratively incorporated and sessions were conducted until no further issues emerged. After the eight cognitive interviews, 4 items were deleted, 2 were added, 4 were split into separate items and 32 were re-phrased (**Table 4**). Some items were also transferred to other QoL domains as necessary. Therefore, at the conclusion of the cognitive interviews (**Table 2, row 5**) the final instrument comprised 314 items.

DISCUSSION

This paper describes the initial stages involved in developing an item bank to assess the impact of DR and DME on all domains of QoL. Following a comprehensive literature review, 1165 items were generated from previously validated patient reported outcome measures; published qualitative papers; focus groups and interviews with patients with DR; and experts in the field of ophthalmology and diabetes. Items were then reduced to a minimally representative set (n=312) via a systematic process of binning and winnowing based on key criteria and expert panel consensus. Item stems and response options for each domain and an overall preceding statement were formulated and the item set was pre-tested via cognitive interviews with DR and DME patients. The final item bank comprised **314 items** across **nine domains** of QoL. The development and coverage of this QoL instrument will result in a comprehensive DR-specific outcome measure available, far surpassing currently available instruments.[10, 11]

By using existing patient reported outcome measures to collate items, we were able to capitalise on previous research, and our use of qualitative sessions allowed us to produce data likely to be

relevant to patients with DR. That more than 50% of our items were generated from original qualitative research is a strength of our study and has resulted in the inclusion of many novel QoL domains such as 'health concerns' and 'convenience'. Our use of panel consensus to guide the subjective process of item deletion reduced the chance for bias and poor decision making. The use of key criteria to guide the panel enhanced consistency and minimised error. Pre-testing of items via cognitive interviews was also a key step in eliminating problems prior to finalising the item set. Finally, development of the QoL domains, item stems and response options was guided by empirical evidence and theoretical frameworks where available; however, we were also able to employ innovative solutions for our novel data.

One potential limitation is that most of the 34 patient reported outcome measures used to create the pool of extant items were not specific to DR and it is therefore possible that some items in the final item set are not relevant to DR patients. However, given that only one DR-specific patient reported outcome measure is currently available, the inclusion of patient reported outcome measures for other ocular conditions and general vision impairment was necessary. Moreover, since the winnowing process preferentially targeted items occurring in patient focus groups, it is unlikely that irrelevant items were included in the final item set. We also changed the wording of items and response options from pre-existing patient reported outcome measures so that they uniformly fit our proposed item format. While these items may function differently than in their original state, all changes are believed to result in overall item improvement. Another limitation is the small number of cognitive interviews conducted. To fully explore all issues with items and response options more cognitive sessions are generally considered necessary.[43] However, because most items were generated from patient data they were already conceptually relevant and most changes involved only minor re-wording. Hence, after eight interviews no new concerns were raised and subsequent sessions were deemed unnecessary. Finally, as diabetic patients may have a number of concurrent systemic complications such as neuropathy or cardiovascular disease that affect QoL, it is possible that some of the items may tap in to impairment unrelated to DR. However, the potential for this

type of contamination has been minimised by employing the preceding phrase *Because of your diabetic retinopathy...* before each item. Moreover, it is not unreasonable to assume that patients can adequately distinguish impairment or emotional reactions arising from their eye disease and vision impairment from that caused by neuropathy or comorbidity. Indeed, any condition-specific patient reported outcome measure is faced with this dilemma and its success in overcoming it may largely depend on the quality of its development.

In conclusion, we have developed the core items and item pools for a DR-specific QoL item bank. The next step is to pilot test the item pools in a large sample of patients across the spectrum of DR and assess the psychometric properties of each using Rasch analysis. Once the item bank is operational, a CAT system will be developed to administer the items. Our group is currently developing QoL item banks for all major ophthalmic diseases (the Eye-tem Bank). This DR-specific item bank represents the DR module of the Eye-tem Bank project. The resulting validated item bank will be useful for clinicians to better understand the impact of DR on patients' QoL; for researchers to assess the QoL impact of new and traditional treatment therapies or interventions from the patients' perspective; and for policy planners to better estimate the resources required by patients with DR.

ACKNOWLEDGEMENTS

National Health and Medical Research Council Centre for Clinical Research Excellence (CCRE) #529923 - Translational Clinical Research in Major Eye Diseases; CCRE Diabetes; Novartis Pharmaceuticals Australia #CRFB002DAU09T; Royal Victorian Eye and Ear Hospital; Operational Infrastructure Support from the Victorian Government.

REFERENCES

1. Hobart, J., et al., (2007). Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurology*, 1094-5.
2. Fenwick, E., et al., (2011). The Impact of Diabetic Retinopathy: Understanding the patient's perspective. *Br J Ophthalmol*, 95(6), 774-82.
3. Fenwick, E., et al., (2011). Social and emotional impact of diabetic retinopathy: a review. *Clin Experiment Ophthalmol*, 40(1), 27-38.
4. Steinberg, E.P., et al., (1994). The VF-14: An index of functional impairment in patients with cataract. *Arch Ophthalmol*, 112(5), 630-8.
5. Pesudovs, K. and D.J. Coster, (1998). An instrument for assessment of subjective visual disability in cataract patients. *Br J Ophthalmol*, 82(6), 617-24.
6. Mangione, C.M., et al., (1992). Development of the 'Activities of Daily Vision Scale'. A measure of visual functional status. *Med Care*, 30(12), 1111-26.
7. Lamoureux, E., et al., (2007). The impact of vision impairment questionnaire: an assessment of its domain structure using confirmatory factor analysis and rasch analysis. *Invest Ophthalmol Vis Sci*, 48, 1001-1006.
8. Lamoureux, E., et al., (2006). The Impact of Vision Impairment Questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci*, 47, 4732-4741.
9. Mangione, C.M., et al., (2001). Development of the 25-item National Eye Institute Visual Function Questionnaire. *Arch Ophthalmol*, 119(7), 1050-8.
10. Woodcock, A., et al., (2004). The influence of diabetic retinopathy on quality of life Interviews to guide the design of a condition-specific, individualised questionnaire: the RetDQoL. *Patient Educ Couns*, 53(3), 365-383.

11. Brose, L. and C. Bradley, (2009). Psychometric Development of the Individualized Retinopathy-Dependent Quality of Life Questionnaire (RetDQoL). *Value Health*, 14(6), 740-54.
12. Vanalphen, A., et al., (1994). Likert or Rasch - Nothing is more applicable than good theory. *J Adv Nurs*, 20(1), 196-201.
13. Finger, R., et al., (2012). Rasch analysis reveals problems with multiplicative scoring in the MacDQoL Ophthalmology, in press.
14. Gothwal, V.K., et al., (2011). Multiplicative rating scales do not enable measurement of vision-related quality of life. *Clin Exp Optom*, 94(1), 52-62.
15. Simo, R. and C. Hernandez, (2008). Intravitreal anti-VEGF for diabetic retinopathy: hopes and fears for a new therapeutic strategy. *Diabetologia*, 51(9), 1574-1580.
16. Wirostko, B., T.Y. Wong, and R. Simo, (2008). Vascular endothelial growth factor and diabetic complications. *Progress in Retinal and Eye Research*, 27(6), 608-621.
17. Lai, J.S., et al., (2005). An item bank was created to improve the measurement of cancer-related fatigue. *J Clin Epidemiol*, 58(2), 190-7.
18. Forkmann, T., et al., (2009). Development of an item bank for the assessment of depression in persons with mental illnesses and physical diseases using Rasch analysis. *Rehabil Psychol*, 54(2), 186-97.
19. Pesudovs, K., (2010). Item Banking: A Generational Change in Patient-Reported Outcome Measurement. *Optom Vis Sci*, 87(4), 1-9.
20. Cella, D., et al., (2007). The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res*, 16 Suppl 1, 133-41.

21. Lai, J.S., et al., (2003). Item banking to improve, shorten and computerize self-reported fatigue: an illustration of steps to create a core item bank from the FACIT-Fatigue Scale. *Qual Life Res*, 12(5), 485-501.
22. Gershon, R.C., (2005). Computer adaptive testing. *J Appl Meas*, 6(1), 109-27.
23. DeWalt, D.A., et al., (2007). Evaluation of item candidates - The PROMIS qualitative item review. *Medical Care*, 45(5), S12-S21.
24. Giesinger, J.M., et al., (2011). Cross-cultural development of an item list for computer-adaptive testing of fatigue in oncological patients. *Health and Quality of Life Outcomes*, 9.
25. Bruce, B., et al., (2009). Better assessment of physical function: item improvement is neglected but essential. *Arthritis Res Ther*, 11(6), R191.
26. Lundstrom, M. and K. Pesudovs, (2011). Questionnaires for measuring cataract surgery outcomes. *J Cataract Refract Surg*, 37(5), 945-59.
27. WHOQOL Group, *Measuring quality of life*. 1997, The World Health Organisation: Geneva. p. 1-13.
28. Pesudovs, K., et al., (2010). Remediating serious flaws in the National Eye Institute Visual Function Questionnaire. *J Cataract Refract Surg*, 36(5), 718-32.
29. Finger, R.P., et al., (2011). The impact of the severity of vision loss on vision-related quality of life in India: an evaluation of the IND-VFQ-33. *Invest Ophthalmol Vis Sci*, 52(9), 6081-8.
30. Massof, R.W. and G.S. Rubin, (2001). Visual function assessment questionnaires. *Surv Ophthalmol*, 45(6), 531-548.
31. Margolis, M.K., et al., (2002). Vision-specific instruments for the assessment of health-related quality of life and visual functioning - A literature review. *Pharmacoeconomics*, 20(12), 791-812.

32. de Boer, M.R., et al., (2004). Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt*, 24(4), 257-73.
33. Devenney, R. and S. O'Neil, (2010). The experience of diabetic retinopathy: A qualitative study. *Br J Health Psychol*, 16(4), 707-21.
34. Coyne, K.S., et al., (2004). The impact of diabetic retinopathy: perspectives from patient focus groups. *Fam. Pract.*, 21(4), 447-453.
35. Oehlergiarratana, J. and R.G. Fitzgerald, (1980). Group-therapy with blind diabetics. *Archives of General Psychiatry*, 37(4), 463-467.
36. Scanlon, P.H., et al., (2006). Reported symptoms and quality-of-life impacts in patients having laser treatment for sight-threatening diabetic retinopathy. *Diabet Med*, 23(1), 60-66.
37. Fenwick, E., et al., (2012). The impact of diabetic retinopathy on quality of life: qualitative findings from an item bank development project. *Qual Life Res*, Epub ahead of print.
38. Rice, P. and D. Ezzy. (1999). *Qualitative research methods : a health focus*. 1999, Oxford: Oxford University Press.
39. Maykut, P. and R. Morehouse. (1994). *Beginning qualitative research: Techniques and procedures for developing grounded theory*, 2nd ed. 1994, Thousand Oaks, California: Sage.
40. Gothwal, V.K., et al., (2009). Visual Activities Questionnaire: assessment of subscale validity for cataract surgery outcomes. *J Cataract Refract Surg*, 35(11), 1961-9.
41. Pesudovs, K., T.A. Wright, and V.K. Gothwal, (2010). Visual disability assessment: valid measurement of activity limitation and mobility in cataract patients. *Br J Ophthalmol*, 94(6), 777-81.

42. Jobe, J. and D. Mingay, (1989). Cognitive research improves questionnaires. *Am J Public Health*, 79(8), 1053-1055.
43. Willis, G. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. 2005, Thousand Oaks California: Sage Publications.

Table 1 – Total number of items generated across four sources of content development

Source of content development	Number of items generated
Validated patient reported outcome measures (n=34)	586
Qualitative articles (n=5)	170
Qualitative sessions with patients (n=57 patients)	661
Qualitative sessions with experts (n=7 experts)	191
<i>Total</i>	1608

Table 2 - The process of instrument development – from the initial item pools at the completion of content development (Phase 1) to the final instrument configuration at the completion of item evaluation (Phase 2)

	SY	TR	OS	AL	MB	EM	HC	SC	CN	EC	Total
Initial pools	101	54	-	460	165	222	-	118	-	45	1165
Binning & winnowing_1	68	-	16	563	107	81	-	63	89	45	1032
Binning & winnowing_2	18	-	8	129	34	73	-	35	35	20	352
Binning & winnowing_3	18	-	8	122	18	48	34	21	31	12	312
Cognitive interviews	18	-	10	120	19	48	36	21	30	12	314

SY=Visual symptoms; TR=Treatment; OS=Ocular surface symptoms; AL=Activity Limitation; MB=Mobility; EM=Emotional; HC=Health concerns; SC=Social; CN=Convenience; EC=Economic.

Table 3 – Examples of items removed during the Wining process

Wining criteria	Quality of life domain	Item stem
Item redundancy	Visual symptoms	<i>'confusing colours?'</i> Item similar to 'distinguishing colours' item in the same pool.
Item clarity	Activity Limitation	<i>'communicating in writing, such as handwriting, type-writing and word-processing?'</i> Item confusing and incorporates several tasks each encompassing their own level of difficulty.
Item applicability	Activity Limitation	<i>'reading school reports?'</i> Item too specific to be broadly applicable.
Item frequency & Item relevance	Convenience	<i>'having to stop moving in order to see better?'</i> Item appeared only once and did not occur during qualitative patient interviews.

Table 4 – Examples of item modifications following the cognitive interviews

Quality of life domain	Item	Type of change	Reason for change
Activity Limitation	‘How much difficulty do you have with attending appointments?’	Deleted	Item similar in content to two other items, namely ‘Accessing medical care’ and ‘Organising how to get to and from your eye appointments’
Health Concerns	‘How concerned are you about being able to access services, e.g. low vision care?’	Added	Participant suggestion
Activity Limitation	‘How much difficulty do you have with driving towards the sun or oncoming headlights?’	Split	Item double-barrelled and the ability required by the two tasks interpreted differently by participants. Item split into two separate items: ‘Driving towards the sun’ & ‘Driving towards oncoming headlights’
Emotional	‘How often do you feel like your emotions go up and down?’	Re-phrased	Item confusing and rephrased as ‘Experience mood swings’ to improve clarity