Patient-Reported Outcomes on Quality of Life in Ophthalmology

Martijn Visser

Patient-Reported Outcomes on Quality of Life in Ophthalmology

Martijn Visser

Printed I Ridderprint I www.ridderprint.nl

© 2022 Martijn Stefan Visser, The Netherlands. All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author. Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming van de auteur.

Patient-Reported Outcomes on Quality of Life in Ophthalmology

Kwaliteit van leven uitkomsten in de oogheelkunde

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit van Rotterdam
op gezag van de rector magnificus

Prof.dr. A.L. Bredenoord

en volgens het besluit van Colleges voor Promoties.

De openbare verdediging zal plaatsvinden op
dinsdag 15 november 2022 om 13:00 uur

door

Martijn Stefan Visser
geboren te Hilversum

Erasmus University Rotterdam

Promotiecommissie

Promotoren
prof.dr. J. J. van Busschbach

Overige leden
prof. dr. H.F. Lingsma
prof.dr. J.R. Vingerling
dr. R.P.L. Wisse

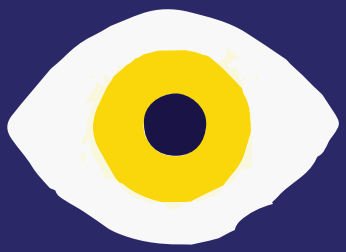
Copromotor
dr. R. Timman

Paranimfen
Mark Visser
Steef Redeker

Table of Contents

CHAPTER 1

General Introduction



1

"Every ophthalmologist is a quality-of-life expert: most of us devote our days not to 'adding years to life', but rather to 'adding life to years'." [1]

The goal of every healthcare professional should be to improve the quality of life (QoL) of his or her patients. Ophthalmic diseases are generally not life-threatening, which makes ophthalmology an area in health that should perfectly suit QoL research. Integrating QoL into ophthalmic care is not as common as one might expect. The outcomes of interventions in ophthalmology are most often expressed in medical terms which focus on clinical visual and refractive outcomes or on the incidence of complications rather than on their meaning for QoL. In other words, the patient is not asked what he or she thinks about the outcome of treatment. This observation does not imply that medical outcomes should be devalued or replaced by QoL measures, but rather that the latter outcomes must be seen as additional to the more familiar medical outcomes. Nevertheless, for now, we can state that the assumption that all ophthalmologists are quality-of-life experts is incorrect, since some measurements undertaken have not met the aims claimed for them. This thesis focuses on the enhancement of QoL measurement in ophthalmology.

The most common ophthalmic diseases are age-related macular degeneration (ARMD) and cataract. ARMD results in blurry vision as a result of leaking blood vessels in the macula of the retina. The deterioration in vision is more or less irrevocable, as photoreceptors are permanently damaged. In cataract the lens is opacified, and treatment focuses on replacing the cloudy lens. In contrast to ARMD with a lasting impact, cataract consequences are generally temporary, as the patient's visual function becomes fully restored. These two major diseases with opposite prognoses make it fascinating to perform QoL research in this area, and thus they are the subject of the research throughout the thesis.

There are several applications in which QoL measurements could be integrated into ophthalmic care, and these measurements can be informative from several stakeholder perspectives.

(i) The first application is in supporting the formulation of clinical guidelines. For instance, in macular degeneration, treatment involves frequent multiple visits, including injections, and there is a risk of complications with every injection. This thesis presents a QoL assessment that helps to determine the optimal injection frequency of macular degeneration treatment.

[1] van den Bos GA Triemstra AH Quality of life as an instrument for need assessment and outcome assessment of health care in chronic patients. Qual Health Care 1999;8247- 252

(ii) Another application is with respect to health policy in relation to reimbursement. When contemplating the level of reimbursement in macular degeneration, there is a challenge in evaluating how to incorporate improvements in the worse seeing eye (WSE). This eye is typically dominated by the other eye, so improvements in this eye have limited overall benefits. In this thesis, we shed light on this health economic issue.
(iii) The final stakeholders are doctors and patients. For example, measuring QoL can facilitate communication between doctors and patients in daily clinical practice, so ophthalmologists can in this situation justifiably state that they are QoL experts. Hence QoL measurement can be appreciated from both doctor and patient perspectives.

Three groups of Health-Related QoL measurements

Measures of health-related QoL (HRQoL) can be i) disease-specific measures, ii) generic health status profiles, or iii) HRQoL utility indexes. These measures serve different purposes.
i) Disease-specific measures are evidently suited for specific diseases, and thus highly sensitive to changes in specific areas of health. They are mainly used to compare different treatments within one specific disease or a specific health condition. In this thesis, measures such as the NEI VFQ-39 and Catquest-9SF instruments fit this description.
ii) Generic health status profiles focus more on patients' general health than on disease-specific considerations. Hence, in addition to mental and physical health, social and emotional health and pain could be incorporated. These generic questionnaires provide a global profile of a patient's health. The SF-36 is an example of such a questionnaire that is discussed in this thesis.
iii) HRQoL utility indexes are questionnaires based on societal preference weights for patients' health states. These preference-weighted QoL scores are summarized in one 'utility score'. Such a score 'U' can be used to 'weight' the health state 'Q'. When this weight is multiplied by the number of life years 'Y' that the health state lasts, this results in the number of Quality Adjusted Life Years (QALYs) = U(Q) x Y. For instance, 4 years in a health state with a QoL utility score of 0.6 = 4 x 0.6 = 2.4 QALYs. As all areas of health can be compared in terms of QoL and life years, QALYs are a generic way of expressing the outcome of health care. Moreover, costs per QALY provide a widely-used generic cost-effectiveness ratio. QALYs are a preferred outcome measure in health economics.
This thesis focuses on the application of all three forms of QoL measure in ophthalmology.

1

Aim

The aim of this thesis is to investigate ways of using QoL measures in ophthalmology in order to aid decision-making in clinical consultations, health policy-making, and for reimbursement purposes.

Outline

The thesis has three parts. The first reports upon two randomized controlled trials (RCTs) aimed at determining the impact of interventions on disease-specific QoL in ophthalmology. **Chapter 2** investigates the effect of different bevacizumab injection frequencies on QoL, using the NEI-VFQ-39 instrument, in mainly elderly patients with ARMD. In **Chapter 3**, low vision spectacles for distance viewing are tested in patients with ARMD. Again, QoL is measured utilizing NEI-VFQ-39 as the primary outcome.

The second part covers the HRQoL utility 'U' measurements described above. **Chapter 4** describes the use of the generic QoL instrument SF-6D in the construction of visual acuity health states based on both the better seeing eye (BSE) and the WSE for patients with age-related macular degeneration. This helps to meet the requirement described above, that economic evaluation should distinguish between improvements in the BSE and WSE.

The third part focuses on the applicability of short, and thus easy-to-use, QoL questionnaires in daily ophthalmic practice, enabling routine outcome monitoring. In **Chapter 5,** a Dutch translation is undertaken for the short patient-reported outcome measure (PROM) Catquest-9SF. This questionnaire is used to quantify benefits in visual functioning from cataract surgery and can be easily used in clinical practice. Validity and test-retest reliability are provided for the Dutch version, and norm scores are calculated. In **Chapter 6,** Catquest-9SF is further investigated in a multicentre study by comparing its results with those of clinical visual and refractive measures. The main factors of significance were performing the surgery in one or two eyes, ocular comorbidity, and pre- and post-operative complications. **Chapter 7** describes the construction of a short version of the general visual QoL questionnaire NEI-VFQ-39. This resulted in a 7-item questionnaire where. by utilizing a computerized system or simply paper-and-pencil, 3 items require to be filled out (VFQ-3oo7) to achieve a representative score with minimal loss of information. A high level of generalizability was reached as a broad range of ophthalmic diseases was studied.

Chapter 8 presents a general discussion, covering a number of applications of HRQoL measures in ophthalmology, and the integration of these measures into ophthalmic care.

1

CHAPTER 2

Six and Eight Weeks Injection Frequencies of Bevacizumab Are Non-Inferior to the Current Four Weeks Injection Frequency For Quality of Life in Neovascular Age-Related Macular Degeneration: A Randomized Controlled Trial

Martijn S. Visser · Sankha Amarakoon · Tom Missotten · Reinier Timman · Jan J. V. Busschbach

Quality of Life Research 2020 29:3305-3313

Abstract

Purpose

Patients with neovascular age-related macular degeneration (nARMD) will not deteriorate on visual acuity and retinal thickness when treated with bevacizumab injection frequencies of 6 or 8 weeks compared to 4 weeks. This study aimed to investigate this non-inferiority in quality of life (QoL). We hypothesized that less frequent bevacizumab injections are not inferior regarding patients reported QoL.

Methods

Patients were randomized to bevacizumab every 4 (n = 64), 6 (n = 63), and 8 weeks (n = 64). Patients were at least 65 years old, have a best-corrected visual acuity of 20/200 to 20/20, no previous ARMD treatment and active leakage. Vision-related QoL questionnaire NEI VFQ-39 was used to assess QoL at baseline and after 1 year. General QoL questionnaire SF-36 was included for secondary analysis. Multilevel analyses were performed, correcting for age, gender and baseline.

Results

The 6 (3.68; 95% CI − 0.63 to 8.00) and 8 (2.15; 95% CI − 2.26 to 6.56) weeks bevacizumab regimens resulted in non-inferior QoL differences compared to 4 weeks on the NEI VFQ-39. Also on the SF-36 the differences were well within the non-inferiority limits.

Conclusion

Non-inferiority of the 6 and 8 weeks frequencies was demonstrated compared to 4 weeks on vision-related and general QoL in patients with nARMD. These results are in line with previously published results of lower frequency injections regarding visual acuity and central retinal thickness. Lower injection frequency may reduce burden, side effects, and treatment costs. In consideration of these results, 8 weeks frequency injections of intravitreal bevacizumab could be considered in patients with nARMD.

Introduction

Age-related macular degeneration (ARMD) is the leading cause of severe vision loss and blindness among people aged over 50 years in Western countries [1, 2]. ARMD affects central retinal function, profoundly impairing the patient's ability to perform daily activities and their quality of life (QoL) [3]. Exudative ARMD, an aggressive form of ARMD [4, 5], progresses rapidly and is characterized by the development of choroidal neovascularization (CNV); hence, it is often described as neovascular ARMD (nARMD). The current standard therapy for nARMD is intravitreal injection of anti-vascular endothelial growth factor (VEGF), a treatment which improves the visual prognosis of nARMD patients considerably.

To enhance effective patient-centered care, there is a trend toward gathering outcome information from the patient's perspective in addition to the clinical outcomes. Since there is interest in the patients' perspective of satisfaction, in terms of outcome, several patient-reported outcome measures (PROMs) have been developed [6]. Several studies have suggested that the use of PROMs have a positive effect on the doctor-patient communication, and consequently patients' satisfaction [7]. The most commonly used anti-VEGF medications are ranibizumab, aflibercept and bevacizumab. The efficacy of ranibizumab and aflibercept has been proven and appear clinically equivalent, and are approved both by the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) for intraocular use in nARMD [8–13]. Bevacizumab has been approved by the FDA and the EMA for the treatment of various tumors, such as colorectal cancer [14], but not specifically for nARMD. However, in recent years, ophthalmologists have been prescribing bevacizumab for off-label use in nARMD because it is a cost-effective substitute for ranibizumab and aflibercept [15–20]. Multiple studies provided RCT evidence supporting the efficacy of bevacizumab in a monthly, pro re nata and treat-and-extend regimes [15–20]. The CATT study also showed that there is no difference in effectiveness in term of vision and side effects between ranibizumab and bevacizumab and is comparably effective when the injection frequency is 4 weeks. Moreover, the IVAN study showed similar results on QoL for bevacizumab and ranibizumab measured with the Euro- Qol-5D [21], macular disease-specific quality of life [22] and treatment satisfaction [23].

The every-four-weeks regimen used in the CATT study was chosen for bevacizumab based on prior ranibizumab trials and is a widely adopted and proven strategy. However, the relatively long half-life of bevacizumab might allow the achievement of a therapeutic effect with less frequent injections, as has been the experience in the clinic [24, 25]. Reduced numbers of injections could have several beneficial effects, including a decrease in the risks associated with intravitreal injection (such

as endophthalmitis and retinal detachment), improved cost-effectiveness, reduced patient burden, and a reduced ophthalmic work-load. A study in nARMD patients comparing an every-four-weeks injection frequency of bevacizumab therapy to an every-six-weeks or every-eight-weeks injection frequency showed no significant difference for lower injection frequencies for visual acuity and central retinal thickness [26]. In the current non-inferiority study, we aimed to determine whether bevacizumab therapy administered every 6 or 8 weeks is also not inferior to an every-four-weeks regimen for QoL outcomes in nARMD patients.

Materials and methods

Study patients

This is a secondary analysis of an RCT comparing three treatment regimens of bevacizumab (Avastin) for the treatment of ARMD on visual acuity and central retinal thickness [26]. A total of 191 patients were enrolled in a 1-year, prospective, open-label RCT which investigated the optimal injection frequency of bevacizumab injection for ARMD treatment at the Rotterdam Eye Hospital from June 2008 to March 2010 (Figure 1). To be eligible, patients had to be at least 65 years old, have a best-corrected visual acuity of 20/200 to 20/20 (Snellen equivalent) in the study eye as assessed using Early Treatment Diabetic Retinopathy Charts (ETDRS), no previous ARMD treatment and active leakage. Patients were only treated in one eye. Fluorescein angiography (FA) and indocyanine green (ICG) angiography were used to observe leakage, and optical coherence tomography (OCT) was used to observe the presence of fluid [26]. Patients who had other significant ocular disorders, had allergies to either FA or ICG dye injections, were immunocompromised, using coumarin-derivatives, had experienced a clinically significant cerebrovascular accident or myocardial infarction or had a planned ocular surgery during the 1-year follow-up, were excluded. Written informed consent was obtained from all participants. After baseline measurements were completed, all eligible patients were randomized to an injection frequency of every 4, 6, or 8 weeks using a computer-based 1:1:1 ratio block randomization procedure.

Treatment

Apart from the difference in frequency, treatment regimens were comparable among the three groups. At each outpatient visit, a dose of 1.25 mg bevacizumab was administered intravitreally. On top of the measures during regular outpatient visits, patients were assessed every 12 weeks by best-corrected visual acuity, spectral-domain OCT and funduscopy. Monthly checks for adverse events took place by questioning patients. Treatment was continuous for 1 year, independent

of visual acuity change, spectral-domain OCT measures, or funduscopy findings. The 4 weeks, 6 weeks, and 8 weeks bevacizumab treatment regimens resulted in totals of 13, 9, and 7 injections and visits a year, respectively.

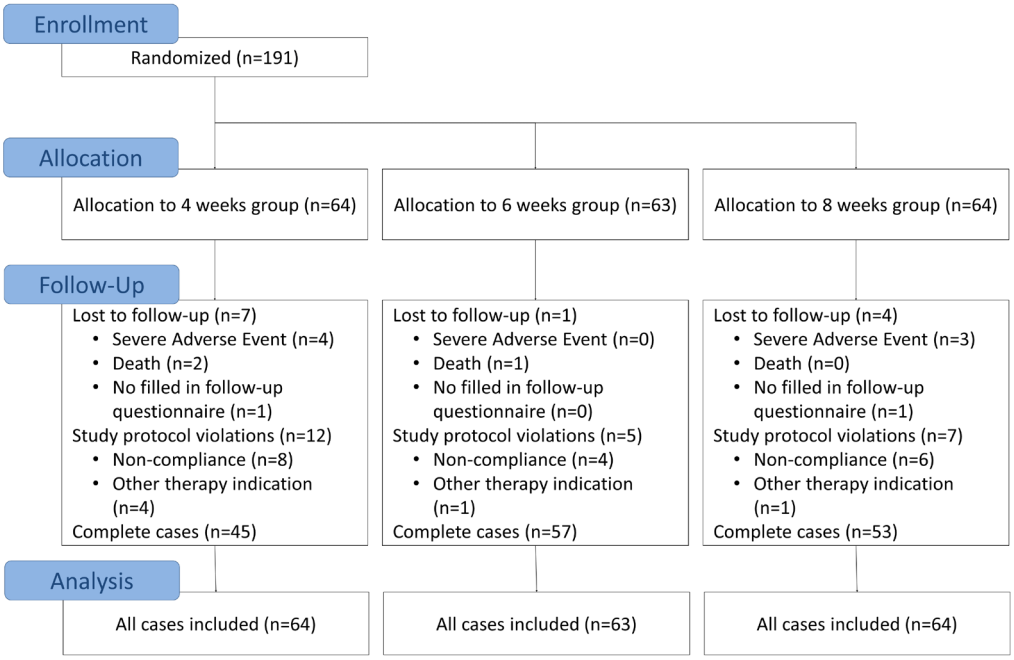


Figure 1 CONSORT flow diagram of enrolment, allocation, follow-up and analysis of the every-four-weeks, every-six-weeks, and every-eightweeks treatment groups [38]

Outcome measures

At baseline and at the final follow-up visit, patients were asked to complete the National Eye Institute 39-Item Visual Function Questionnaire (NEI VFQ-39) [27] and the 36-item Medical Outcomes Study Short-Form General Health Survey (SF-36) [28, 29]. The NEI VFQ-39 assesses visionrelated QoL, while the SF-36 evaluates general QoL. Given the nature of the disease, both questionnaires were presented in a larger font size and often administered in the presence and sometimes with support of a caregiver and/or family member.

Vision-related quality of life: NEI VFQ-39

The primary outcome was vision-related QoL, measured as the composite score on the NEI VFQ-39 [27]. The NEI VFQ-39 consists of a 25-item base set of questions and 14 supplemental items. All items use a Likert-type scaling and five response categories, with occasionally a sixth category to opt out, except for two items that have 10 response options. Responses are converted into 12 vision-targeted multi-item

subscales (0–100): general health, general vision, ocular pain, near activities, distant activities, social functioning, mental health, role limitations, dependency, driving, color vision, and peripheral vision. These 12 subscales can be summarized as a single composite score. A 10-point difference in either the sub-scales or the composite score of the NEI VFQ-39 is deemed clinically important, and thus considered a clinically meaningful change [30, 31]. The reliability of the NEI VFQ-39 in age-related macular degeneration varies from a Cronbach's alpha of 0.86 to 0.96 [32, 33].

General quality of life: SF-36

Another outcome measure was general QoL measured by the SF-36 [29]. This is a self-report questionnaire comprising 36 questions measuring different aspects of general health. All items use a Likert based scaling and use two to six response options. The responses are converted into eight multi-item subscales: physical functioning, role functioning physical, bodily pain, general health, vitality, social functioning, role functioning emotional, and mental health. These scales can be summarized as a psychometrically based 'physical component summary' (PCS), in which the first four scales are most heavily weighted, and a 'mental component summary' (MCS), in which the last four scales are most heavily weighted [34]. These summaries are transformed into T-scores with a mean of 50 and standard deviation of 10. Higher scores on SF-36 scales indicate a better quality of life. The UK version reliability of the physical subscale is 0.92, and the mental subscale is 0.89 [34]. Following the approach provided by Jacobson & Truax, the clinical significant change is 7.84 and 9.19 for the respective subscales [35].

Data analysis and statistical methods

Differences between dropouts and retained patients were analyzed with Student's *t*- and chi square-tests. Baseline differences for continuous variables between the three groups were analyzed with One-way ANOVA with Bonferroni correction for pairwise differences. Chi square-tests were applied for binary variables and when significant, standardized residuals were evaluated to determine the deviating groups. The non-inferiority limit for the 6 weeks and 8 weeks groups comparison with the 4 weeks group was based on the 10-point clinical significant difference of the NEI VFQ, composite score and the subscales near vision, distance vision and role limitations. This negative 10-point difference indicated the lower end of the 'region of therapeutic equivalence' and, together with the maximum possible difference, enclosed the 'region of non-inferiority' [36]. The region of non-inferiority ranged from − 10 to 100. Non-inferiority was assumed whenever the 95% confidence interval of the difference in change fell entirely within this region

[36]. Note that only the right-hand side of the distribution was relevant, Figure 2.
In addition, differences between treatment groups were tested for the secondary SF-36 subscales. We applied multilevel linear regression analyses to evaluate differences in change in QoL between the three randomization groups. The patients formed the upper level, their repeated measures the lower level. These analyses can handle data with missing time points efficiently, i.e. data of patients without a followup can be included, without a need for imputation. For each outcome we applied a separate model. The random parts of the models only included the intercept. The fixed parts of the models included time (follow-up vs. baseline), centered baseline score, 6-weeks and 8-week frequencies and the interaction of time with baseline, six and eight weeks frequencies. The four-week frequency group served as reference group. In all analyses, gender and age were included as control variables.
The study was originally designed to detect differences in visual acuity, and subsequently powered with a noninferiority limit of seven letters [26]. When testing QOL, a power analysis for non-inferiority was performed on the NEI VFQ-39 composite score. The clinical important difference for the NEI VFQ-39 is 10 and the standard deviation is 20, the one-sided alpha was set at 0.05 and power at 0.80, for which a sample size of 50 persons per group is needed. This implies that the sample size of 63–64 is sufficient. All other analyses were performed with IBM SPSS version 24.0 "IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp." This study was approved by the Erasmus Medical Research Ethics Committee (MEC-2007-254) in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and was registered in the Dutch Trial Register (NTR 1174).

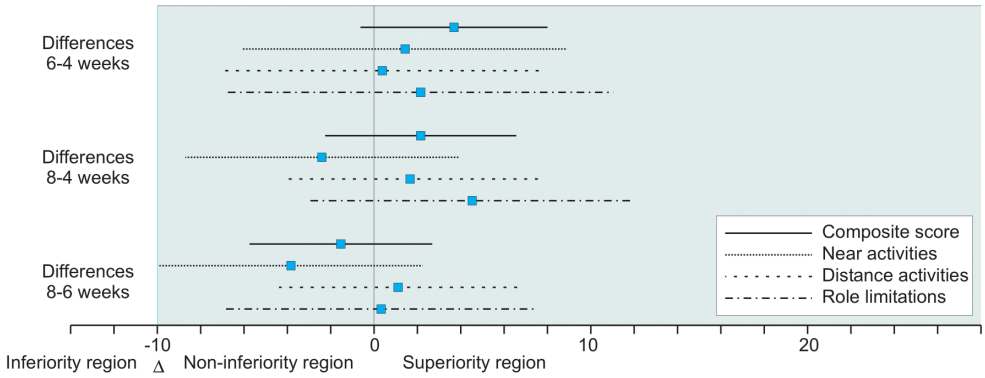


Figure 2 Forest plot of 95% confidence intervals of differences between treatment groups. The sensitivity analysis was based on a matched sample

Table 1 Baseline characteristics

	Every 4 weeks (*n* = 64)	Every 6 weeks (*n* = 63)	Every 8 weeks (*n* = 64)	*p* value
Characteristics				
Age in years at baseline, mean ± SD	76.5 ± 6.8	77.4 ± 6.7	78.1 ± 6.1	0.436
Gender, male *n* (%)	18 (28.1)	25 (39.7)	21 (32.8)	0.382
Race, Caucasian *n* (%)	63 (98.4)	63 (100)	64 (100)	0.369
Visual acuity score (no. letters)	66 ±12	65 ±13	62 ±15	0.230
Total thickness at fovea, µm ± SD	369 ±85	371 ± 97	371 ± 97	0.990
Patients treated in worse eye, *n* (%)[a]	30 (56.6)	31 (56.4)	30 (53.6)	0.955
NEI VFQ-39, mean ± SD[b, e]				
Composite score	72.0 ± 17.6	67.8 ± 20.0	63.1 ± 19.4	0.032 [f]
Near Activities	60.5 ± 24.2	57.1 ± 24.7	49.4 ± 26.7	0.041 [f]
Distant Activities	67.8 ± 23.6	64.3 ± 25.2	57.9 ± 25.1	0.073
Role Limitations	64.2 ± 25.9	60.2 ± 27.2	52.6 ± 25.5	0.042 [f]
SF-36, mean ± SD[b, c, e]				
Physical component	44.8 ± 10.9	42.1 ± 11.1	42.2 ± 9.2	0.288
Mental component	50.9 ± 9.1	51.5 ± 11.5	48.4 ± 11.1	0.239
Lost to follow-up, *n* (%)				
Exit reason				
SAE[d]	4 (6.3)	0 (0.0)	3 (4.7)	0.150
Death	2 (3.1)	1 (1.6)	0 (0.0)	0.364
Non-compliance	8 (12.5)	4 (6.3)	6 (9.4)	0.495
No filled in follow-up questionnaire	1 (1.6)	0 (0.0)	1 (1.6)	0.608
Other therapy indication	4 (6.3)	1 (1.6)	1 (1.6)	0.217
Total	19 (29.7)	6 (9.5)	11 (17.2)	0.013 [g]

[a] The treatment eye was defined as the worse-seeing eye when the visual acuity letter score at baseline was worse by five or more letters compared to that for the fellow eye. Patients with missing visual acuity (VA) scores or similar VA scores within a 5 letter range, were omitted, resulting in *n* = 53, 55, and 56, respectively [39].

[b] Higher scores indicate a better quality of life.

[c] In the every 4 weeks group $n = 63$.
[d] SAE = Severe Adverse Event
[e] The baseline scores were included in the multilevel model to adjust for potential differences
[f] The difference was between the every 4-8 weeks $p<0.05$ (Bonferroni correction)
[g] Every 4 weeks is overrepresented

Results

Demographic and clinical characteristics

After randomization, 64 patients were treated in the 4 weeks group, 63 in the 6 weeks group, and 64 in the 8 weeks group. Treatment arms were well balanced with regard to baseline demographic characteristics, visual acuity, and other characteristics of the affected eye (Table 1). However, significant baseline differences were present for the NEI VFQ-39 as the 8 weeks group had lower scores than the 4 weeks group.

Dropouts

Patients lost to follow-up were subdivided based on their exit reasons (Table 1). The highest drop-out rate in the 4 weeks treatment group (29.7%) and the lowest in the 6 weeks group (9.5%) significantly differed, $p = 0.004$. Patients who dropped out had significantly worse baseline scores than retained patients on the physical component summary of the SF-36: $t(176) = -2.95$, $p = 0.004$ (not in Table 1). No other statistical significant differences were found.

NEI VFQ-39

The changes and differences estimated by the multilevel models are presented in Table 2, the total models are presented in Table 3. Observed differences are presented in Appendix 1 and the observed means and standard deviations in Appendix 2. The 95% confidence intervals of the difference in change scores showed that the composite score interval was well inside the [− 10, 100] point difference interval that represented the non-inferiority region for the three treatment comparisons (Figure 2). For the subscales near activities, distant activities, role limitations, visual functioning and socio-emotional functioning the 95% confidence intervals of the differences were also entirely within the region of non-inferiority. This also barely holds for the near activities estimate for gain within 6 weeks (10.26) compared to gain within 8 weeks (6.43). This 95% confidence interval of − 9.91 to 2.24 is just within the limit.

Table 2 Estimated changes in the NEI VFQ-39 and SF-36 scores, age, gender and baseline controlled score.

Every	4 weeks	6 weeks	8 weeks	Differences 6-4 weeks		Differences 8-4 weeks		Differences 8-6 weeks	
	Estimate [95% CI]				*p* value	Estimate [95% CI]	*p* value	Estimate [95% CI]	*p* value
NEI VFQ-39									
Composite score	2.91 [-0.25, 6.07]	6.59 [3.65, 9.53]	5.05 [2.04, 8.07]	3.68 [-0.63, 8.00]	0.094	2.15 [-2.26, 6.56]	0.338	-1.53 [-5.75, 2.68]	0.474
Near Activities	8.83 [4.33, 13.34]	10.26 [6.01, 14.51]	6.43 [2.10, 10.75]	1.43 [-4.76, 7.62]	0.650	-2.41 [-8.71, 3.89]	0.452	-3.84 [-9.91, 2.24]	0.215
Distance Activities	5.13 [0.90, 9.37]	5.51 [1.56, 9.47]	6.80 [2.74, 10.85]	0.38 [-5.41, 6.17]	0.898	1.66 [-4.24, 7.56]	0.580	1.28 [-4.39, 6.60]	0.656
Role Limitations	1.63 [-3.73, 6.98]	5.83 [0.84, 10.82]	6.15 [1.04, 11.27]	4.21 [3.11, 11.52]	0.259	4.53 [-2.94, 11.99]	0.234	0.32 [-6.83, 7.48]	0.929
SF-36									
Physical component	-0.88 [-3.00, 1.23]	-0.53 [-2.49, 1.42]	-0.25 [-2.26, 1.76]	0.35 [-2.52, 3.22]	0.810	0.63 [-2.29, 3.56]	0.670	0.28 [-2.52, 3.09]	0.842
Mental component	3.20 [1.13, 5.27]	0.85 [-1.09, 2.78]	2.66 [0.66, 4.66]	-2.35 [-5.18, 0.47]	0.103	-0.54 [-3.42, 2.35]	0.715	1.82 [-0.97, 4.60]	0.201

Note: non-inferiority means that the lower boundary of the 95% CI is not lower than minus 10 for the NEI-VFQ composite score or subscales and *p*-values indicate whether the difference of changes is different from zero

Table 3 Multilevel VFQ-39 and SF-36 models

	NEI VFQ-39								SF-36			
	Composite score		Near activities		Distance activities		Role limitations		Physical component		Mental component	
	Estimate	*p* value	Estimate	*p* value	Estimate	*p* value	Estimate	*p* value	Estimate	*p* value	Estimate	*p* value
Intercept	67.02	<0.001	54.70	<0.001	62.37	<0.001	58.02	<0.001	42.94	<0.001	50.13	<0.001
Male	2.21	0.024	3.40	0.014	3.36	0.012	3.49	0.030	0.56	0.383	0.21	0.739
Age	0.00	0.991	-0.09	0.371	-0.07	0.511	0.00	0.984	0.05	0.303	-0.08	0.080
Time	2.91	0.071	8.83	<0.001	5.13	0.018	1.63	0.551	-0.88	0.412	3.20	0.003
Baseline	1.00	<0.001	0.99	<0.001	0.99	<0.001	0.99	<0.001	1.00	<0.001	1.00	<0.001
Time * baseline	-0.32	<0.001	-0.29	<0.001	-0.31	<0.001	-0.38	<0.001	-0.28	<0.001	-0.54	<0.001
Every 6 weeks	-0.21	0.885	-0.25	0.907	-0.21	0.917	-0.37	0.881	-0.14	0.887	0.09	0.920
Time * 6 weeks	3.68	0.094	1.43	0.650	0.38	0.898	4.21	0.259	0.35	0.810	-2.35	0.103
Every 8 weeks	-0.13	0.930	-0.14	0.946	-0.20	0.920	-0.18	0.942	-0.12	0.900	0.16	0.869
Time * 8 weeks	2.15	0.338	-2.41	0.452	1.66	0.580	4.53	0.234	0.63	0.670	-0.54	0.715

SF-36

The treatment did not significantly affect the SF-36 component summaries. All treatment effects of different injection frequencies were well within the non-inferiority limits (Table 2).

Discussion

To study non-inferiority of a less frequent injection schedule for bevacizumab therapy, we tested QoL in 191 ARMD patients who were randomly assigned to receive 1 year of continuous treatment with intravitreal bevacizumab
injections every 4, 6, or 8 weeks. In this study we showed that 6 weeks and 8 weeks injection regimens were not inferior to the four-week regimen in QoL assessments. The eight-week regimen was also not inferior to the six-week regimen. Thus, regarding patient satisfaction there is no objection to reduce the frequency of the injection to eight instead of 4 weeks. This is in line with the former results of our study group, where no effects of a lower injection frequency on visual acuity and central retinal thickness were observed [26].

In daily ophthalmic care the fixed regimen as examined in this study is not routine clinical practice. The treat-andextent regimen is accepted as the preferred practice, in which, after an initial induction phase, the next treatment interval is extended as long as the patient shows no symptoms of relapse. A lower injection frequency may reduce the burden for patient and doctor, the chances of injectionrelated side effects, and treatment costs. Hereby, the biggest fear of extending treatment interval is that in the meanwhile the dormant disease will flame up and cause irreversible vision loss. The current challenge is to find the right balance in treating, waiting and adjusting. Another way to reduce burden is to determine whether the initial 4 weeks injection interval used with treat-and-extend could be perhaps 6 or 8 weeks. This current study implicates that there is room to investigate this statement. For an 8 weeks pro re nata, on demand, versus a 4 weeks pro re nata regimen no significant difference was shown [39]. In consideration of these results, low frequency injections (in particular every 8 weeks) of intravitreal bevacizumab should not be withheld from patients with nARMD.

Strengths and limitations

The every-four-weeks regimen group had the highest dropout rate. However, it is unlikely that this higher drop-out rate jeopardizes the conclusion, as drop-outs tended to have the same baseline values. The main reasons for treatment discontinuation in all groups were compliance-related study visit violations. The noncompliance is not only an issue in this study but a problem also in clinical

practice [37]. In this study, we see a slightly higher, though not significant, nonadherence rate with the most rigorous treatment schedule, which may be a justification for considering a lower treatment frequency as alternative, as this may increase patient compliance. But where some see frequent visits as a hassle, others will see it as a welcome social benefit. In the end, again, more personalized care might be the answer. Imbalances were found in the vision-related QoL baseline scores. Principally these differences are a coincidental result of randomization, but as it might have affected the results, the positive effect of the treatment was larger in the eightweek group, we corrected for baseline in the model. In this analysis the interaction between baseline and time confirms the influence of an imbalanced baseline. Apparently, patients with lower baseline scores on average have larger increase in QoL. This could logically be a result of regression to the mean. This same situation occurred in the previous study where the difference of 4 letters on baseline was equalized at follow-up [26]. It is obviously more difficult to improve more if you already have a high QoL.

Conclusion

Non-inferiority of the 6 and 8 weeks frequencies to 4 weeks was demonstrated on vision-related and general QoL in patients with nARMD. These results are in line with previously published results of these frequency injections. Lower injection frequency may reduce burden, side effects, and treatment costs. In consideration of these results, 6 and in particular 8-week frequency injections of intravitreal bevacizumab could be considered in patients with nARMD.

References

1. Bressler NM (2004). Age-related macular degeneration is the leading cause of blindness. JAMA, 291, 1900–1901.

2. Klaver CC, Wolfs RC, Vingerling JR, Hofman A & de Jong PT (1998). Age-specific prevalence and causes of blindness and visual impairment in an older population: The Rotterdam Study. Archives of Ophthalmology, 116, 653–658.

3. Taylor DJ, Hobby AE, Binns AM & Crabb DP(2016). How does age-related macular degeneration affect real-world visual ability and quality of life? A systematic review. British Medical Journal Open, 6, e011504.

4. Bird AC (1993). Choroidal neovascularisation in age-related macular disease. The British Journal of Ophthalmology, 77, 614–615.

5. Bird AC (1999). Age-related macular disease: Aetiology and clinical management. Community Eye Health, 12, 8–9.

6. McAlinden C, Gothwal VK, Khadka J et al. (2011). A head-to-head comparison of 16 cataract surgery outcome questionnaires. Ophthalmology, 118, 2374–2381.

7. Nilsson E, Orwelius L & Kristenson M (2016). Patientreported outcomes in the Swedish National Quality Registers. Journal of Internal Medicine, 279, 141–153.

8. Brown DM, Michels M, Kaiser PK et al. (2009). Ranibizumab versus verteporfin photodynamic therapy for neovascular age-related macular degeneration: Two-year results of the ANCHOR study. Ophthalmology, 116(57–65), e5.

9. Heier JS, Brown DM, Chong, V et al. (2012). Intravitreal aflibercept (VEGF trap-eye) in wet age-related macular degeneration. Ophthalmology, 119, 2537–2548.

10. Regillo CD, Brown DM, Abraham P et al. (2008). Randomized, double-masked, sham-controlled trial of ranibizumab for neovascular age-related macular degeneration: PIER Study year 1. American Journal of Ophthalmology, 145, 239–248.

11. Rosenfeld PJ, Brown DM, Heier JS et al. (2006). Ranibizumab for neovascular age-related macular degeneration. New England Journal of Medicine, 355, 1419–1431.

12. Schmidt-Erfurth U, Eldem B, Guymer R et al. (2011). Efficacy and safety of monthly versus quarterly ranibizumab treatment in neovascular age-related macular degeneration: The EXCITE study. Ophthalmology, 118, 831–839.

13. Stewart MW, Rosenfeld PJ, Penha FM et al. (2012). Pharmacokinetic rationale for dosing every 2 weeks versus 4 weeks with intravitreal ranibizumab, bevacizumab, and aflibercept (vascular endothelial growth factor Trap-eye). Retina, 32, 434–457.

14. Cohen MH, Gootenberg J, Keegan P & Pazdur R (2007). FDA drug approval summary: Bevacizumab plus FOLFOX4 as second-line treatment of colorectal cancer. The Oncologist, 12, 356–361.

15. Berg K, Hadzalic E, Gjertsen I et al. (2016). Ranibizumab or bevacizumab for neovascular age-related macular degeneration according to the Lucentis compared to Avastin study treat-andextend protocol: Two-year results. Ophthalmology, 123, 51–59.

16. Berg K, Pedersen TR, Sandvik L & Bragadottir R (2015). Comparison of ranibizumab

and bevacizumab for neovascular age-related macular degeneration according to LUCAS treat-andextend protocol. Ophthalmology, 122, 146–152.

17. CATT Research Group, Martin DF, Maguire MG et al. (2011). Ranibizumab and bevacizumab for neovascular agerelated macular degeneration. New England Journal of Medicine, 364, 1897–1908.

18. Investigators IS, Chakravarthy U, Harding SP et al. (2012). Ranibizumab versus bevacizumab to treat neovascular age-related macular degeneration: One-year findings from the IVAN randomized trial. Ophthalmology, 119, 1399–1411.

19. Kodjikian L, Souied EH, Mimoun G et al. (2013). Ranibizumab versus bevacizumab for neovascular age-related macular degeneration: Results from the GEFAL noninferiority randomized trial. Ophthalmology, 120, 2300–2309.

20. Martin DF, Maguire MG & Fine SL (2011). Bevacizumab: Not as good with more adverse reactions? Response. Clinical and Experimental Ophthalmology, 39, 718–720.

21. Dolan P (1997). Modeling valuations for EuroQol health states. Medical Care, 35, 1095–1108.

22. Mitchell J & Bradley C (2004). Design of an individualized measure of the impact of macular disease on quality of life (the MacDQoL). Quality of Life Research, 13, 1163–1175.

23. Mitchell J, Brose L & Bradley C (2007). Design of a measure of satisfaction with treatment for macular degeneration (MacTSQ). Quality of Life Research, A120, 2.

24. Krohne TU, Eter N, Holz FG & Meyer CH (2008). Intraocular pharmacokinetics of bevacizumab after a single intravitreal injection in humans. American Journal of Ophthalmology, 146, 508–512.

25. Krohne TU, Liu Z, Holz FG & Meyer CH (2012). Intraocular pharmacokinetics of ranibizumab following a single intravitreal injection in humans. American Journal of Ophthalmology, 154(682–6), e2.

26. Lushchyk T, Amarakoon S, Martinez-Ciriano JP et al. (2013). Bevacizumab in age-related macular degeneration: A randomized controlled trial on the effect of injections every 4 weeks, 6 weeks and 8 weeks. Acta Ophthalmologica, 91, e456–e461.

27. Mangione CM, Lee PP, Gutierrez PR et al. (2001). Development of the 25-item National Eye Institute Visual Function Questionnaire. Archives of Ophthalmology, 119, 1050–1058.

28. Aaronson NK, Muller M, Cohen PD et al. (1998). Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. Journal of Clinical Epidemiology, 51, 1055–1068.

29. Ware JE (2000). SF-36 health survey update. Spine (Phila Pa 1976), 25, 3130–3139.

30. Lindblad AS & Clemons TE (2005). Responsiveness of the National Eye Institute Visual Function Questionnaire to progression to advanced age-related macular degeneration, vision loss, and lens opacity: AREDS Report no. 14. Archives of Ophthalmology, 123, 1207–1214.

31. Miskala PH, Hawkins BS, Mangione CM et al. (2003). Responsiveness of the National Eye Institute Visual Function Questionnaire to changes in visual acuity: Findings in patients with subfoveal choroidal neovascularization—SST Report No. 1. Archives of

Ophthalmology, 121, 531–539.

32. Clemons TE, Chew EY, Bressler SB et al. (2003). National Eye Institute Visual Function Questionnaire in the Age-Related Eye Disease Study (AREDS): AREDS Report No. 10. Archives of Ophthalmology, 121, 211–217.

33. Revicki DA, Rentz AM, Harnam N et al. (2010). Reliability and validity of the National Eye Institute Visual Function Questionnaire-25 in patients with agerelated macular degeneration. Investigative Ophthalmology & Visual Science, 51, 712–717.

34. Ware JE, Kosinski M & Keller SD (1994). SF-36 physical and mental health summary scales: A user's manual. Boston, MA: New England Medical Center, The Health Institute.

35. Jacobson NS & Truax P (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology, 59, 12–19.

36. Lesaffre E (2008). Superiority, equivalence, and non-inferiority trials. Bulletin of the NYU Hospital for Joint Diseases, 66, 150–154.

37. Obeid A, Gao X, Ali FS et al. (2018). Loss to follow-up among patients with neovascular age-related macular degeneration who received intravitreal anti-vascular endothelial growth factor injections. JAMA Ophthalmology, 136, 1251–1259.

38. Piaggio G, Elbourne DR, Pocock SJ et al. (2012). Reporting of noninferiority and equivalence randomized trials: Extension of the CONSORT 2010 statement. JAMA, 308, 2594–2604.

39. Amarakoon S, Martinez-Ciriano JP, Ingeborg van den Born L et al. (2019). Bevacizumab in agerelated macular degeneration: a randomized controlled trial on the effect of on-demand therapy every 4 or 8 weeks. Acta Ophtalmologica, 97, 107–112.

Appendix 1 Parameters of multilevel models of NEI VFQ-25 Quality of Life scores between the E-Scoop and control group

Rasch score			
Effect	**estimate**	**[95% CI]**	**p-value**
Intercept	-.079	[-.480 – .322]	0.678
Time	-.157	[-.355 – .041]	0.119
E-Scoop	.168	[-.164 – .501]	0.320
Time*E-Scoop	.115	[-.169 – .399]	0.425
Visual Acuity	-3.746	[-5.139 – -2.354]	<0.001
Time*Visual Acuity	.683	[-.433 – 1.799]	0.228
E-Scoop*Visual Acuity	-.709	[-2.578 – 1.160]	0.455
Time*E-Scoop*Visual Acuity	-.264	[-1.824 – 1.296]	0.739
Visual Acuity²	5.054	[1.701 – 8.406]	0.003
Time*Visual Acuity²	.289	[-2.399 – 2.976]	0.832
E-Scoop*Visual Acuity²	.408	[-4.515 – 5.332]	0.870
Time*E-Scoop*Visual Acuity²	-.990	[-5.022 – 3.042]	0.628
Age	.005	[-.016 – .027]	0.624
Time*Age	-.004	[-.021 – .014]	0.684
E-Scoop*Age	-.010	[-.038 – .019]	0.510
Time*E-Scoop*age	.000	[-.023 – .023]	0.993
Gender	-.807	[-1.224 – -.389]	<0.001
Time*Gender	.268	[-.072 – .609]	0.122
E-scoop*Gender	.619	[.044 – 1.194]	0.035
Time*E-Scoop*Gender	-.169	[-.647 – .309]	0.486
Classical score			
Intercept	55.275	[49.102 – 61.447]	<0.001
Time	3.830	[.830 – 6.831]	0.013
E-scoop	-2.224	[-7.490 – 3.043]	0.406
Time*E-Scoop	-3.662	[-7.968 – .643]	0.095
Visual Acuity	51.880	[29.846 – 73.915]	<0.001
Time*Visual Acuity	-6.889	[-23.805 – 10.027]	0.422
E-Scoop*Visual Acuity	9.790	[-19.700 – 39.279]	0.514
Time*E-Scoop*Visual Acuity	-2.601	[-26.254 – 21.052]	0.828
Visual Acuity²	-65.068	[-118.035 – -12.102]	0.016

Time*Visual Acuity²	-13.285	[-54.022 – 27.452]	0.520
E-Scoop*Visual Acuity²	-13.365	[-90.885 – 64.155]	0.734
Time*E-Scoop*Visual Acuity²	37.165	[-23.978 – 98.307]	0.232
Age	.000	[-.344 – .343]	0.998
Time*Age	-.077	[-.344 – .190]	0.571
E-Scoop*Age	.095	[-.353 – .543]	0.677
Time*E-Scoop*Age	.156	[-.194 – .507]	0.379
Gender	11.309	[4.724 – 17.893]	0.001
Time*Gender	-4.227	[-9.387 – .933]	0.108
E-Scoop*Gender	-6.936	[-16.002 – 2.130]	0.133
Time*E-Scoop*Gender	4.334	[-2.915 – 11.584]	0.239

The base group is without E-scoop, the base gender is female and we centered age at 79.5 and visual acuity 0.30 LogMAR. Visual Acuity^2 is entered in the model as a quadratic term.

CHAPTER 3

Randomized Controlled Trial of a Spectacle Lens for Macular Degeneration

Martijn S. Visser · Reinier Timman · Joanneke Kampen-Smalbrugge · Karin Buis · Jan Roelof Polling · Jan J. van Busschbach

Optometry and Vision Science 2020;97:889-897

Abstract

Significance

E-Scoop, a spectacle lens, provides no clinically relevant improvements on quality of life, visual acuity, and contrast sensitivity for patients with AMD. Because patients' burden is high and therapeutic options are scarce, the incentive to develop effective vision rehabilitation interventions remains.

Purpose

Patients with AMD experience low quality of life due to vision loss, despite angiogenesis inhibitor interventions that slow down progression for some patients. E-Scoop, which includes low-power prisms, 6% magnification, yellow tint, and antireflection coating, might aid in daily activities by improving distance viewing. Separately, these features have little proven effectiveness. E-Scoop has not been formally tested. This study aimed to determine the impact of E-Scoop on quality of life and the effect on visual acuity and contrast sensitivity.

Methods

In this randomized controlled, open-label trial, 190 of 226 eligible patients were included. The primary outcome was quality of life measured with the 25-item National Eye Institute Visual Function Questionnaire. Secondary outcomes were visual acuity and contrast sensitivity. The follow-up for quality of life was after 6 weeks for controls and after 3 weeks of use for E-Scoop wearers. The visual measures were repeated after 6 weeks, with optimal refractive correction, with and without E-Scoop.

Results

Randomization resulted in 99 E-Scoop and 86 control group patients for intention-to-treat analysis. No differential change was found between the E-Scoop and control groups on the 25-item National Eye Institute Visual Function Questionnaire using Rasch analysis (Cohen $d = -0.07$, $p = .53$). Statistically significant but small effects were found in favor of E-Scoop on binocular visual acuity (mean difference, 0.05 logMAR [2.5 letters, $p < .001$]) and contrast sensitivity (mean difference, 0.10 logCS [2 letters, $p < .001$]).

Conclusion

No effect of E-Scoop on quality of life was found. E-Scoop showed effects that were statistically significant, although not clinically meaningful and within typical variability, on visual measures.

Introduction

In patients with AMD, quality of life decreases especially when the center of the macula is moderately or severely affected [1]. AMD also increases the risk of depression, falls, fractures, a shorter independent life, and many other limitations in daily life [2]. Although pharmacological interventions have improved over the last decades, patients can still experience irreversible vision loss [3]. In patients with AMD, a decrease in visual acuity is associated with an increase in the use of low-vision aids [4,5]. These aids are relatively effective but are more commonly intended for tasks at near distances. Effective options for distance viewing do exist, however, for example, handheld, head-mounted, or spectacle-mounted telescopic systems [6]. Nonetheless, these aids are often rejected by the patients because of their appearance and limited field of view. Amore user-friendly alternative might be E-Scoop. This patented spectacle lens shown in Figure 1 has a low-power prism of 4^{Δ}, 6^{Δ}, or 8^{Δ}, intended to redirect the image of the object of interest to less damaged parts of themacula; 6% magnification; a yellow tint; and an antireflection coating. E-Scoop is already being prescribed to patients in the hope of improving vision. E-Scoop lenses have not been studied specifically, but previous studies have reported the effects of the separate features of these lenses. In the past, it had been proposed that spectacles with prisms might be useful to improve distance acuity in AMD, redirecting the light to the undamaged or less damaged outer surface of the macula [7-12]; however, this approach is now criticized because prisms simply move the image, but patients' eyes move to negate the prismatic image shift and place their preferred retinal locus on the new eccentric focal point [13,14]. The concept of prism spectacles was not supported by a large randomized controlled trial, which found no benefit from prisms in quality of life, visual acuity, or other visual function tasks [15]. The preferred retinal locus for near reading lies mostly left or right of the macula [16-21], where, for distance viewing, the preferred retinal locus below is most common [9.15,18,22-25]. Some prism prescriptions, mostly with high powers, also result in various reported adverse events, for example, discomfort and dizziness [7-12]. The 6% magnification is only expected to improve visual acuity by one letter, 0.025 logMAR = $\log^{10}(1.06)$. The magnification aspect of E-Scoop would therefore have minimal impact and will not result in a clinically meaningful change on its own. Regarding the colored lenses, yellow/ orange lenses seemed to have only a minimal positive effect on contrast sensitivity [26-28]. Besides, yellow filters block some light, darken the image somewhat, and alter color perception. The antireflection coating helps to reduce reflections, which might improve contrast sensitivity as well [29].

Figure 1 E-Scoop is glazed to either a spectacle frame or fitted to a clip-on placed over a regular spectacle frame. E-Scoop is mostly monofocal, although bifocal and multifocal are available with addition up to +3.50, which can be used for near vision.

Previous studies on these separate components have therefore provided modest to no effects. A way forward might be to combine these distinct features in one device, such as E-Scoop. Currently, no data from controlled trials exist regarding the effectiveness of E-Scoop. It would be helpful to study the effects on visual acuity and contrast sensitivity but more importantly on quality of life. Because quality of life also depends on the value that a patient attaches to improved vision, improved vision alone is not a sufficient condition to conclude that the quality of life is improved. Such a line of reasoning makes quality of life a relevant outcome for reimbursement decisions. In that respect, the Dutch National Health Care Institute indicated that the evidence concerning the effectiveness of medical devices, such as spectacles, is not yet at the same level as pharmacology. For instance, in pharmacology, the National Health Care Institute expects that the effectiveness is expressed in terms that are relevant for the patient, such as health-related quality of life. However, such evidence is usually not available when the National Health Care Institute has to decide on the reimbursement of medical devices such as spectacles. Arguably, therefore, generic quality-of life questionnaires should be used when testing the effectiveness of medical devices. The aims of this study were to determine the effect of E-Scoop on quality of life in a real-life setting and to examine whether E-Scoop enhanced visual acuity and contrast sensitivity. A second aim was to measure the burden of disease in patients with AMD with low-vision complaints.

Methods

Study Design

This randomized controlled trial was conducted to test the impact of E-Scoop on quality of life in a real-life setting. Subjects were recruited by eight low-vision specialists between September 2014 and July 2016 in low-vision clinics in the Netherlands. Inclusion criteria were as follows: (1) previously diagnosed with

AMD and referral by an ophthalmologist, (2) age older than 18 years, (3) reported benefit from E-Scoop during the initial baseline visit, and (4) a signed informed consent form. Whether the AMD was atrophic or exudative was not an exclusion criterion. Subjects who could not understand Dutch or had physical impairments impeding participation were excluded. Subjects could leave the study at any time for any reason without consequences. Randomization took place after the baseline visit by a coordinator at the central office using a 1:1.15 computerized block randomization. We anticipated a higher dropout in the E-Scoop group because they would receive the quality-of-life questionnaires at home. Low-vision specialists were unaware of the group distribution at baseline. The study was registered in the Dutch Trial Register: NTR 6126. For the current study, all subjects completed the following quality-of-life questionnaires during the initial baseline visit (T0) at the clinic: the 25-item National Eye Institute Visual Function Questionnaire and the EuroQol 5-Dimension 5-Level. Visual acuity in logMAR and contrast sensitivity in log were measured monocularly and binocularly using an optical trial frame with optimum refraction and a trial frame with both optimum refraction and E-Scoop lens combined. All subjects received E-Scoop 6 weeks later, during a visit to the clinic at which measurements of visual acuity and contrast sensitivity with the optical trial frame with optimum refraction were repeated, both with and without wearing E-Scoop. Subjects in the control group completed the follow-up quality-of-life questionnaires using the 25-item National Eye Institute Visual Function Questionnaire in the clinic before receiving the E-Scoop lenses (T1, 6 weeks after baseline). The E-Scoop group completed the follow-up quality-of-life questionnaires using the 25-item National Eye Institute Visual Function Questionnaire at home after using E-Scoop for 3 weeks and then returned them by post (T1, 9 weeks after baseline). The control group filled in the questionnaire in the clinic, whereas the E-Scoop group completed it at home. This choice was made to keep the number of appointments the same for each group.

Fitting E-Scoop

Patients were eligible for the study only when they indicated that they experienced benefits during an initial test from wearing an optical trial frame with a yellow tint and prisms. Other features from E-Scoop were not tested, such as the 6% magnification. Whether patients report an advantage from wearing E-Scoop was determined via a fixed protocol: viewing using an optical trial frame with optimum refraction binocularly, followed by the addition of a yellow tint. When subjects reported that the yellow tint improved their vision in terms of visual acuity or subjectively, the prisms were added. The test compared a base-up versus a base-

down prismand the three available prism powers, starting with the lowest. Only when the patient reported benefits was the E-Scoop prescribed in consultation with the patient. Most E-Scoop prescriptions were monofocal, although bifocal and multifocal are available with addition up to +3.50, which can be used for near vision. Patients were then informed about participating in the study.

Outcome Measures

TWENTY-FIVE-ITEM NATIONAL EYE INSTITUTE VISUAL FUNCTION QUESTIONNAIRE

The primary outcome was quality of life as determined with the 25-item National Eye Institute Visual Function Questionnaire. The 25-item National Eye Institute Visual Function Questionnaire is a vision-specific quality-of-life questionnaire consisting of a 25-item base set of questions and a supplement of 14 additional items [30]. Originally, this questionnaire consisted of multiple subscales; however, most subscales were found to lack validity and reliability and have been reorganized [31]. Summing the item results into a "composite score" produces less bias, but the unidimensionality of the 25-item National Eye Institute Visual Function Questionnaire is questioned [31]. This composite score ranges from 0 to 100, where higher scores represent a higher quality of life [30]. The most statistically sound way is to construct Rasch scores, which are extensively described in the statistical analyses hereinafter [31]. For this study, both outcomes were evaluated.

VISUAL ACUITY AND CONTRAST SENSITIVITY

The secondary outcomes were visual acuity and contrast sensitivity measured with optimum refractive correction, both with and without E-Scoop. Visual acuity in logMAR was measured monocularly and binocularly with the Early Treatment Diabetic Retinopathy Study chart at 3musing the letter-by-letter scoringmethod [32]. Contrast sensitivity wasmeasured with the Pelli-Robson contrast sensitivity chart at a 1-m distance using the letter-by-letter scoring method, with an increase of 0.05 logCS per additional letter [33,34].

BURDEN OF DISEASE

Burden of disease was determined by comparing the vision-specific and general burden with available normative data. Using standardized questionnaires enables us to compare the quality of life of our subjects with the quality of life of the general population. These comparisons give insight into the burden for patients withAMDin this study. To determine the vision-specific burden of disease, we used the 25-item National Eye Institute Visual Function Questionnaire and matched our cases with

a Dutch sample from the general population (n = 910) collected in August and September 2017 [35]. The general burden of disease was measured using the generic quality-of-life questionnaire EuroQol 5-Dimension 5-Level, including the EuroQol Visual Analogue Scale. The EuroQol 5-Dimension 5-Level represents the societal perspective on quality of life on a scale from 0 to 1, where higher scores indicate lower burden. The EuroQol Visual Analogue Scale represents the patient perspective on quality of life on a scale ranging from 0 to 100, with higher scores indicating a better quality of life. These results were compared with two published reference values [36-38].

Statistical Analyses

DEMOGRAPHICS

Descriptive statistics were used to describe patient characteristics such as sex, age, visual acuity, contrast sensitivity, and qualityof- life baseline scores, representing the burden of disease.

E-SCOOP: QUALITY OF LIFE

For the primary outcome, the 25-item National Eye Institute Visual Function Questionnaire Rasch analysis was applied using a generalized partial credit model. This is a two-parameter Rasch model for ordered categories. The generalized partial credit model assumes equal differences between the answer categories over the items. This makes an ordering of the items on the latent trait possible, based on the item measure, and provides itemdifferentiation parameters. Rasch analysis also allows the respondent's performance to be expressed on this same latent trait, the person measure [39]. Rasch analysis requires unidimensionality of the items. First, we removed items with too many missing values (>10%), for example, on driving a car. Next, a principal component analysis was performed to check unidimensionality with the following criteria: (1) the eigenvalue of the component should be higher than the eigenvalue found with aMonte Carlo principal component analysis for parallel analysis; (2) the first eigenvalue should be at least five times higher than the second one; (3) the explained variance of the first component should be at least 50%; (4) items should load at least 0.50 on the first component; and (5) item loadings on the first component should be higher than on the second [40,41].

All items complying with these unidimensionality requirements were included in the generalized partial creditmodel. Infit and outfit measures are mean squares provided by Winsteps (Winsteps. com, Beaverton, OR), to detect poorly fitted items. Mean squares greater than 1.0 indicated an underfit to the model, and

mean squares less than 1.0 indicated an overfit, whereas values between 0.7 and 1.3 were considered acceptable [42]. Differential item functioning may occur when a test item does not have the same relationship to a latent variable across two or more groups [39]. That means that persons from different groups who have the same position on the latent trait will have a different outcome. In this study, differential item functioning was discerned for the two randomization groups. A differential item functioning *t* value of more than ±2 was considered significant. The person measures derived from the generalized partial credit model analysis were applied in the multilevel models.

Change in quality of life as measured with the 25-item National Eye Institute Visual Function Questionnaire was analyzed with multilevel regression models after an intention-to-treat analysis. For both outcomes, a model was constructed with three levels: optometrist as the upper level, the subjects as the intermediate level, and their baseline and follow-upmeasure as the lower level. The necessity of the optometrist or clinic as the upper level was determined with a deviance test using restricted maximum likelihood.[43] Treatment group, time, and their interaction were defined as fixed effects. In addition, because age and sex have significant relations to several items, these were entered as covariates. To compensate for potential baseline differences, we also included baseline visual acuity. On top of that, for possible nonlinear effects, we also included a quadratic effect to the model. For the two primary outcomes, we applied a Bonferroni correction, considering $p < .03$ as significant.

E-SCOOP: VISUAL ACUITY AND CONTRAST SENSITIVITY

The secondary outcomes best-corrected visual acuity and contrast sensitivity were tested for normality. In case of a nonnormal distribution, the differences within patients between with and without E-Scoop were analyzed using nonparametric Wilcoxon signed rank tests. These analyses were performed separately for the better-seeing eye, the worse-seeing eye, and both eyes together, for visual acuity and contrast sensitivity. In case both eyes had equal sight, the eyes were randomly assigned to better- or worse-seeing eye. Medians and interquartile ranges or mean differences with standard deviations were presented, as appropriate. Analyses were performed with SPSS version 24 (IBM SPSS Statistics for Windows; IBM Corp., Armonk, NY), and generalized partial credit model was performed with Winsteps version 4.1.0 [44]. For the exploratory secondary outcomes, we applied a Bonferroni correction considering $p < .05$ as significant. In addition, in light of minimal clinically meaningful difference, we hold onto a test-retest threshold for visual acuity of 0.20 logMAR and for contrast sensitivity 0.35 logCS [45].

QUALITY OF LIFE: AMD VERSUS THE GENERAL POPULATION

To determine the vision-specific ad general burden of disease, we matched our cases with the general samples by sex and age. We applied Excel (Microsoft Corporation, Redmond, WA) to sort the sex and age of both the E-Scoop group and the general sample. Older subjects were better represented in the E-Scoop group than in the general sample. To compensate for this, younger subjects from the E-Scoop group were matched with older subjects from the general sample. This procedure resulted in equal means but a larger standard deviation for the E-Scoop group. The vision-specific burden of disease measured by the 25-item National Eye Institute Visual Function Questionnaire was compared with the matched general sample using a paired *t* test. The general burden of disease, as measured with the EuroQol 5-Dimension 5-Level and the EuroQol Visual Analogue Scale, was compared with the general population reference using the Welch *t* test [36-38].

SAMPLE SIZE CALCULATION

In a pilot study, we solely focused on the responsiveness of quality-of-life questionnaires. Twenty-three E-Scoop users were included. For the 25-item National Eye Institute Visual Function Questionnaire, we addressed the subscale (distance activities) that best suited the aim of the spectacles, that is, "improving distance viewing". We observed a difference of 6.9, a standard deviation of 21.0, and a correlation of 0.80 between the first and the second measurement (data not published). For an α of 0.05 and a power of 0.90, 80 effective cases are needed in each group. Anticipating a general dropout of 10% and higher dropout in the E-Scoop group, a total number of 178 subjects are needed to be included in the study.

Results

Demographics

The eight low-vision specialists recruited 226 eligible subjects. A total of 190 subjects agreed to participate and were randomized into the control group (n = 86) or the E-Scoop group (n = 99; Table 1, Figure 2). There were no significant differences between the control and E-Scoop group for any baseline variables reported in Table 1. At baseline, three participants in the E-Scoop group did not administer questionnaires. Follow-up data were available for 78 participants in the control group and 79 in the E-Scoop group, in line with the predicted higher dropout in the E-Scoop group. All dropouts were a result of loss to follow-up, with the biggest group failing to resend the questionnaires being the E-Scoop group. Baseline characteristics are presented in Table 1.

Table 1 Demographics and baseline scores

	Control group (*n* = 86)		E-Scoop group (*n* = 99)		Total (*n* = 190)*	
Demographics						
Female, n (%)	50	(58)	55	(56)	109	(57)
Age (y), mean ± SD	79.6	± 9.6	79.2	± 10.8	79.5	± 10.2
EQ-5D-5L, mean ± SD	0.67	± 0.30	0.70	± 0.22	0.68	± 0.26
EQ-VAS, mean ± SD	65.6	± 19.7	69.1	± 15.0	67.5	± 17.2
NEI VFQ-25, mean ±SD						
Rasch score	0.12	± 1.35	0.22	± 1.18	0.17	± 1.26
composite score	52.6	± 20.1	51.2	± 18.1	51.8	± 18.9
Visual acuity in LogMAR, mean ±SD						
BSE	0.28	± 0.16	0.31	± 0.19	0.29	± 0.18
WSE	0.10	± 0.14	0.12	± 0.14	0.11	± 0.14
Both	0.29	± 0.18	0.31	± 0.20	0.30	± 0.19
Contrast sensitivity in logCS, mean ±SD						
BSE	0.98	± 0.43	0.95	± 0.43	0.97	± 0.43
WSE	0.51	± 0.55	0.50	± 0.54	0.50	± 0.54
Both	0.98	± 0.42	0.98	± 0.44	0.98	± 0.42

*The total includes 5 patients who were not randomized, but did have quality of life and visual scores. Both = vision with both eyes; BSE = betterseeing eye; EQ-5D-5L = EuroQol 5 dimension 5 level; EQ-VAS = EuroQol Visual Analogue Scale; NEI VFQ-25 = 25-item National Eye Institute Visual Function Questionnaire; SD = standard deviation; WSE = worse-seeing eye.

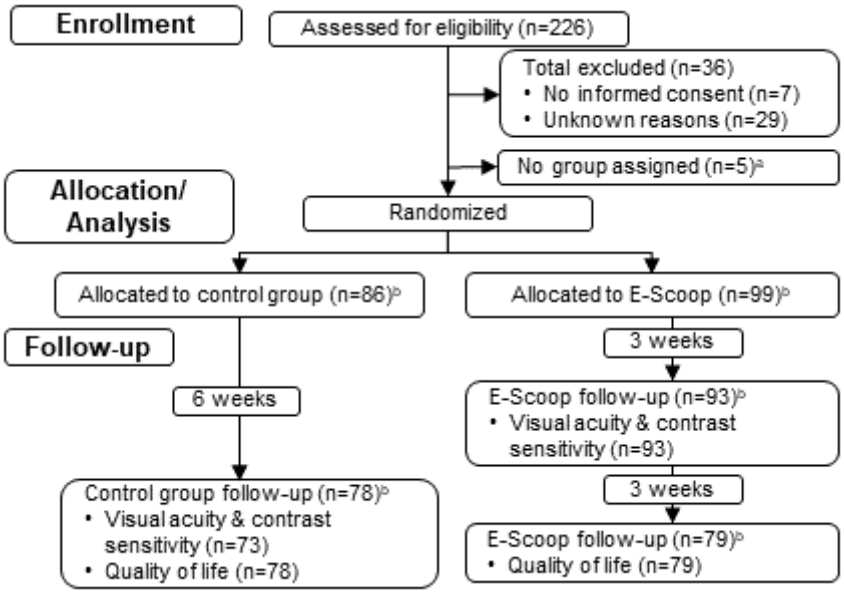


Figure 2 Flow diagram of enrolment, allocation and analysis.
a These 5 patients were not randomized, and therefore, could not be included in the quality of life analysis. However, they did report visual scores and were, in line with the intention to treat principle, included in the analyses of the visual measures.
b As the intention to treat principle was followed, all available cases were included in the analysis for the NEI VFQ-25 in both control and E-Scoop group, regardless of missing data on baseline or follow-up. All drop outs were a result of loss to follow up, with the biggest group failing to resend the questionnaires in the E-Scoop group.

Table 2 Multilevel models of changes in NEI VFQ-25 Quality of Life scores between the E-Scoop and control group

	Control group		E-Scoop group			
	Baseline (*n* = 85)	Follow-up (*n* = 76)	Baseline (*n* = 96)	Follow-up (*n* = 77)	Effect size Cohen's *d*	*p* value
Primary outcomes						
Rasch score	0.09	0.05	-0.08	-0.24	-0.11	0.43
Composite score	53.1	53.2	55.3	59.1	0.23	0.09

Estimates are at mean of covariates visual acuity, age, and sex. *Because of incomplete and/or missing questionnaires, 4 baseline measures are unavailable (1 in the control group and 3 in the E-Scoop group) and 28 follow-up measures (10 in the control group and 22 in the E-Scoop group). NEI VFQ-25 = 25-item National Eye Institute Visual Function Questionnaire

E-Scoop: Quality of Life

We removed items 14, 15, and 16 because these had too many missing values. The Monte Carlo principal component analysis for parallel analysis showed that a significant first component should have an eigenvalue higher than 1.59 and a second higher than 1.50. The first component of the principal component analysis had an eigenvalue of 9.93, and the second had 1.25; thus, the first two requirements for unidimensionality were met. However, the explained variance of the first component was only 45.1%, and items 1, 4, and 19 had loadings of, respectively, 0.49, 0.34, and 0.51 on the first component and −0.17, 0.82, and 0.67 on the second. Item 1 concerned general health, and items 4 and 19 concerned pain. Removing these three items resulted in a one-component solution with all factor loadings >0.50. We decided to ignore the two-item pain scale, as two items are insufficient for Rasch analysis. The iteminfit and outfit means were 1.00 and 0.99, and the person infit and outfit means were 1.01 and 1.00, respectively, which are acceptable measures. No differential item functioning t value was larger than 1.96 for any item. Greater Rasch scores were related to poorer sight.

The deviance test pointed out that the three-levelmodel showed a better fit than did the two-level model (= 14.69, $p < .001$). Thus, we included optometrist as an upper level in the analyses. No statistically significant differences were observed between the E-Scoop and control groups on the 25-item National Eye Institute Visual Function Questionnaire Rasch score ($d = -0.07$, $p = .53$) and the composite score (d=0.14, $p = .22$; Table 2, Figure 3; Appendix 1). No significant effects of age and sex were observed.

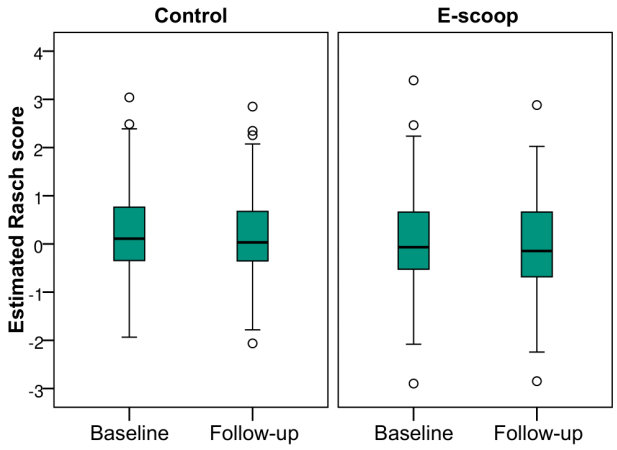


Figure 3 Box & whisker plot of Rasch person scores estimated by the linear mixed model of the NEI-VFQ-25 questionnaire.

E-Scoop: Visual Acuity and Contrast Sensitivity

With E-Scoop, the visual acuity chart indicated higher values for 86 of 168 patients in the better-seeing eye, for 90 of 171 patients in the worse-seeing eye, and for 38 of 168 patients in binocular visual acuity (Table 3). Within subjects, this change in visual acuity was statistically significant with averages of 0.05 logMAR for the better-seeing eye (2.5 letters, $p < .001$), 0.01 logMAR for the worse-seeing eye (1 letter, $p < .001$), and 0.05 logMAR for both eyes (2.5 letters, $p < .001$). However, the 0.05 logMAR difference was lower than the test-retest threshold of 0.20 logMAR, and so it was deemed not clinically meaningful [45].

The Wilcoxon signed rank test for contrast sensitivity with E-Scoop showed similar results. Higher scores were seen with E-Scoop in the better-seeing eye for 76 of 169 patients, in the worse-seeing eye for 38 of 169 patients, and in binocular contrast sensitivity for 83 of 170 patients. Change with E-Scoop compared with without E-Scoop was seen in the better-seeing eye (1.5 letter, $p < .001$) with a mean change of 0.09 logCS, the worse-seeing eye (1 letter, $p < .001$) with a mean of 0.05 logCS, and both eyes (2 letters, $p < .001$) with a mean of 0.10 logCS. In addition, the 0.10 logCS change was lower than the test-retest threshold of 0.35 logCS [45].

Quality of Life: AMD versus the General Population

For the vision-specific burden of disease, indicated by the 25-itemNational Eye Institute Visual Function Questionnaire composite score, we matched 180 cases by sex and age. Mean ages were 79.5 (10.1) years in the sample of the E-Scoop group and 79.5 (8.4) years in the sample of the general population. Themean 25-itemNational Eye Institute Visual Function Questionnaire composite score for our study population, 51.8 (18.9), was significantly lower ($t_{181} = 20.8$, $p < .001$) than the general population, 86.0 (12.3). The general burden of disease at baseline was estimated, with the mean (standard deviation) of the EuroQol 5-Dimension 5-Level being 0.68 (0.26) and that of the EuroQol Visual Analogue Scale being 67.5 (17.2). When these results were compared with a sample of the general population older than 75 years from the 2006 data, significantly higher scores were found, 0.83 (0.27) for the EuroQol 5-Dimension 5-Level ($p < .001$) and 72.9 (24.3) for the EuroQol Visual Analogue Scale ($p < .01$) [36,38]. Also, a more recent sample of the general population older than 70 years from the 2016 data reported a significantly higher score of 0.85 (0.15) for the EuroQol 5-Dimension 5-Level ($p < .001$).

Table 3 Visual Acuity and Contrast Sensitivity with and without E-Scoop (both E-Scoop group and control group) at baseline

	Without E-Scoop			With E-Scoop		Difference Mean (SD)	Ranks			Wilcoxon test	
	n*	Median	IQR	Median	IQR		Negative	Ties	Positive	*z*	*p* value
Visual acuity in LogMAR											
BSE	168	0.27	[0.16 – 0.40]	0.30	[0.16 – 0.48]	0.05 (0.09)	1	81	86	8.00	<0.001
WSE	168	0.05	[0.02 – 0.16]	0.05	[0.02 – 0.18]	0.01 (0.04)	2	128	38	5.48	<0.001
Both	171	0.28	[0.16 – 0.40]	0.30	[0.16 – 0.48]	0.05 (0.08)	2	79	90	8.06	<0.001
Contrast sensitivity in logCS											
BSE	169	1.05	[0.75 – 1.20]	1.05	[0.80 – 1.33]	0.08 (0.15)	3	90	76	7.15	<0.001
WSE	169	0.15	[0.00 – 0.90]	0.15	[0.00 – 1.05]	0.05 (0.12)	1	130	38	5.04	<0.001
Both	170	1.05	[0.75 – 1.20]	1.10	[0.90 – 1.35]	0.10 (0.16)	5	82	83	7.46	<0.001

Ranks: positive is in favor of E-Scoop. *Only complete caseswere included in this analysis. The five patients who were not randomized but did have visual
scores were also included in these analyses according to the intention-to-treat principle. Both = vision with both eyes; BSE = better-seeing eye;
IQR = interquartile range; SD = standard deviation; WSE = worse-seeing eye.

Discussion

This study aimed to determine the effect of E-Scoop on quality of life in a real-life setting in patients with AMD who initially report benefit from the E-Scoop trial frame with the yellow tint and prism. No effect was found on the vision-specific quality-of-life measure 25-item National Eye Institute Visual Function Questionnaire. The effect of E-Scoop on secondary outcomes visual acuity and contrast sensitivity were statistically significant, but these were deemed to not be clinically meaningful. The secondary aim to define the vision-specific and general burden of disease resulted in a lower quality of life than what one would expect in this elderly population. Apparently, these patients experience high burden. Therefore, the incentive to develop an effective vision rehabilitation intervention remains. On this latter note, more promising developments are made in electronic glasses [46].

In this research, we used the vision-related quality-of-life questionnaire 25-item National Eye Institute Visual Function Questionnaire, as this seemed the most suitable in AMD. Some suggest modifications for the 25-item National Eye Institute Visual Function Questionnaire, as it shows notable variation between different age groups.[47] However, this ought not to be an issue in the current study because of the homogenous age distribution. It is possible that other questionnaires, such as the Low Vision Quality of Life Questionnaire or the Veterans Affairs Low Vision Visual Functioning Questionnaire, are more sensitive [48-50]. However, most current quality-of-life questionnaires in ophthalmology tend to emphasize performance in near vision tasks.

As expected, the dropout was larger in the E-Scoop group. This could have influenced the results when, for example, the dropouts have particular characteristics such as a worse sight or lack of improvement. However, according to Little and Rubin [51], multilevel regression analysis is robust for selective dropout on the condition that the aspects related to the dropout are included in the model. In our analyses, we included many covariates. So we are confident that bias was not based on selective dropout by age, sex, or visual acuity. Although we observed a statistical effect on visual acuity and contrast sensitivity, we could not reveal a positive effect on quality of life. Quality of life is not only determined by a possible improvement in vision but also influenced by the value patients attribute to their vision and improvement in vision, as compared with the influence of other (sometimes nonvisual) factors on quality of life. This makes it more challenging to measure the effect of any vision enhancement on quality of life than it is to measure any enhancement of visual acuity and contrast sensitivity. Furthermore,

the relationship between visual acuity and quality of life is stronger in patients with AMD with better visual acuity [52].

Finally, we showed that subjects eligible for E-Scoop report a lower score on the 25-item National Eye Institute Visual Function Questionnaire and the EuroQol 5-Dimension 5-Level compared with an age-matched general population sample, suggesting that their low vision is a much higher burden, not only for specific vision-related quality of life but also for general quality of life. This suggests that the impact of the low vision of patients with AMD affects a wide range of daily life, including social functioning and role difficulties. Obviously, the unbiased estimate of this burden can only be done if we could correct for other differences in study population, for example, comorbidity or social economic status, which was not done in this study because we solely matched for age and sex. Nevertheless, the context is that E-Scoop tries to help patients in considerable need when there are few other alternatives to aim for distance vision.

It ultimately is the general aim of health care to improve the quality of life of patients. We have shown that it is possible to conduct studies thatmeasure quality of life in a real-life setting. In that respect, this study breaks ground for introducing "patient-related outcome measures" in a randomized setting in optometry in the Netherlands. The challenge in this study was to test E-Scoop in a real-life setting, where we asked subjects to reveal the possible effect of E-Scoop on quality of life in daily functioning. This is an improvement compared with the laboratory settings in which it is unclear whether the results have meaning in subjects' lives. However, research outside the laboratory does have a number of logistical challenges.

A first challenge of the design we encountered was the "short" test period of 3 weeks. From a research point of view, it could be preferable to have a longer test period with the result that subjects have a longer period to get used to E-Scoop. It could be argued that a greater effect on quality of life would have been found with a longer follow-up period, as there was a small effect found on visual acuity and contrast sensitivity. A limitation was that the usage of E-Scoop was not measured, which would have given more insight into the compliance. Another related question is that perhaps visual performance might also improve with a longer test period. Another challenge in the design was keeping the circumstances of both the E-Scoop and control groups the same during the study, apart from the intervention itself. We did not fully succeed in that respect: the timing of the administration of the second quality-oflife questionnaire differed between groups. The control group filled in the questionnaires in the clinic, whereas the E-Scoop group received them by post at home. This choice was being made to keep the number of appointments between groups the same. One might argue that the control group, who completes

the second quality-oflife questionnaire in the clinic, is more optimistic owing to the presence of a professional and the prospect of receiving the new spectacles. If this is the case, this would restrict the potential positive effect of E-Scoop. This could have been overcome by having the experimenter administer the instruments, both in the clinic and by telephone.

Third, we did not test the improvement of visual acuity and contrast sensitivity with a placebo in the form of sham glasses. Because participants had already seen the lenses during the initial eligibility assessment, it would have been difficult for sham lenses (e.g., in a different color) to effectively maintain the masking. A fourth challenge lies in selection and inclusion criteria. Although the subjects were recruited to wear E-Scoop only when they report benefits, there was no meaningful change on visual acuity and contrast sensitivity and no effect on quality of life. Apart from issues in generalizability, this subjective preference may therefore not be an appropriate inclusion criterion, and theremay be a better way to target suitable patients who may show greater benefit. One additional criterion could focus on patients with a particular level of visual acuity. In the multilevel model in Appendix 1, the term visual acuity squared hinted that we could gain more in the middle range of visual acuity.

CONCLUSIONS

In this first formal study testing the impact of E-Scoop in daily life, no effect on quality of life was found. Furthermore, E-Scoop showed no clinically meaningful positive effects on visual acuity and contrast sensitivity. Combining separate features within a single spectacle lens did not result in a measurable effect. Furthermore, research in this patient group is encouraged, as the burden of disease is shown to be much worse compared with an age-matched normal population sample, for both specific vision related quality of life and general quality of life.

References

1. Mangione CM, Gutierrez PR, Lowe G et al. (1999). Influence of Age-related Maculopathy on Visual Functioning and Health-related Quality of Life. American Journal of Ophthalmology, 128, 45–53.

2. Cruess A, Zlateva G, Xu X et al. (2007). Burden of Illness of Neovascular Age-related Macular Degeneration in Canada. Canadian Journal of Ophthalmology, 42, 836–843.

3. Nguyen CL, Oh LJ, Wong E et al. (2018). Anti-vascular Endothelial Growth Factor for Neovascular Age-relatedMacular Degeneration: A Meta-analysis of Randomized Controlled Trials. BMC Ophthalmology, 18, 130.

4. Schmier JK, Covert DW, Matthews GP et al. (2009). Impact of Visual Impairment on Service and Device Use by Individuals with Diabetic Retinopathy. Disability and Rehabilitation, 31, 659–65.

5. Schmier JK, Halpern MT, Covert DW et al. (2006). Impact of Visual Impairment on Service and Device Use by Individuals with Age-related Macular Degeneration (AMD). Disability and Rehabilitation, 28, 1331–1337.

6. Dickinson C (1998). Low Vision: Principles and Practice: Butterworth. Oxford, UK: Butterworth-Heinemann.

7. Bertrand T, Junginger H, Romanet JP et al. (1997). Description of the Prismation Method in the Rehabilitation of Low Visions of Macular Origin. Journal of French Ophtalmology, 20, 271-276.

8. Parodi MB, Toto L, Mastropasqua L et al. (204). Prismatic Correction in Patients Affected by Age-related Macular Degeneration. Clinical Rehabilitation, 18, 828–832.

9. Romayananda N, Wong SW, Elzeneiny IH et al. (1982). Prismatic ScanningMethod for Improving Visual Acuity in Patients with Low Vision. Ophthalmology, 89, 937–945.

10. Rosenberg R, Faye E, Fischer M et al. (1989). Role of Prism Relocation in Improving Visual Performance of Patients with Macular Dysfunction. Optometry and Vision Science, 66, 747–750.

11. Verezen CA, Volker-Dieben HJ, Hoyng CB (1996). Eccentric Viewing Spectacles in Everyday Life, for the Optimum Use of Residual Functional Retinal Areas, in Patients with Age-related Macular Degeneration. Optometry and Vision Science, 73, 413–417.

12. Vettard S, Dubois E, Quaranta M et al. (2004). Prismatic Treatment in Low-vision Rehabilitation of Patients with Age-related Macular Degeneration. Journal of French Ophtalmology, 27, 589–596.

13. Bailey IL (1983). Can Prisms Control Eccentric Viewing? Optometry Monthly, 74, 360–362.

14. Leat SJ, Campbell MC, Woo GC et al. (2001). Changes in Fixation in the Presence of Prism Monitored with a Confocal Scanning Laser Ophthalmoscope. Clinical and Experimental Optometry, 84, 132–138.

15. Smith HJ, Dickinson CM, Cacho I et al. (2005). A Randomized Controlled Trial to Determine the Effectiveness of Prism Spectacles for Patients with Age-related Macular Degeneration. Archives of Ophthalmology, 123, 1042–1050.

16. Cacho I, Dickinson CM, Reeves BC et al. (2007) Visual Acuity and Fixation Characteristics in Age-related Macular Degeneration. Optometry and Vision Science, 84, 487–495.

17. Crossland MD, Culham LE, Kabanarou SA et al. (2005). Preferred Retinal Locus Development in Patients with Macular Disease. Ophthalmology, 112, 1579–1585.

18. Fletcher DC, Schuchard RA (1997). Preferred Retinal Loci Relationship to Macular Scotomas in a Low-vision Population. Ophthalmology, 104, 632–638.

19. Guez JE, Le Gargasson JF, Rigaudiere F et al. (1993). Is there a Systematic Location for the Pseudo-fovea in Patients with Central Scotoma? Vision Research, 33, 1271–1279.

20. Rubin GS, FeelyM (2009). The Role of EyeMovements During Reading in Patients with Age-related Macular Degeneration (AMD). Neuroophthalmology, 33, 120–126.

21. Sunness JS, Applegate CA, Haselwood D et al. (1996). Fixation Patterns and Reading Rates in Eyes with Central Scotomas from Advanced Atrophic Age-related Macular Degeneration and Stargardt Disease. Ophthalmology, 103, 1458–1466.

22. Bowers AR, Woods RL, Peli E (2004). Preferred Retinal Locus and Reading Rate with Four Dynamic Text Presentation Formats. Optometry and Vision Science, 81, 205–213.

23. NilssonUL, Frennesson C, Nilsson SE (2003). Patients with AMD and a Large Absolute Central Scotoma Can Be Trained Successfully to Use Eccentric Viewing, as Demonstrated in a Scanning Laser Ophthalmoscope. Vision Research, 43, 1777–1787.

24. Peli E (1986). Control of Eye Movement with Peripheral Vision: Implications for Training of Eccentric Viewing. American Journal of Optometry and Physiological Optics, 63, 113–118.

25. Greenstein VC, Santos RA, Tsang SH et al. (2008). Preferred Retinal Locus in Macular Disease: Characteristics and Clinical Implications. Retina, 28, 1234–1240.

26. Bailie M, Wolffsohn JS, Stevenson M et al. (2013). Functional and Perceived Benefits of Wearing Coloured Filters by Patients with Age-related Macular Degeneration. Clinical and Experimental Optometry, 96, 450–454.

27. LangagergaardU, GanerHJ, Baggesen K (2003). Age-related Macular Degeneration: Filter Lenses Help in Certain Situations. Acta Ophthalmologica, 81, 455–458.

28. Wolffsohn JS, Dinardo C, Vingrys AJ (2002). Benefit of Coloured Lenses for Age-related Macular Degeneration. Ophthalmic and Physiological Optics, 22, 300–311.

29. Ross J, Bradley A (1997). Visual Performance and Patient Preference: A Comparison of Anti-reflection Coated and Uncoated Spectacle Lenses. Journal of the American Optometric Association, 68, 361–366.

30. Mangione CM, Lee PP, Gutierrez PR et al. (2001). Development of the 25-item National Eye Institute Visual Function Questionnaire. Archives of Ophthalmology, 119, 1050–1058.

31. Pesudovs K, Gothwal VK, Wright T et al. (2010). Remediating Serious Flaws in the National Eye Institute Visual Function Questionnaire. Journal of Cataract and Refractive Surgery, 36, 718–732.

32. Kitchin JE, Bailey I (1981). Task Complexity and Visual Acuity in Senile Macular Degeneration. Australian Journal of Optometry, 64, 235–242.

33. Pelli D, Robson JG, Wilkins AJ (1988). The Design of a New Letter Chart for Measuring

Contrast Sensitivity. Clinical Vision Science, 2, 187–199.

34. Elliott DB, Bullimore MA, Bailey IL (1991). Improving the Reliability of the Pelli-Robson Contrast Sensitivity Test. Clinical Vision Science, 6, 471–475.

35. Visser MS, Timman R, Nijmeijer KJ et al. (2020). A Very Short Version of the Visual Function Questionnaire (VFQ-3oo7) for Use as a Routinely Applied Patient-reported Outcome Measure. Acta Ophthalmology, 98, 618–626.

36. Lamers LM, McDonnell J, Stalmeier PF et al. (2006). The Dutch Tariff: Results and Arguments for an Effective Design for National EQ-5D Valuation Studies. Health Economics, 15, 1121–1132.

37. Versteegh MM, Vermeulen KM, Evers SMAA et al. (2016). Dutch Tariff for the Five-level Version of EQ-5D. Value in Health, 19, 343–352.

38. Szende S, Janssen B (2014). Cross-country Analysis of EQ-5D Data. In: Szende A, Janssen B, Cabases J, eds. Self-reported Population Health: An International Perspective Based on EQ-5D. Dordrecht, the Netherlands: Springer, 31–36.

39. Embretson SE, Reise SP (2013). Item Response Theory. Chineham, UK: Taylor & Francis.

40. Watkins MW (2006). Determining Parallel Analysis Criteria. Journal of Modern Applied Statistical Methods, 5, 344–346.

41. Reeve BB, Hays RD, Bjorner JB et al. (2007). Psychometric Evaluation and Calibration of Health-related Quality of Life Item Banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Medical Care, 45, S22–31.

42. Wright BD, Linacre JM, Gustafson J et al. (1994). Reasonable Mean-square Fit Values. Rasch Measurement Transactions, 8, 370.

43. Singer JD, Willett JB (2003). Applied Longitudinal Data Analysis—Modeling Change and Event Occurrence. Oxford, UK: Oxford University Press.

44. Linacre JM. Winsteps.com. WinstepsRasch Measurement [computer program]. Available at: https://www.winsteps.com/index.htm. Accessed September 29, 2020.

45. Kiser AK, Mladenovich D, Eshraghi F et al. (2005). Reliability and Consistency of Visual Acuity and Contrast Sensitivity Measures in Advanced Eye Disease. Optometry and Vision Science, 82, 946–954.

46. Wittich W, Lorenzini MC, Markowitz SN et al. (2018). The Effect of a Head-mounted Low Vision Device on Visual Function. Optometry and Vision Science, 95, 774–784.

47. LangelaanM, van Nispen RM, Knol DL et al. (2007). Visual Functioning Questionnaire: Reevaluation of Psychometric Properties for a Group of Working-age Adults. Optometry and Vision Science, 84, 775–784.

48. Wolffsohn JS, Cochrane AL (2000). Design of the Low Vision Quality-of-Life Questionnaire (LVQOL) and Measuring the Outcome of Low-vision Rehabilitation. American Journal of Ophthalmology, 130, 793–802.

49. Stelmack JA, Massof RW (2007). Using the VA LV VFQ-48 and LV VFQ-20 in Low Vision Rehabilitation. Optometry and Vision Science, 84, 705–709.

50. Stelmack JA, Szlyk JP, Stelmack TR et al. (2004). Psychometric Properties of the Veterans Affairs Low-Vision Visual Functioning Questionnaire. Investigative Ophthalmology and Visu-

al Science, 45, 3919–3928.

51. Little RJ, Rubin DB (2002). Statistical Analysis with Missing Data. 2nd ed. Hoboken,NJ: John Wiley & Sons, Inc.

52. Elshout M, van der Reis MI, de Jong-Hesse Y et al. (2016). Distinguishing between Better and Worse Visual Acuity by Studying the Correlation with Quality of Life in Neovascular Age-related Macular Degeneration. Ophthalmology, 123, 2408-2412.

CHAPTER 4

Sf-6d Utility Values for the Better- and Worse-Seeing Eye for Health States Based on the Snellen Equivalent in Patients With Age-Related Macular Degeneration

Martijn S. Visser · Sankha Amarakoon · Tom Missotten · Reinier Timman · Jan J. van Busschbach

Plos One 2017 12(2)

Abstract

Objective

Economic evaluations in wet age-related macular degeneration (ARMD) is hampered as often utility values for solely one eye are used, mostly the better-seeing eye (BSE). Moreover, frequently chosen methods rely on patient values and/or disease specific measures, while economic evaluations prefer generic quality of life (QoL) measures based on societal preferences. The generic QoL utility instrument EQ-5D has shown to be insensitive for differences in visual acuity. The aim of this study was therefore to provide societal utility values, using the generic SF-6D, for health states acknowledging both BSE and worse-seeing eye (WSE).

Methods

SF-6D utility values of 191 ARMD patients (≥65 years) with 153 follow-up measures at 1 year were used to fill health states defined by the combination of BSE and WSE using Snellen equivalents; no visual loss (≥20/40), mild-moderate (<20/40->20/200) and severe (≤20/200).

Results

QoL utilities were estimated for the SF-6D, ranging from 0.740 for ARMD patients without visual loss to 0.684 for patients with a combination of mild-moderate visual loss in their BSE and severe visual loss in their WSE.

Conclusion

Societal utility values are provided for ARMD patients using the generic QoL instrument SF-6D for visual acuity health states based on both BSE and WSE. The range of the values is smaller than previous elicited utilities with the disease-specific VisQoL. Besides, the utility values are placed on a more realistic position on the utility scale, and SF-6D utility values avoid the problem associated with the interpretation of disease-specific utility values.

Introduction

Age-related macular degeneration (ARMD) is the leading cause of severe vision loss and legal blindness among people over the age of 50 years in Western countries [1, 2]. Health economic evaluations in ARMD often assume that an outcome of interventions in the worse seeing eye (WSE) can be valued, in terms of quality of life (QOL), as if it was an outcome in the better seeing eye (BSE). Thus one assumes that an increase in visual acuity in the WSE has the same utility gain as an increase in visual acuity of the BSE. This assumption seems unlikely, as the WSE is assumed to be dominated by the BSE. On the other hand, economic evaluations sometimes do not even value the WSE at all, and assume that an increase in visual acuity in the WSE results in no utility gain, unless in improves beyond the BSE. This assumption seems also unlikely as this neglects the effect of loss of depth perception and loss of visual field. Therefore changes in the WSE, improvements or deteriorations, is expected to have a smaller influence in visual acuity than changes in the BSE, but should not be set to zero. Using utility of the BSE for effective interventions in the WSE might therefore lead to an overestimation of the benefit of interventions in ARMD, where not using them might lead to an underestimation of the benefit. The problem originates from using utility values based on a cross sectional sample, in which the health states were defined on the basis of the BSE [3, 4]. However, in clinical practice 60 to 72% of the eyes treated are WSEs [5, 6].

Several attempts have been made to overcome this flaw. One way is to use utilities valued by patients instead of utilities from the general public. However, in health economics values from the general public are preferred over patient values [7]. Finger et al. were the first to provide societal utilities for health states of visual acuity defined as the visual acuity in both the WSE and the BSE [8]. Their first attempt to do so failed, as a standard utility instrument, namely the EuroQol-5 dimensions (EQ-5D), seemed not sensitive enough. They were more successful using a disease-specific utility instrument, the Vision and Quality of Life Index (VisQoL) measuring disease-specific QoL. Because QoL instruments measure the subjective burden experienced by an individual, the utilities found were related to WSE and BSE eyes combined. After classifying the patients by their WSE and BSE visual acuity, Finger et al. could relate the utilities to visual acuity in both eyes. Finger et al. claimed that they had obtained utilities that could be used in health economic evaluation, notably Quality Adjusted Life Years (QALYs). However, it is still a discussion if the values of disease-specific instrument can be considered valid values for the estimation of QALYs. That is because values of the different stages of the `disease' are measured in isolation of co-morbidity, as shown in Table 2 of the publication of Finger et al., which presents utilities for various stages of visual acuity between .95

and .85, which is above the average utility of the general population that is usually between .90 and .80 [9]. It is not clear how these utilities in isolation are related to utilities that were elected in the presence of co-morbidity issues and it is unknown whether they are on the same scale [10-12]. Moreover, there are indications that disease-specific instrument tend to overestimate effects of the chosen morbidity, because respondents who provide the utilities focused too much on the chosen morbidity [13].

A solution to all this may be the use of generic utility instruments. Yet, the most used generic QoL instrument, the EQ-5D, was not significant related to difference in visual acuity in both eyes [14, 15]. Alternative generic instruments that can be used are the HUI, the AQoL-7D and the short form 6-D (SF-6D), which is based on the popular SF- 36 [16-18]. The SF-6D showed to be able to differentiate between patients with ARMD [14, 15, 19], and is found to be more sensitive in mild health states than the EQ-5D [20].

In health economics, models are built on health states defined in the terms of the primary clinical outcome. In ARMD that would be visual acuity measured using `Snellen charts' in a clinical settings or Early Treatment of Diabetic Retinopathy Study (ETDRS) charts in a research settings.[21] The aim of this study is to link SF-6D utilities to health states of visual acuity of the BSE and the WSE combined, defined in terms of the Snellen charts and the ETDRS charts.

Methods

Study design

The data in this analysis was used from a randomized controlled trial comparing three treatment regimens of bevacizumab (Avastin) for the treatment of ARMD [22]. In this trial a total of 191 ARMD patients, 65 years of age or older who had a visual acuity of 20/200 to 20/20 Snellen equivalents in the treatment eye, were randomized to treatment regimens of every 4, 6 or 8 weeks. The bevacizumab treatment consisted of a dose of 1.25 mg in a 0.05-ml solution. Treatment was continuous for one year, with visual acuity and QoL measures during their hospital visits at baseline and one year later. Written informed consent was obtained from all participants. The study was approved by the Erasmus Medical Research Ethics Committee (MEC- 2007-254) and was registered in the Dutch Trial Register (NTR 1174). No difference in visual acuity [22] nor in QoL [23] was found between the three treatment regimens.

Main outcome measures

Visual acuity in the treatment eye was measured with the Early Treatment Diabetic

Retinopathy Study (ETDRS) chart by the letter-by-letter scoring method [24]. The Snellen charts were used in the non-treatment eye. The outcome of both charts were divided in no visual loss (≥20/40), mild-moderate visual loss (<20/40->20/200) and severe visual loss (≤20/200). The two cut-offs chosen in line with a previous study [25]. In that study the cut-off of 20/40 was related to `legal driving vision' and the 20/200 cut-off to `legal blindness'. The latter cut-off correspond with the exclusion criterion for the treatment eye. The two cut-offs together with the differentiation between BSE and WSE define 6 health states, of which one is excluded (BSE worse than 20/200).

Two instruments to measure QoL were used; one generic utility instrument, the SF-6D, and one disease-specific questionnaire, the NEI VFQ-39. A Dutch value set to compute utilities for the SF-6D is currently not available, and therefore the UK value sets was used [16, 17]. NEI VFQ-39 consists of a 25-item base set of questions and a supplement of 14 additional items measuring vision-related QoL, which can be summarized into a `total component score'; range 0-100 [26]. The NEI VFQ-39 will provide information whether and how the health states differ, when measured with a sensitive disease-specific instrument. Several studies proposed a 10-point difference to be minimal clinically important [27-29]. Note that this NEI VFQ-39 total component score is not a `utility score' that can be used in QALY-analysis.

Data analysis

BETTER- AND WORSE-SEEING EYE

Better- and worse-seeing eyes were determined by the difference in letters. When baseline visual acuity in both eyes was equal or higher than 50 letters (20/100), a 5 letter difference was used as a minimal difference threshold to distinguish a BSE from a WSE [21, 30]. With a lower visual acuity, a minimal difference of 10 letters was established. When the minimal difference between the better- and worse-seeing eye did not met the criteria above, equal visual acuity was assumed.

STATISTICAL ANALYSES

Descriptive statistics analyses were performed. The observed utility means were displayed using the three health states for BSE and WSE; no visual loss (≥20/40), mild-moderate visual loss (<20/40->20/200) and severe visual loss (≤20/200). To account for the dependency of the two measures per patient and to straighten out possible inconsistencies due to illogical orderings, the utilities of the SF-6D and total component score of the NEI VFQ-39 were also analyzed with multilevel regression analyses. Patients formed the upper level of that analysis, the repeated measures the lower level. Included covariates were: dummies reflecting mild-moderate

vision in the BSE, mild-moderate vision in the WSE, severe vision problems in the WSE, gender, and age. The difference between the baseline and follow-up was not applied as a covariate as that is not of interest for the estimation of the utilities.

Results

Descriptive statistics

Characteristics of the included patients are shown in Table 1. Of the 191 patients enrolled at baseline, 3 patients were excluded from analyses because visual acuity was evaluated in only one eye. In total 39% *(n* = 73) of the patients were treated in the BSE against 48% *(n* = 91) patients in the WSE. For the remaining 13% *(n* = 24) equal visual acuity was assumed. Patients treated in the BSE had lower values compared to patients treated in the WSE on both SF-6D (.68 vs .74) and NEI VFQ (56 vs 75) at baseline. Patients treated in the BSE were slightly older than those treated in the WSE, 81 and 77 years respectively.

After one year of treatment one patient was excluded from analyses as visual acuity was evaluated in only one eye and a total of 34 patients dropped out of the study: eighteen patients dropped out as a result of non-compliance, seven because of serious adverse events, three died and six dropped out for other reasons.

Health states values

The observed number of participants, means and standard deviations of the SF-6D and NEIVFQ are presented in Table 2. The SF-6D utility of the state was [BSE Mild-moderate; WSE Mild-moderate; 0.745] seems illogical ordered, and suggests an interaction between BSE and WSE. Patients with no visual loss in the BSE reported a lower mean (0.727) than patients with mild-moderate visual loss (0.745). The multilevel regression model revealed no such interaction effect (p = 0.953) between the BSE and WSE, so this effect was deleted from the models.

The remaining parameters are presented in Table 3. The SF-6D utility of the health state with no visual loss in the BSE and mild-moderate visual loss in the WSE was estimated at 0.740, and the health state mild-moderate visual loss in the BSE and severe loss in the WSE was estimated at 0.684. The other health states had an intermediate position. Men had overall 0.04 point higher scores. The estimated values of the NEI VFQ-39 were 78.6 and 51.0 respectively. Older patients had overall 0.3 lower scores for each year they were older. Notable is the difference in parameter weight in the mild-moderate loss state between BSE and WSE. Where the NEIVFQ- 39 shows the BSE has a larger weight than the WSE (-16.2 vs. -4.4), the SF-6D shows the opposite direction with a smaller BSE weight than WSE weight (.008 vs. .033). As the parameters for the model are logical decrements, differences

in modelled SF-6D utilities and scores of the NEI VFQ-39 are in the expected direction: better health states are associated with better scores ranging from 0.740 to 0.684 (Table 4).

Table 1 Baseline characteristics relevant to patient-reported outcomes and visual acuity distribution according to better- or worse-seeing eye at baseline

	Study eye		
	BSE (*n* = 73)	**WSE (*n* = 91)**	**BSE = WSE (*n* = 24)**
Mean age, years ±SD***	81.1 ±5.4	77.4 ±6.7	75.9 ±6.7
Male gender, *n* (%)	24 (33)	31 (34)	7(29)
VA BSE, Snellen equivalent, *n* (%)			
>20/40	33 (45)	80 (88)	16[1](67)
20/40-20/80	29 (40)	11 (12)	7 (29)
20/80-20/200	11 (15)	0	1 (4)
VA WSE, Snellen equivalent, *n* (%)			
>20/40	9 (12)	28 (31)	15 (63)
20/40-20/80	4 (5)	38 (42)	8[1](33)
20/80-20/200	11 (15)	25 (27)	1 (4)
20/200-20/400	14 (19)	0	0
<20/400	35 (48)	0	0
BSE VA at 2 meter, mean ±SD***	.34 ±.23	.12 ±.16	.23 ±.25
WSE VA at 2 meter, mean ±SD***	1.48 ±.89	.48 ±.25	.27 ±.25
SF 6D, mean ±SD**	.683 ±.095	.740 ±.128	.746 ±.095
NEI VFQ-25, mean ±SD***	56.37 ±17.22	75.14 ±17.52	75.22 ±15.02

* p <0.05 ** p < 0.01 *** p <0.001

[1] In the same eye group, one person fell in different categories.

Abbreviations: BSE=better seeing eye; WSE = worse seeing eye; VA= visual acuity; SD = standard deviation

Table 2 Observed means of SF-6D and NEI-VFQ and numbers per health state

		WSE			
	BSE	No visual loss	Mild-moderate	Severe	Overall
		Mean ±SD (*n*)	Mean ±SD (*n*)	Mean ±SD (*n*)	Mean ±SD (*n*)
SF-6D	No visual loss	.755 ±0.114 (92)	.727 ±0.138 (88)	.717 ±0.125 (44)	.736 ±0.127 (224)
	Mild-moderate		.745 ±0.110 (38)	.668 ±0.114 (47)	.702 ±0.118 (85)
	Overall	.755 ±0.114 (92)	.732 ±0.130 (126)	.691 ±0.121 (91)	.727 ±0.125 (309)
NEI-VFQ	No visual loss	82.4 ±11.9 (98)	74.8 ±17.1 (100)	67.1 ±17.6 (50)	76.2 ±16.3 (248)
	Mild-moderate		58.9 ±16.1 (41)	49.3 ±14.1 (52)	53.5 ±15.7 (93)
	Overall	82.4 ±11.9 (98)	70.2 ±18.3 (141)	58.0 ±18.2 (102)	70.1 ±19.1 (341)

Abbreviations: BSE=better seeing eye; WSE = worse seeing eye; sd = standard deviation no visual loss (≥20/40), mild-moderate visual loss (<20/40 - >20/200) and severe visual loss (≤20/200)

Table 3 Multilevel regression parameters

	Estimate [95% CI]	Significance
SF6D		
Intercept	.740 [.711, .769]	<.001
BSE – mild-moderate loss	-.008 [-.040, .025]	.645
WSE – mild moderate loss	-.033 [-.063, -.003]	.034
WSE – severe loss	-.049 [-.088, -.010]	.015
Male gender	.039 [.004, .075]	.031
Age	-.002 [-.005, .000]	.108
NEI-VFQ		
Intercept	78.618 [74.973, 82.263]	<.001
BSE – mild-moderate loss	-16.163 [-20.291, -12.035]	<.001
WSE – mild moderate loss	-4.433 [-8.559, -.306]	.035
WSE – severe loss	-11.474 [-16.540, -6.407]	<.001
Male gender	2.856 [-1.293, 7.004]	.176
Age	-.342 [-.654, -.029]	.032

Table 4 Quality of life utilities [95% CI] per health state

		WSE			
	BSE	No visual loss	Mild-moderate	Severe	Mean
SF-6D *n* = 309	No visual loss	.740 [.711, .769]	.707 [.680, .734]	.691 [.658, .725]	.713 [.690, .735]
	Mild-moderate		.699 [.666, .732]	.684 [.649, .718]	.691 [.662, .721]
	Mean	.740 [.711, .769]	.703 [.678, .728]	.687 [.658, .717]	.704 [.685, .725]
NEI-VFQ *n* = 341	No visual loss	78.6 [75.0, 82.3]	74.2 [70.9, 77.5]	67.1 [63.0, 71.3]	73.3 [70.6, 76.0]
	Mild-moderate		58.0 [53.9, 62.2]	51.0 [46.8, 55.2]	54.5 [50.9, 58.1]
	Mean	78.6 [75.0, 82.3]	66.1 [63.0, 69.2]	59.1 [55.4, 62.7]	65.8 [63.4, 68.2]

Estimates at mean age and for women.
Abbreviations: BSE=better seeing eye; WSE = worse seeing eye
no visual loss (≥20/40), mild-moderate visual loss (<20/40 - >20/200) and severe visual loss (≤20/200)

Discussion

Principal findings

Societal utility values are provided for ARMD patients using the generic QoL instrument SF-6D for visual acuity health states based on both BSE and WSE.

Imbedding in the literature

One of the implications of this research is that it can be confirmed that visual acuity changes in the WSE should be taken into account in health economic evaluations. Simply stating that changes in the WSE have no effect on QoL is not supported by the results of this study: difference in visual acuity in the WSE does have a relation with utility. This implies that setting utility changes for the WSE to zero in health economic models will result in an underestimation of the cost effectiveness of effective treatments. Additionally, assuming that effects in the WSE are as valuable as in the BSE is an unjust simplification as the results of the multilevel regression of the SF-6D shows an even higher value for the WSE. Our finding stresses the need for realistic utilities for both changes in the BSE as well as for changes in the WSE.
It has been suggested that, in order to be sensitive in ARMD, a utility measure should have a dimension which is directly linked to vision [18]. Our findings state

that the SF-6D is sensitive in ARMD, although it has no such direct link with vision. On one hand, it can be assumed that the SF-6D underestimates the effects. On the other hand, the use of the generic SF-6D avoids the difficulty associated with disease-specific measures, like the focused effect, scale problems, the exclusion of side effects and comorbidities [13]. All those complications tend to overestimate the effects. Indeed, as theory predicts, our modelled range of the effects [0.740, 0.684] is smaller and at a lower place at the scale than Finger et al. measured with the disease specific instrument VisQoL [0.95, 0.84] [8]. Moreover, the place on the scale of our SF-6D is in line with the EQ-5D range found by Finger et al. [0.70, 0.67]. The range of the EQ-5D values of Finger et al. is a smaller, which might reflect the better sensitivity of the SF-6D in the higher region of the utility scale, as compared to the EQ-5D [20]. Indeed Finger et al. did not find statistical significant results when modeling the EQ-5D utilities to visual acuity, while we did find a relation between visual acuity and the utilities of the SF-6D. Therefore, we argue that our results represent a conservative estimate of differences between health states of visual acuity, without the complication that utilities estimated from disease-specific instruments bring.

It is too early to conclude that the disease-specific utilities of the VisQoL provide more sensitive outcomes than the SF-6D utilities reported in the present paper, as the variation in the utilities of both instruments are in the extreme health states.

Limitations

A limitation in our study is that the raw scores suggest that some health states have imprecise values, as the ordering is illogical. In the observed means (Table 2) we found an inconsistency within the group of patients with a mild-moderate visual loss in the WSE. Patients with no visual loss in the BSE reported a lower mean (0.727) than patients with mild-moderate visual loss (0.745). By using the multilevel model, these flaws are modelled out, but the model also reduces the range. This contributes to the idea that our results are an underestimation the real effects. More values per health states from further research could provide more precise values. When the study was initiated, the main focus was on the study eye, measured with an ETDRS LogMar chart. As the non-treatment eye was measured according to daily clinical practice, the Snellen chart was used. This difference in approach might complicate comparability of the visual acuity between both eyes. We think we were able to compensate for this difference, as we accounted for the known measurement errors and sensitivity of the used visual acuity charts. When no distinction could be made between the BSE and WSE, eyes were framed as equal visual acuity.

A limitation in the generalizability of our results is that we did not include patients with worse than 20/200 visual acuity in their BSE. So values from this investigation need consideration if and when many patients belong to that population.
It might be suggested that the reduction of the number of health states to five health states may be too rigorous. However, a larger number of health states will make the estimation of stable transition- and cost-parameters more complex. In economic modeling, results of studies with more health states than the five states used here are likely to be rearranged to fewer health states anyway. Of course in larger patient cohorts it would relatively easy to increase the number of separate health states.

Conclusion

Societal utility values are provided for ARMD patients using the generic QoL instrument SF-6D for visual acuity health states based on both BSE and WSE. The range of the values is smaller than that of the utilities elicited with the disease-specific instrument VisQoL, but the utilities are placed on a more realistic position on the utility scale. Furthermore, the utilities of the SF-6D avoid the problem associated with the interpretation of disease-specific utilities.

References

1. Bressler NM (2004). Age-related macular degeneration is the leading cause of blindness. JAMA, 291, 1900-1901.

2. Klaver CC, Wolfs RC, Vingerling JR et al. (1998). Age-specific prevalence and causes of blindness and visual impairment in an older population: the Rotterdam Study. Archives of Ophthalmology, 116, 653-661.

3. Brown GC (1999). Vision and Quality-of-Life. Transactions of the American Ophthalmological Society, 97, 473-511.

4. Larouche K, Rochon S (2005). Evaluation of photodynamic therapy for the treatment of exudative age-related macular degeneration (ARMD) with subfoveal neovascularization. Montreal (QC): Agence d'évalution des technologies et des modes d'intervention en santé. Available from: http://www.inesss.qc.ca/fileadmin/doc/AETMIS/Rapports/Ophtalmologie/2004_05_en.pdf.

5. Bressler NM, Chang TS, Fine JT et al. (2009). Improved vision-related function after ranibizumab vs photodynamic therapy: a randomized clinical trial. Archives of Ophthalmology, 127, 13-21.

6. Chang TS, Bressler NM, Fine JT et al. (2007). Improved vision-related function after ranibizumab treatment of neovascular age-related macular degeneration: results of a randomized clinical trial. Archives of Ophthalmology, 125, 1460-1469.

7. Drummond MF, Sculpher MJ, Claxton K et al. (2015). Methods for the economical evaluation of health care programmes. Fourth Edition ed. Oxford, England: Oxford University Press.

8. Finger RP, Fenwick E, Hirneiss CW et al. (2013). Visual impairment as a function of visual acuity in both eyes and its impact on patient reported preferences. PLoS One, 8, e81042.

9. Luo N, Johnson JA, Shaw JW et al. (2005). Self-reported health status of the general adult U.S. population as assessed by the EQ-5D and Health Utilities Index. Medical Care, 43, 1078-1086.

10. Brazier J, Rowen D, Yang Y, Tsuchiya A (2102). Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. European Journal of Health Economics, 13, 575-587.

11. Fryback DG, Lawrence WF Jr (1997). Dollars may not buy as many QALYs as we think: a problem with defining quality-of-life adjustments. Medical Decision Making, 17, 276-284.

12. Stolk EA, Busschbach JJ (2003). Validity and feasibility of the use of condition-specific outcome measures in economic evaluation. Quality of Life Research, 12, 363-371.

13. Brazier J, Tsuchiya A (2010). Preference-based condition-specific measures of health: what happens to cross programme comparability? Health Economics, 19, 125-129.

14. Longworth L, Yang Y, Young T et al (2014). Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. Health Technology Assessment, 18, 1-224.

15. Bozzani FM, Alavi Y, Jofre-Bonet M, Kuper H (2012). A comparison of the sensitivity of EQ-5D, SF-6D and TTO utility values to changes in vision and perceived visual function in

patients with primary open-angle glaucoma. BMC Ophthalmology, 12, 43.

16. Brazier J, Roberts J, Deverill M (2002). The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics, 21, 271-292.

17. Ware JE, Snow KK, Kosinski M, Gandek B (1993). SF-36 health survey: manual and interpretation guide. Boston, MA: New England Medical Center, The Health Institute.

18. Richardson J, Iezzi A, Peacock S et al. (2012). Utility weights for the vision-related Assessment of Quality of Life (AQoL)-7D instrument. Ophthalmic Epidemiology, 19, 172-182.

19. Tosh J, Brazier J, Evans P, Longworth L (2012). A review of generic preference-based measures of healthrelated quality of life in visual disorders. Value in Health, 15, 118-127.

20. Brazier J, Roberts J, Tsuchiya A, Busschbach J (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. Health Economics, 13, 873-884.

21. Falkenstein IA, Cochran DE, Azen SP et al. (2008). Comparison of visual acuity in macular degeneration patients measured with snellen and early treatment diabetic retinopathy study charts. Ophthalmology, 115, 319-323.

22. Lushchyk T, Amarakoon S, Martinez-Ciriano JP et al. (2013). Bevacizumab in age-related macular degeneration: a randomized controlled trial on the effect of injections every 4 weeks, 6 weeks and 8 weeks. Acta Ophthalmologica, 91, e456-461.

23. Visser MS, Amarakoon S, Missotten T et al. (2020). Six and eight weeks injection frequencies of bevacizumab are non-inferior to the current four weeks injection frequency for quality of life in neovascular age-related macular degeneration: a randomized controlled trial. Quality of Life Research, 29, 3305–3313.

24. The Early Treatment Diabetic Retinopathy Study Research Group (1991). Early Treatment Diabetic Retinopathy Study design and baseline patient characteristics. ETDRS report number 7. Ophthalmology, 98, 741-756.

25. Sahel JA, Bandello F, Augustin A et al. (2007). Health-related quality of life and utility in patients with age-related macular degeneration. Archives of Ophthalmology, 125, 945-951.

26. Mangione CM, Lee PP, Gutierrez PR et al. (2001). Development of the 25-item National Eye Institute Visual Function Questionnaire. Archives of Ophthalmology, 119, 1050-1058.

27. Lindblad AS, Clemons TE (2005). Responsiveness of the National Eye Institute Visual Function Questionnaire to progression to advanced age-related macular degeneration, vision loss, and lens opacity: AREDS Report no. 14. Archives of Ophthalmology,123, 1207-1214.

28. Miskala PH, Bressler NM, Meinert CL (2004). Is adjustment of National Eye Institute Visual Function Questionnaire scores for general health necessary in randomized trials? American Journal of Ophthalmology, 137, 961-963.

29. Miskala PH, Hawkins BS, Mangione CM et al. (2003). Responsiveness of the National Eye Institute Visual Function Questionnaire to changes in visual acuity: findings in patients with subfoveal choroidal neovascularization - SST Report No. 1. Archives of Ophthalmology, 121, 531-539.

30. Bressler NM, Chang TS, Suner IJ et al. (2010). Vision-related function after ranibizumab treatment by better- or worse-seeing eye: clinical trial results from MARINA and ANCHOR. Ophthalmology, 117, 747-756 e4.

CHAPTER 5

Validation, Test–Retest Reliability and Norm Scores for the Dutch Catquest-9SF

Martijn S. Visser · Myrthe Dieleman · Stijn Klijn · Reinier Timman · Mats Lundström · Jan J. V. Busschbach · Nicolaas J. Reus

Acta Ophthalmologica 2017: 95: 312–319

Abstract

Purpose

The Catquest-9SF questionnaire is a unidimensional, reliable, valid and short patient-reported outcome measure for quantifying benefits in visual functioning from cataract surgery. Our aim was to develop a formal Dutch translation, calculate norm scores, assess its validity and test–retest reliability and provide an easy way for use in clinical practice.

Methods

Translation of the questionnaire was performed according to guidelines of the International Society for Pharmacoeconomics and Outcomes Research. Catquest-9SF was obtained in 657 patients pre- and postcataract surgery. We applied Rasch and classical analyses to determine the questionnaire performance with characteristics such as unidimensionality, reliability, separation and differential item functioning. Test–retest reliability was assessed in another group of 145 patients. A cut-off value to discriminate between people with and without cataract, norm scores and a reliable change index (RCI) were calculated using data from a sample of 916 'healthy' persons from the normal population.

Results

The Dutch Catquest-9SF was unidimensional, and both person and item reliability were high; 0.87 and 0.99, respectively. Cronbach's alpha was 0.94, test– retest reliability was 0.85 and the intraclass correlation coefficient was 0.93. Catquest- 9SF showed to be responsive to the effect of cataract surgery (effect size = 1.27; $p < 0.001$). The cut-off value was-1.90, and RCI was 2.27.A quick-access tablewith norm scores and percentiles was established to facilitate clinical interpretation.

Conclusion

This investigation provides validity and reliability of the Dutch Catquest-9SF as well as norm scores and a new tool to facilitate the clinical interpretation of patient scores. This makes Catquest-9SF suitable for routine use in clinical practice.

Introduction

Cataract surgery is the most frequently performed ophthalmic surgical procedure in the world. Traditionally, visual acuity has been regarded to be an important variable in determining whether cataract surgery is required and in quantifying the surgical outcome. However, the quality of vision as perceived by the patient may be influenced by parameters other than visual acuity. Several questionnaires have been developed to quantify the subjective quality of vision with cataract. The Catquest nine-item short-form (Catquest- 9SF) is one of such diseasespecific questionnaires. It measures problems due to reduced visual functioning as perceived by the patient in his or her daily-life activities [1,2]. Seven items (i.e. questions) relate to performance in specific dailylife activities, and two items evaluate the patient's general perception of difficulties and satisfaction with vision [2]. The original Catquest questionnaire included 17 items on four areas of visual functioning (i.e. daily-life activity level, perceived difficulties in performing specific activities of daily living, cataract symptoms and items about difficulties and satisfaction with vision in general) [3]. To increase its validity and reliability, Lundström and Pesudovs [1] applied Rasch analysis to this questionnaire. This resulted in the Catquest-9SF with a reduced number of nine items. Previous studies in Sweden, Australia, Germany/ Austria, Italy and China have shown that the Catquest- 9SF is a valid measure for investigating perceived visual functioning with cataract surgery [4-8].

The Catquest-9SF has been adopted as a patient-reported outcome measure (PROM) by the International Consortium for Health Outcomes Measurement (ICHOM) for their Standard Set for Cataracts measurement of risk factors and outcomes [9]. The Catquest-9SF has also been advocated to be used in daily clinical practice in determining the indication for surgery as it relates to clinical outcome measures and can be used to improve quality of care [10]. An expert review that rated 48 PROMs, the Catquest-9SF has been recommended as a higher-quality PROM for use with cataract surgery [11]. Currently, however, it is not being used routinely. Several factors might contribute to this. First, up to now, a patient's score on the Catquest-9SF questionnaire could only be calculated using a spreadsheet in which the given answers for each of the nine items were entered and a value (i.e. logit) was calculated based on specific weighting factors. Because this necessitates a computer to be available and time to enter the data in the spreadsheet, it may impede the use of the questionnaire in daily clinical practice. Second, Catquest-9SF scores of people in the normal population are not available. However, these so-called norm scores are important when the questionnaire is used as a tool in determining the indication for surgery [12,13]. Furthermore, the minimum amount of change that the Catquest-9SF can detect reliably has not been published. This is

important when determining the number of patients that truly benefit from surgery. Finally, test–retest reliability of the Catquest-9SF has not yet been evaluated.
In order to facilitate its use in clinical practice, this study aimed to provide norm scores for the Catquest- 9SF for patients as well as the normal population, and an easy way to interpret its outcome. In addition, because a localized Catquest-9SF questionnaire was not available in the Netherlands, we set out to provide a formal Dutch translation.

Subjects and Methods

Translation

The translation was carried out in two steps. No validated English version of the Catquest-9SF existed at the time this study was initiated. Because English is often used as a source language for translations, it was decided to first create an (British) English translation and linguistic validation of the Swedish version. Next, the English version was translated into Dutch and linguistically validated. To achieve the greatest extent of both semantic and conceptual equivalence between the source and target language versions, we complied with the International Society for Pharmacoeconomics and Outcomes Research's (ISPOR) 'principles of good practice for the translation and cultural adaption process for patient-reported outcomes'. To this end, we involved the instrument's developer (ML) [14].
Furthermore, according to the suggested principles, study teams included one key in-country consultant and two independent bilingual translators for Sweden, the United Kingdom and the Netherlands. During the translation, a concept elaboration document, approved by the instrument's developer, was used defining each item of the Catquest-9SF and the conceptual meaning behind each item. The translation involved the ten steps as suggested by Wild et al. [14]. These are preparation, forward translation, reconciliation, backward translation, backward translation review, harmonization, cognitive debriefing, review of the cognitive debriefing results and finalization, proofreading and the final report. All steps were completed for the translation from Swedish to English as well as the translation from English to Dutch.

Data collection

Data of cataract patients were prospectively collected in the Rotterdam Eye Hospital, Rotterdam, the Netherlands. Data of 657 consecutive patients, who completed the Catquest-9SF before surgery as well as 3 months after surgery, were used for validation using Rasch analysis. A second group of 145 consecutive patients was included for assessment of the test–retest reliability. These patients completed

the questionnaire twice before surgery, with an interval of 1 week. Additionally, to establish norm scores, the questionnaire was set out through an online survey among a panel representative for the Dutch population aged 50 years or older, resulting in a third group of 1048 people. Of the respondents, 132 were excluded because of previously having undergone cataract surgery (30 had surgery in one eye and 102 in both eyes). This resulted in a 'normal' population, of 916 people.

Rasch analysis

The characteristics of the Catquest-9SF data were assessed by Rasch analysis using WINSTEPS software (version 3.80.1) with an Andrich rating scale model [15,16]. Other statistical analyses (i.e. Pearson's correlation coefficient, paired two-tailed ttest and analysis of variance) were performed using SPSS Version 21.0 (IBM Corp., Armonk, NY, USA). Eight of the nine items in the Catquest-9SF have the following response categories: very great difficulty; great difficulty; some difficulty; no difficulty; can't say. For the remaining item (i.e. satisfaction with vision), the response categories are as follows: very dissatisfied; rather dissatisfied; fairly satisfied; very satisfied; can't say. These two different response categories are treated separately. In addition, according to the instructions of the original questionnaire guidelines, the response category can't say is treated as missing data in the analysis [3]. As the Catquest-9SF is used both pre- and postoperatively, it should be valid and have comparable performance for measurements in both situations. Therefore, the Rasch analysis was performed on both pre- and postoperative data stacked as a single data set, and an additional analysis was performed to compare the pre- and postoperative data. Incomplete questionnaires (defined as less than seven completed items) were excluded, resulting in a total of 1089 remaining observations pre- or postoperative (respectively 639 and 450 observations). Patient characteristics are presented in Table 1.

SEPARATION AND RELIABILITY

Key indicators for questionnaire performance in Rasch analysis include a measure of separation and reliability for items as well as persons. Separation is defined as the ratio of the true spread of the measurement to the error of the measurement. A separation of 2.00 is considered a minimum level of performance. For each item, all possible response categories are weighted in Rasch analysis. Item and person reliability then provide how much confidence can be placed in the consistency of the estimates of these weights. Reliability may range from 0 to 1; coefficients >0.8 represent 'good', and >0.9 'excellent' reliability [17]. The combination of good person separation and good person reliability suggests that the instrument is able to accurately distinguish between low and high performers. In case of the Catquest-

Table 1 Patient characteristics

Characteristics	Pre-post	Test-retest	Normal population
Patients, *n*	657	145	916
Age in years, mean ±SD	69.9 ±11.1	71.2 ±10.0	63.3 ±8.8
Female *n* (%)	366 (56%)	80 (56%)	473 (52%)
One-eye surgery *n* (%)	225[1] (49%)	-	-
Both-eye surgery *n* (%)	237[1] (51%)	-	-

[1] Data of 195 patients was unavailable mostly due to missing post-surgery measures.

9SF, this means that the measure is able to accurately distinguish between people with and without cataract- related visual disturbances. In addition, it provides evidence that the ordering of the response categories of the items is valid, that is that people who responded that they have very great difficulty with a certain task indeed have more difficulty with that task than people who responded that they have no difficulty. Furthermore, Cronbach's alpha and the item rest correlations are provided. Finally, test– retest reliability was calculated with Pearson's correlation coefficient based on the 145 repeated administrations and the intraclass correlation coefficient was produced [18].

SCALE DIAGNOSTIC

Scale diagnostics are expressed in logit values, which are used to verify the correct ordering of response categories, check item hierarchy (i.e. assessing the ranking of the nine questions with regard to their difficulty), and target the person mean ability. The logit value is the natural log-odds of a positive reaction to an item, where logit = 0 is the mean item difficulty. In case of the Catquest-9SF, a more positive value means more visual disability. As the items have four response categories, each item thus has three thresholds, which are the tipping points at which an increased disability leads to a higher chance of a higher response category. Item hierarchy and person mean ability are depicted by a person–item map (Figure 1).

UNIDIMENSIONALITY

An important assumption of Rasch analysis is unidimensionality. For the ordering of the items on an underlying trait, it is necessary that aspects not relevant to that

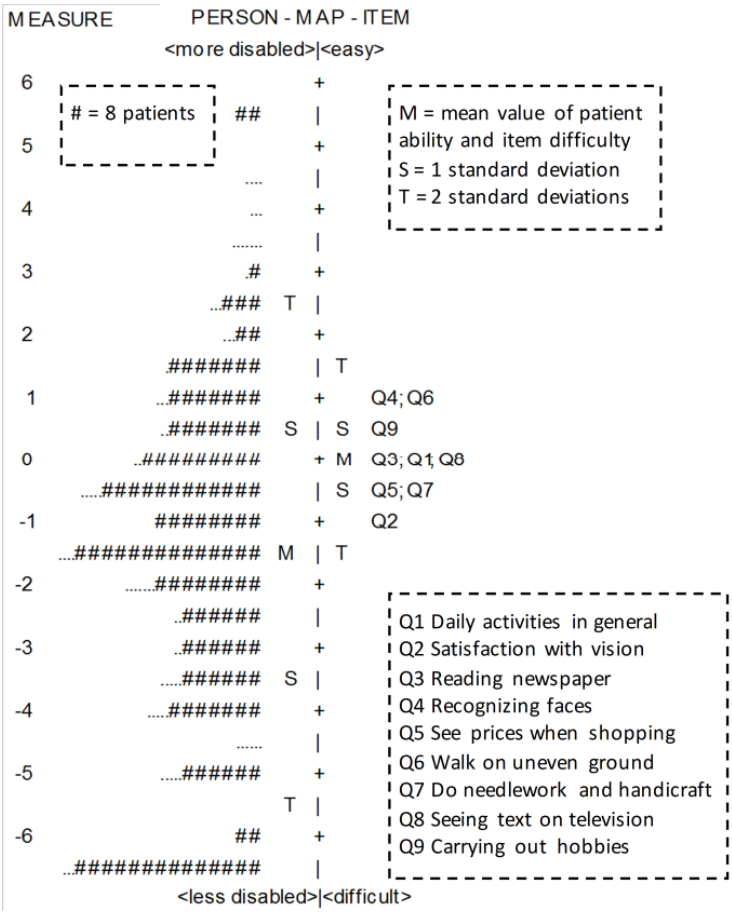


Figure 1 Person-item map of Catquest-9SF. The distribution of Rasch-calibrated patient scores are displayed on the left side. Item locations are displayed on the right side. The map contains both preoperative and postoperative data, resulting in a skewed distribution of patient ability towards the less disabled part, with a mean patient ability of -1.64 logits.

trait do not disturb it. In other words, it is important that the severity of disability that a patient perceives on a certain item (question) depends only on the severity of cataract and not on another (unknown) variable. Both the information-weighted (infit) and the outlier-sensitive (outfit) mean-square statistic (MnSq) are indicators of unidimensionality. The infit gives information on items that are within the person's ability level; the outfit provides information on items that are far beyond a person's ability level, such as failure on easy tasks or success on difficult ones. Both statistics have an ideal value of 1.0; a range between 0.7 and 1.3 is considered an acceptable fit [19]. A value of >1.3 implies too much variance, meaning that the item is measuring something beyond the dimension of interest, resulting in 'underfitting'. A value of <0.7 implies too little variance, meaning that this item is too

predictable and thus does not provide any additional information, which results in 'overfitting' [19]. However, solely using the fit statistic may be inadequate to confirm unidimensionality. Therefore, principal component analysis was also applied [20]. For this purpose, three criteria were used: (1) the variance explained by the first component should be at least 40%; (2) the first eigenvalue should be at least five times greater than the second eigenvalue; and (3) all items should load 0.3 or higher on the first component [21].

DIFFERENTIAL ITEM FUNCTIONING

Items can be more or less difficult for different groups of responders. In Rasch terminology, this means that different groups of responders can have a dissimilar position on the underlying trait, a phenomenon referred to as differential item functioning (DIF). An item shows DIF when subjects from different groups, although equal in the level of their underlying trait, do not have the same probability of endorsing an item [22]. In case of the Catquest-9SF, one would expect people with similar visual disturbances to answer similarly to an item. However, if an item were to show DIF, this would mean that people with similar visual disturbances but of, for example, different age would answer differently to that item. Of note, items that show DIF are not necessarily considered unsuitable. Scores between 0.5 and 1.0 are regarded as minimal DIF, and >1.0 as notable DIF. We analysed DIF between the preoperative and postoperative situation, as well as DIF by gender and age, with cut-offs at 65 and 75 years [1].

Responsiveness

Responsiveness is defined as the ability of a questionnaire to detect a change in the trait that it is supposed to measure. Responsiveness of the Catquest-9SF to cataract surgery was assessed by calculating the difference in a person's ability between the pre- and postoperative situation. The difference was tested for statistical significance with a paired two-tailed t-test. Cohen's d effect size was also calculated. Analysis of variance (ANOVA) was used to determine whether this difference depended on whether one eye or both eyes of a patient were treated (i.e. unilateral versus bilateral surgery).

Quick-access percentile table

For clinical use, we supplied a quickaccess percentile table that can be used by calculating a summary score. This summary score can be calculated by assigning an ordinal score (1, 2, 3 or 4) to each of the response categories and addition of

the nine items of the Catquest-9SF. The relation between the Rasch score and the classical test theory summary score is supplied by WINSTEPS, and the percentiles were calculated in the respective populations. Because the Catquest-9SF provides the opportunity to skip answers (the response 'cannot say/decide'), items remain unanswered relatively often. To make an estimation when one or two of the items are missing, we performed a regression analysis with the Rasch score as dependent and the logit of the sumscores as covariate, see eqn 1.

$$\text{Rasch score} = a \cdot LN\left(\frac{p}{1-p}\right) + b \ , \ (1)$$

where p = (summaryscore–8)/28 – 1/56 in case all items are valid. We then applied the calculated parameters taking for p = (summaryscore–7)/25 – 1/50 in case of eight valid items and p = (summaryscore– 6)/22 – 1/44 in case of seven valid items. With the estimated Rasch parameters, the percentile ranks are calculated.

Clinical relevance: Jacobson and Truax

To determine whether the visual functioning of cataract patients as measured with the Catquest-9SF clinically significantly changed after cataract surgery, the approach of Jacobson and Truax was used [23]. In this approach, the Catquest-9SF scores of both the normal and the cataract population (preoperative) result in a cut-off value and a RCI. The cut-off value distinguishes whether a patient is part of the normal or of the cataract population. It therefore can be used in estimating whether significant improvement with cataract surgery is to be expected. The RCI shows whether a change in Catquest- 9SF score for a patient due to surgery is statistically reliable (without having to cross the cut-off value). Note that the approach from Jacobson and Truax is different from the common definition of 'minimal clinically important difference', which is determined by expert panels or by asking patients to label differences in relevant and not relevant.

Results

Rasch analysis

SEPARATION AND RELIABILITY

Person reliability was 0.87, and item reliability was 0.99. Person separation was 2.56 and item separation was 13.48, in line with the minimum level of performance of 2.00. Cronbach's alpha was 0.945. The item rest correlation of the nine items ranged from 0.694 thru 0.837. Removing any of the items resulted in a decrease of Cronbach's alpha. Test–retest reliability was 0.849, and intraclass correlation coefficient was 0.93.

SCALE DIAGNOSTIC

For each item, the three response category thresholds were correctly ordered (Figure 2). The item hierarchy (i.e. ranking of the questions) had a spread in item difficulty from _1.67 to 1.16 logits (Table 2), and patient ability had a spread of _6.14 to 5.71 logits, with a mean patient ability of _1.64 logits (Figure 1).

UNIDIMENSIONALITY

The infit statistic showed that the questionnaire does not contain misfitting items, while the outfit statistic showed a misfit in one item (Table 2); the 1.50 logit of item 'Recognizing faces' was larger than the set boundary of 1.30. Principal component analysis showed an explained variance of the first component of 69.6% with an eigenvalue of 6.264, which is 11 times larger than the second component (eigenvalue = 0.562). All items had a minimum load of 0.755 on the first component, so all three criteria for unidimensionality based on PCA are met.

DIFFERENTIAL ITEM FUNCTIONING (DIF)

Five items showed minimal DIF between the pre- and postoperative situation. One item ('Satisfaction with vision') had notable DIF (Table 2); compared to other items, patients reported an exceptionally high degree of satisfaction on this item after cataract surgery. This implies that patients with similar visual disturbances, answer differently to this item presurgery compared to post-surgery, in this case they report more dissatisfaction pre- compared to postsurgery. The DIF analysis of patients' sex displayed one item ('Do needlework and handicraft') with minimal DIF. No DIF was found for the different age groups.

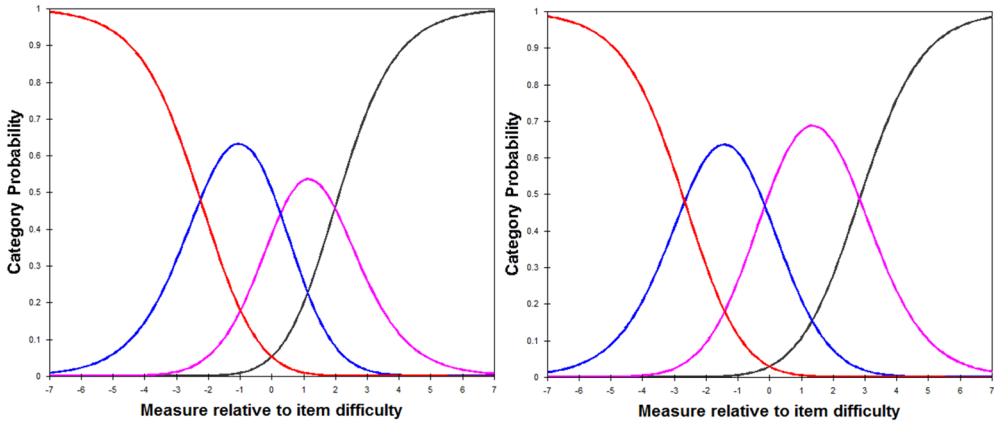


Figure 2 Category probability curves of the "satisfaction with vision" item on the left, and of the remaining eight items with response levels concerning "difficulty" on the right.

Table 2 The Dutch Catquest-9SF questionnaire with item calibration, infit and outfit mean square and standardized fit statistics.

	Item Calibration (SE)	Infit MnSq	Outfit MnSq	DIF Preop to Postop (SE)	DIF Male to Female (SE)	DIF 65- to 65-75 years (SE)	DIF 65-75 to 75+ years (SE)
Daily activities in general	-.17 (.06)	.80	.80	**-.74** (.14)	-.25 (.12)	-.26 (.15)	-.24 (.15)
Satisfaction with vision	-1.67 (.06)	.98	.98	**-1.26 (.13)**	-.21 (.12)	-.08 (.15)	-.26 (.14)
Reading news-paper	-.22 (.06)	1.07	1.09	**.60 (.12)**	-.12 (.11)	-.17 (.14)	.26 (.14)
Recognizing faces	1.16 (.06)	1.24	**1.50**	.20 (.14)	-.10 (.12)	.21 (.14)	-.27 (.15)
See prices when shopping	-.34 (.06)	.89	.89	**.66** (.12)	.00 (.11)	-.17 (.13)	.07 (.13)
Walk on uneven ground	1.12 (.06)	1.09	1.28	.00 (.14)	.42 (.12)	.23 (.15)	.12 (.14)
Do needlework and handicraft	-.33 (.06)	.95	.96	**.69 (.12)**	**.65 (.12)**	.11 (.14)	.10 (.14)
Seeing text on television	.02 (.05)	1.09	1.02	**-.55** (.13)	-.12 (.11)	.21 (.13)	.19 (.13)
Carrying out hobbies	.43 (.06)	.86	.85	.32 (.14)	-.23 (.12)	-.14 (.14)	-.03 (.15)

MNSQ (mean square) should be between 0.70 and 1.30.
DIF [0.5, 1.0] minimal, >1.0 notable DIF

Responsiveness

The pre- and postoperative Rasch scores are displayed in Figure 3. The mean Rasch score improved from -0.56 ± 1.86 to -3.37 ± 2.53 logits, which was statistically significant ($p < 0.001$). Cases below the diagonal line in Figure 3 represent 383 cases (86.7%) reporting a better score on the postoperative Catquest-9SF; the 45 cases (10.2%) above the diagonal line represent a worse score. The change in score after cataract surgery represents an effect size of 1.27. A larger mean improvement in Rasch scores was found when both eyes (mean improvement 3.32) were treated by means of cataract surgery instead of only one eye (mean improvement 2.07; $F_{(1,436)} = 30.179$, $p < 0.001$).

Quick-access percentile table

Regression modelling of the logit of the Rasch scores lead to the equation Rasch = 1.586 * LN(p/(1-p)) – 0.00570, with an explained variance of 0.995. The found percentiles are reported in Table 3.

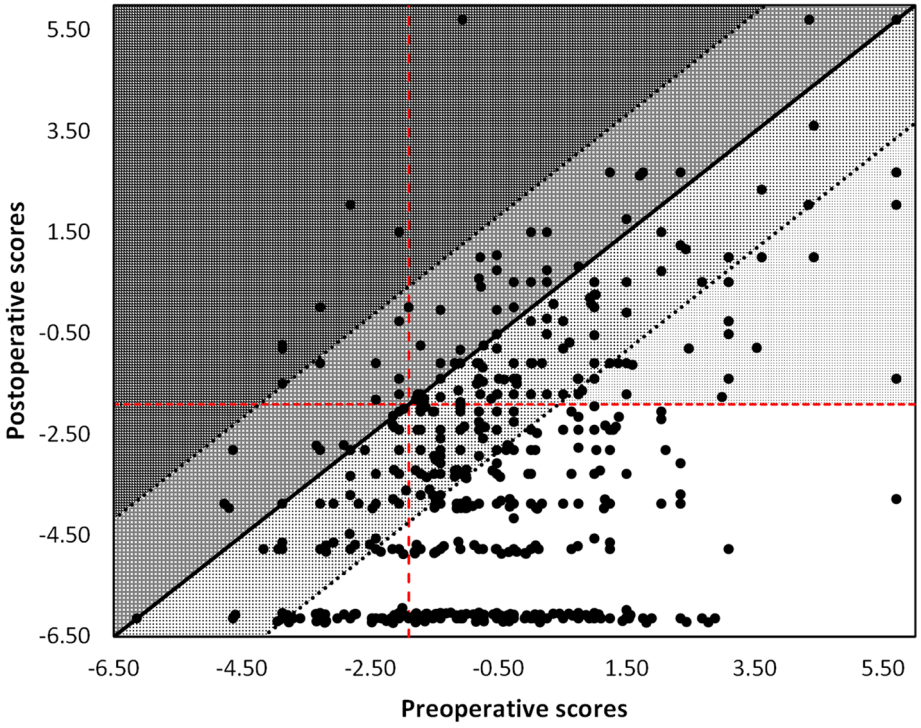


Figure 3 Scatterplot of preoperative versus postoperative Rasch scores. Solid diagonal represents the line of equality. Dashed diagonals represent boundaries of test-retest reliability, which are based on the reliable clinical change index of 2.27(RCI). Cases below the bottom dashed diagonal significantly improved (n=250 [57%]), whereas cases above the top dashed diagonal significantly deteriorated (n=7 [2%]).
The horizontal and vertical dashed lines (at -1.90) resemble the cut-off between cataract patients and the normal population. Note: for the purpose of displaying separate datapoints that are in fact overlapping, small random variation was applied to these datapoints.

1 = patients improving with RCI and fitting the cataract patients preoperative but the normal population postoperative.
1*= patients improving with RCI and fitting the normal population both pre and postoperative.
2 = patients improving with RCI but do not fit the normal population postoperative.
3 = patients with unclear improvement/deterioration, as it is insufficient.
4 = patients deteriorate with RCI and fitting the cataract patients postoperative.

Table 3 Score and percentile ranks for pre, post and norm scores.

		No missings or N/A			1 missing or N/A			2 missings or N/A		
Sum[1]	Rasch[1]	Percentile ranks			Percentile ranks			Percentile ranks		
		Pre	Post	Norm	Pre	Post	Norm	Pre	Post	Norm
7	.							0.01	0.28	0.15
8	.				0.01	0.28	0.15	0.02	0.40	0.33
9	-6.14	0.01	0.28	0.15	0.02	0.39	0.33	0.04	0.51	0.45
10	-4.77	0.01	0.37	0.33	0.04	0.50	0.45	0.09	0.58	0.66
11	-3.87	0.04	0.49	0.45	0.08	0.57	0.58	0.16	0.67	0.76
12	-3.28	0.08	0.56	0.57	0.13	0.63	0.69	0.23	0.72	0.83
13	-2.81	0.12	0.62	0.68	0.18	0.68	0.77	0.31	0.78	0.87
14	-2.41	0.15	0.66	0.76	0.24	0.73	0.83	0.43	0.84	0.93
15	-2.05	0.23	0.72	0.83	0.33	0.78	0.87	0.50	0.87	0.95
16	-1.71	0.30	0.77	0.87	0.43	0.84	0.93	0.57	0.88	0.96
17	-1.40	0.36	0.81	0.90	0.49	0.86	0.94	0.61	0.90	0.97
18	-1.09	0.43	0.84	0.93	0.56	0.88	0.95	0.69	0.92	0.97
19	-0.80	0.47	0.85	0.94	0.61	0.89	0.96	0.73	0.94	0.98
20	-0.52	0.55	0.88	0.95	0.66	0.91	0.97	0.77	0.94	0.99
21	-0.26	0.60	0.89	0.96	0.69	0.92	0.98	0.84	0.96	0.99
22	0.01	0.65	0.91	0.97	0.74	0.94	0.99	0.87	0.97	1.00
23	0.26	0.69	0.92	0.97	0.77	0.95	0.99	0.90	0.97	1.00
24	0.51	0.72	0.94	0.98	0.84	0.96	0.99	0.91	0.98	1.00
25	0.75	0.76	0.94	0.99	0.87	0.97	1.00	0.95	0.99	1.00
26	1.00	0.80	0.96	0.99	0.89	0.97	1.00	0.96	0.99	1.00
27	1.24	0.84	0.96	0.99	0.91	0.98	1.00	0.97	1.00	1.00
28	1.50	0.86	0.97	1.00	0.93	0.99	1.00	1.00	1.00	1.00
29	1.76	0.89	0.97	1.00	0.95	0.99	1.00			
30	2.04	0.91	0.98	1.00	0.96	0.99	1.00			
31	2.34	0.93	0.98	1.00	0.97	1.00	1.00			
32	2.68	0.94	0.99	1.00	1.00	1.00	1.00			
33	3.09	0.96	0.99	1.00						
34	3.61	0.97	0.99	1.00						
35	4.42	0.98	1.00	1.00						
36	5.71	1.00	1.00	1.00						

Sum=CatQuest-9SF summary score; Rasch= Rasch score; N/A not applicable
A 2.27 Rasch score improvement or relapse is assumed a reliable clinical change.
The dashed line resembles the cut-off between cataract patients and the normal population
[1]Lower scores resemble better quality of vision.

Clinical relevance: Jacobson and Truax

Using the method of Jacobson and Truax, a Rasch score cut-off value between the normal population and the cataract population of -1.90 was found, which resulted in a total of 343 (78%) respondents reporting worse than the cut-off value before surgery compared to 124 (28%) after surgery. This last figure was similar to the 24% found in the normal population. A reliable change index of 2.27 was found, which resulted in a total of 250 (57%) patients who improved significantly, while 7 (2%) patients worsened significantly (Figure 3). Of the 99 (22%) patients who were already below the cut-off value before operation (i.e. patients with the least reported problems), 42 (42%) improved significantly and six (6%) deteriorated significantly. Norm scores for cataract patients as well as the normal population are shown in Table 3. Note that, because of the non-linear relationship between the summary score and the Rasch score, when expressed as summary scores, the RCI is about 10 points in the middle range, and 3 or 4 points at the extremes.

Discussion

The aims of this study were to translate the Catquest-9SF into Dutch and to investigate its reliability and validity and provide norm scores, in order to enhance its practical use for ophthalmologists. We confirmed that the Catquest-9SF is a unidimensional questionnaire for assessing visual disability in Dutch cataract patients with a good responsiveness to cataract surgery. The test–retest reliability of the original long version of the Catquest has been established in one of the first publications [3]. To the best of our knowledge, we are the first to show that the short version, the Catquest-9SF, also has a good test–retest reliability and intraclass correlation. We provided a quickaccess table, with relevant percentiles, improving the clinical interpretation of the Catquest-9SF scores. In addition, we proposed norm scores based on visual functioning in the normal population, which can be used as a reference when evaluating visual disability in cataract patients. Finally, we recommended a reliable clinical change score as well as a cut-off value between the normal population and cataract patients.

Similar to our findings, validation studies of the Catquest-9SF in Swedish, Australian, German/Austrian and Chinese populations also showed good performance of the questionnaire. The targeting of item difficulty to person ability and responsiveness to cataract surgery are comparable 2,6, as well as the item and person spread [1,4-7]. The ordering of item difficulty is also comparable, although the item 'walk on uneven ground' seems to be valued slightly different; in the Dutch version, it was associated with a worse vision compared to other translated versions. The

largest, albeit not significant, DIF between the pre- and postoperative situation was observed in the item 'satisfaction with vision' which is in line with the original studies in Sweden [1,5]. Patients report lower satisfaction pre- compared to postsurgery, where a tentative explanation could be that patients presurgery have to cope with the bad news of having cataract, which might causes patients to be unreasonably dissatisfied with their vision. After the operation, they might be positively surprised by their increased vision and reported a relatively extra high satisfaction. The misfit of the item about 'recognizing faces' with a variance of more than 50% in the outfit suggests some people have more trouble with this item than expected, meaning that the item is measuring something beyond the dimension of interest. In other words, the severity of trouble that patients perceive on this item depends not only on the severity of cataract but also on another (unknown) variables. This finding was also reported in these studies.

According to the guidelines developed for the original Catquest, the response category 'cannot say' or 'cannot decide' was treated as missing data [3]. In one study, this resulted in up to 40% of missing values within one item [6]. It is not known how this may influence the results. However, the reliability of cases with many missing values may be reduced. In our samples, the percentage of 'cannot say/ decide'-responses per item varied between 0% and 8% in the patient group and 0–3% in the normal population. For the calculation of the RCI, we used a 'normal population' including a selection of people with healthy vision. This means this population might be somewhat healthier than the general population. However, when comparing patients after cataract surgery with the normal population (Table 3), there is little difference after all.

Patients who reported a relatively good visual functioning before surgery were less likely to report improved visual functioning on the Catquest- 9SF after surgery, which might be the result of regression to the mean or a bottom effect. A bottom effect might be attenuated by adding items with low difficulty, around the -1.90 cut-off, as also suggested by Lin et al. [7]. Another sign of a possible bottom effect postoperatively is the deviation in the targeting of item difficulty to person ability. All of this suggests that the Catquest-9SF could benefit from the inclusion of a number of items which would allow differentiation at the better end of the visual functioning spectrum. However, the main goal of the Catquest-9SF is to differentiate within the cataract population. Additionally, one must always keep in mind that there are some limitations of such a cut-off value. Although the Catquest-9SF contains the most reliable and valid items, some specific tasks, which only few

patients might have trouble with, are not included in the Catquest- 9SF. Therefore, the Catquest-9SF score should be seen as an informative measure, but should not be leading in whether to perform surgery or not. In daily clinical practice, scientific evidence is lacking to aid in identifying patients who are, at that moment, more or less likely to significantly benefit from cataract surgery [24]. Such information can be valuable in patient counselling as it helps to determine and explain the progression of the disease and the optimal time of intervention. It may also help to manage expectations. Only the NIKE (National Indication model for Cataract Extraction) system found an association which might help predicting the outcome of cataract surgery [25]. The cut-off value provided in this paper may also be informative for this purpose. In addition, a measure of the treatment outcome from the patient's perspective can quickly be obtained, which fits the development and promotion of patient-reported outcome measures in the evaluation of treatment [26].

In conclusion, the Catquest-9SF proved to be a suitable measure of subjective visual functioning in the Dutch cataract population. The questionnaire is valid, reliable, unidimensional and responsive to changes after cataract surgery. Applicability in daily clinical practice may be improved without loss of accuracy using summary scores and convert them to percentiles, using the quick-access table provided in this paper. In addition, cut-off values may aid in identifying patients eligible for cataract surgery.

References

1. Lundstrom M & Pesudovs K (2009). Catquest- 9SF patient outcomes questionnaire: nine-item short-form Rasch-scaled revision of the Catquest questionnaire. Journal of Cataract and Refractive Surgery, 35, 504–513.

2. McAlinden C, Gothwal VK, Khadka J et al. (2011). A head-to-head comparison of 16 cataract surgery outcome questionnaires. Ophthalmology, 118, 2374–2381.

3. Lundstrom M, Roos P, Jensen S & Fregell G (1997). Catquest questionnaire for use incataract surgery care: description, validity, and reliability. Journal of Cataract and Refractive Surgery, 23, 1226–1236.

4. Gothwal VK, Wright TA, Lamoureux EL et al. (2009). Catquest questionnaire: re-validation in an Australian cataract population. Clinical and Experimental Ophthalmology, 37, 785–794.

5. Lundstrom M, Behndig A, Kugelberg M et al. (2011). The outcome of cataract surgery measured with the Catquest-9SF. Acta Ophthalmologica, 89, 718–723.

6. Harrer A, Gerstmeyer K, Hirnschall N et al. (2013). Impact of bilateral cataract surgery on vision-related activity limitations. Journal of Cataract and Refractive Surgery, 39, 680–685.

7. Lin X, Li M, Wang M et al. (2014). Validation of Catquest-9SF questionnaire in a Chinese cataract population. PLoS ONE, 9, e103860.

8. Skiadaresi E, Ravalico G, Polizzi S et al. (2016). The Italian Catquest-9SF cataract questionnaire: translation, validation and application. Eye Vision (London), 3, 12.

9. International Consortium for Health Outcomes Measurement (2013). The ICHOM Standard Set for Cataracts.

10. Lundstrom M & Stenevi U (2013). Analyzing patient-reported outcomes to improve cataract care. Optometry and Vision Science, 90, 754–759.

11. Khadka J, McAlinden C & Pesudovs K (2013). Quality assessment of ophthalmic questionnaires: review and recommendations. Optometry and Vision Science, 90, 720–744.

12. Fan XT (1998). Item response theory and classical test theory: an empirical comparison of their item/person statistics. Educational and Psychological Measurement, 58, 357–381.

13. MacDonald P & Paunonen SV (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. Educational and Psychological Measurement, 62, 921–943.

14. Wild D, Grove A, Martin M et al. (2005). Principles of good practice for the translation and cultural adaptation process for patientreported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value in Health, 8, 94–104.

15. Andrich D (1978). A rating scale formulation for ordered response categories. Psychometrika, 43, 561–573.

16. Linacre JM (2014). Winsteps_. Beaverton, Oregon. Winsteps.com.

17. Bond TG & Fox CM (2007). Applying the Rasch Model: Fundamental measurement in the human sciences. 2nd edn. Mahwah, NJ: Lawrence Erlbaum.

18. Bartlett JW & Frost C (2008). Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound in Obstetrics & Gynecology, 31, 466–475.

19. Tesio L (2003). Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. Journal of Rehabilitation Medicine, 35, 105–115.

20. Smith EV Jr. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. Journal of Applied Measurement, 3, 205–231.

21. Hambleton RK, Swaminathan J & Rogers HJ (1991). Fundamentals of item response theory. Newbury Park, CA: Sage.

22. Embretson SE & Reise SP (2000). Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum Associates.

23. Jacobson NS & Truax P (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology, 59, 12–19.

24. Kessel L, Andresen J, Erngaard D et al. (2016). Indication for cataract surgery. Do we have evidence of who will benefit from surgery? A systematic review and meta-analysis. Acta Ophthalmologica, 94, 10–20.

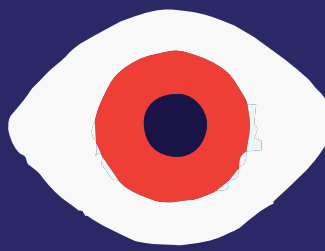
25. Lundstrom M, Albrecht S, Hakansson I et al. (2006). NIKE: a new clinical tool for establishing levels of indications for cataract surgery. Acta Ophthalmologica, 84, 495–501.

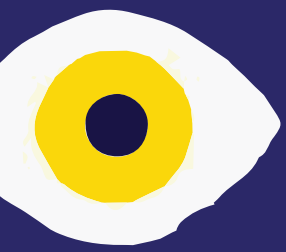
26. Coronini-Cronberg S, Appleby J & Thompson J (2013). Application of patient-reported outcome measures (PROMs) data to estimate cost-effectiveness of hernia surgery in England. Journal of the Royal Society of Medicine, 106, 278–287.

CHAPTER 6

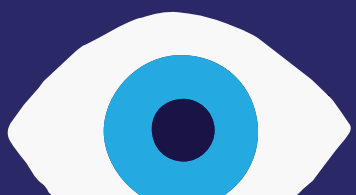
Effects of Clinical Parameters on Patient-Reported Outcome in Cataract Patients: A Multicentre Study

Aline C. Stolk-Vos · Martijn S. Visser · Stijn Klijn · Reinier Timman · Peter Lansink · Rudy Nuijts · Khiun Tjia · Bart Zijlmans · Leonieke W. Kranenburg · Jan V. Busschbach · Nicolaas J. Reus





Acta Ophthalmologica 2018: 96: 586–591



Abstract

Purpose

Ophthalmologists tend to evaluate the results of cataract surgery by focusing on the clinical visual and refractive outcomes and the incidence of complications, where patients' main interest might be their ability to perform daily activities. Therefore, there appears to be a need for optimizing effective communication between patients and ophthalmologist about the outcome of cataract surgery. The aim of this multicentre study was to determine the effects of whether the surgery was performed in one or two eyes, ocular comorbidity and per- and postoperative complications on visual function experienced by patients measured with the Catquest-9SF.

Methods

To measure patient-reported outcomes, Catquest-9SF data were collected between 2014 and 2015 in five Dutch hospitals. Data from 870 pairs of questionnaires - completed before and after cataract surgery - were compared with clinical data. Clinical data, retrieved from patients' medical files, consisted of one or two eye surgery, ocular comorbidity and per- and postoperative complications.

Results

Quality of vision improved more in patients who had surgery in both eyes and had fewer postoperative complications (both $p < 0.001$). We found a nonsignificant trend that quality of vision was worse when ocular comorbidity was present. No significant effect of peroperative complications was observed.

Conclusion

Our results emphasize the added value of the Catquest-9SF as a tool for visual function experienced by patients; the additional information can complement clinical parameters to improve patient-centered approaches in clinical practice.

Introduction

Ophthalmologists tend to evaluate the results of cataract surgery by focusing on clinical parameters such as refractive outcome, postoperative visual acuity (VA) and the incidence of complications. These parameters might be less interesting to patients *per se*, as their main interest is their ability to perform daily activities [1, 2]. These differences in viewpoints may obscure their patient-doctor communication, and thereby, obscure a patient's expectations and satisfaction levels about the outcome of the cataract surgery. As a result, ophthalmologists may be satisfied with the clinical outcomes, whereas patients may not be satisfied with their experienced visual function after cataract surgery. Throughout the manuscript, we apply the term 'visual function' as experienced by the patient, where, in fact, we mean 'vision-related activity limitations'.

To enhance effective patient-centered care, there is a trend towards gathering outcome information from the patient' perspective in addition to the clinical outcomes. As there is interest in patients' side of satisfaction, in terms of outcome, several patient-reported outcome measures (PROMs) have been developed [3]. Several studies have suggested that the use of PROMs has a positive effect on the doctor-patient communication and consequently patients' satisfaction [4]. A validated and short PROM for cataract surgery is the Catquest-9SF, which was developed in Sweden and measures patients' vision-related activity limitations in daily life [5], and has recently been translated into Dutch [6]. It has been shown to be the best-fitting questionnaire to measure the visual function experienced by cataract patients [7].

Several studies have compared patient-reported outcomes measured by the Catquest-9SF and clinical parameters in relation to outcome monitoring [2, 8, 9]. However, this comparison has not been made for the Dutch situation. The aim of this study was to determine the effects of whether surgery was performed on one or both eyes, ocular comorbidity (affecting or not affecting visual function) and per- and postoperative complications on visual function experienced by patients measured with the Catquest-9SF. Our research question was as follows: Is the Catquest-9SF a tool that is of added value to clinical parameters in cataract care? We used the Dutch version of the Catquest-9SF at five eye centers in the Netherlands.

Materials and Methods

Design

Earlier, we performed a study on the linguistic and clinical validity of the Catquest-9SF for the Dutch situation [6]. The current prospective study on the clinical validity was performed as a follow-up of this study. To determine the clinical validation, that

Figure 1 The Catquest-9SF questionnaire

A. Do you find that your sight at present in some way causes you difficulty in your everyday life?

Yes, very great difficulty	Yes, great difficulty	Yes, some difficulty	No, no difficulty	Cannot decide
☐	☐	☐	☐	☐

B. Are you satisfied or dissatisfied with your sight at present?

Very dissatisfied	Fairly dissatisfied	Fairly satisfied	Very satisfied	Cannot decide
☐	☐	☐	☐	☐

C. Do you have difficulty with the following activities because of your sight?

If so, to what extent? In each row place just one tick in the box which you think best corresponds to your situation.

	Yes, very great difficulty	Yes, great difficulty	Yes, some difficulty	No, no difficulty	Cannot decide
Reading text in newspapers	☐	☐	☐	☐	☐
Recognising the faces of people you meet	☐	☐	☐	☐	☐
Seeing the prices of goods when shopping	☐	☐	☐	☐	☐
Seeing to walk on uneven surfaces, e.g. cobblestones	☐	☐	☐	☐	☐
Seeing to do handicrafts, woodwork etc.	☐	☐	☐	☐	☐
Reading subtitles on TV	☐	☐	☐	☐	☐
Seeing to engage in an activity/hobby that you are interested in	☐	☐	☐	☐	☐

Thank you very much for taking part.

is whether the Catquest-9SF is sufficiently sensitive for various clinical situations, we collected follow-up data in cataract patients. The Catquest- 9SF was administered before surgery and 3 months after surgery. This study was approved by the research committee of the Rotterdam Eye Hospital.

Study population

We collected patient-reported outcomes and clinical information of patients undergoing cataract surgery in five Dutch clinics between 2014 and 2015. All patients who had surgery on their first eye during a period of 3 months were invited by care professionals or desk clerks in five clinics; some of these patients had surgery in only one eye, while others had surgery in both eyes. Patients who already had surgery in one eye before the start of the study period were excluded. The participating clinics were as follows: The Rotterdam Eye Hospital, Medisch Spectrum Twente (Enschede), Maastricht University Medical Center, Isala clinics (Zwolle) and VU University Medical Center, Amsterdam.

Patient-reported measures

Visual function experienced by patients was assessed with the Dutch version of the Catquest-9SF [6]. The Catquest-9SF is a nine-item selfreport scale that comprises two parts: the first part contains a global question about difficulties in general to perform daily life activities and a general item about satisfaction with vision. The second part evaluates performances in specific daily activities and patient's general perceptions of difficulties by seven items [3]. Eight items have the following response categories: *very great difficulty; great difficulty; some difficulty; no difficulty; can't say*. One item, satisfaction with vision, scores: very dissatisfied; rather dissatisfied; fairly satisfied; very satisfied; can't say (Figure 1). The Rasch scores range from -5.73 to 5.50; where a score of -5.73 suggests the best visual function and a score of 5.50 the worst visual function, in other words, a lower score indicates fewer problems in performing daily life activities, and a higher score indicates more problems in performing daily life activities. The category can't say is treated as missing data in the analysis according to the instructions of the original guidelines [10]. The Catquest- 9SF has been recommended as a PROM for use in cataract surgery in a review with 48 PROMS [7]. In addition, the International Consortium for Health Outcomes Measurement (ICHOM) has adopted this questionnaire as PROM as part of their standard set for cataract measurement of risk factors and outcomes [11].

Clinical measures

Clinical measures were derived from patients' records and included gender, age, date of surgery, surgery in both eyes (first eye and second eye), visual acuity before and after surgery and type of implanted intraocular lens [i.e. (toric) monofocal or (toric) multifocal]. Also included were data on the presence of ocular comorbidity. We attempted to distinguish comorbidity affecting and not affecting visual function. Comorbidity affecting visual functioning included the following: retinal detachment, amblyopia, diabetic retinopathy, age-related macular degeneration, glaucoma or macular pucker. Comorbidity not affecting visual function included the following: Fuchs' endothelial dystrophy or previous corneal refractive surgery. In addition, peroperative complications (yes, no) were noted, if yes: anterior capsule rupture, nuclear drop, zonulolysis ≥3 hr, conversion into extracapsular cataract extraction, anterior chamber bleeding, iris prolapse and other. Also, postoperative complications (yes, no) were registered, if yes: endophthalmitis, cystoid macular oedema (CME), corneal oedema, anterior chamber tingle, posterior capsule opacification, retinal detachment, wound leakage, vitreous haemorrhage and other. Four of the clinical measures were considered of primary interest: (1) surgery in one versus two eyes, (2) ocular comorbidity, (3) peroperative complications and (4) postoperative complications.

Statistical analysis

Descriptive statistics were used to present patient characteristics. Catquest- 9SF scores were estimated with a Rasch model, specifically the generalized partial credit model (gPCM) model, using Winsteps 4.0.0 [12]. Generalized partial credit model (gPCM) is a model derived from Item Response Theory that can handle items with ordered categories and takes into account the place on the latent trait and the discriminative value of the items. The person measures estimated by this Rasch model are used in a multilevel regression analysis. Three potential levels were included in the models, institute as upper level, patients as middle level and their two repeated measures as the lowest level. The need for the institute level was tested with a deviance test using restricted maximum likelihood [13]. Time, gender, age, operated on one or both eyes, ocular comorbidity, per- and postoperative complications and their interactions with time were postulated as fixed effects. For answering our research question, the interaction effect of these covariates on the difference between pre- and postoperative Catquest-9SF scores are of primary interest. Gender and age are included as control variables. Cohen's *d* effect sizes were calculated using the standard deviation derived from the model's variance estimations. Effect sizes >0.20 are considered small, >0.50 medium and >0.80 large

[14]. Women at mean age were taken as reference group. Visual acuity was not included in this analysis as many data were missing in the clinical patient records. Multilevel regression analyses were performed with IBM SPSS statistics 21.0.1 (Armonk, NY, USA: IBM Corp). As we test the potential influence of four covariates, we applied a Bonferroni correction and considered a level of $p < 0.0125$ significant.

Results

The study population consisted of 870 cataract patients in the five clinics. The majority was female *(n* = 474, 53%), had surgery in both eyes *(n* = 509, 59%) and had ocular comorbidity *(n* = 505, 58%). Table 1 shows the descriptive characteristics of the study sample. In total, 3% of the patients had peroperative complications *(n* = 27) and 6% had postoperative complications *(n* = 50). The most reported postoperative complication was CME *(n* = 16, 2%). Patients with ocular comorbidity had three times more postoperative complications (7.9%) compared to patients without ocular comorbidity (2.8%). More details are shown in Table 1.

The deviance test pointed out that it was not necessary to apply a three-level model. The two-level model, without the institute level, was not significantly worse (=3.287, $p = 0.07$). Pre- and postoperative Catquest-9SF scores had a medium-sized correlation ($r = 0.360$, $p < 0.001$).

For all patients, large improvements in Catquest-9SF scores were observed after surgery (Tables 2 and 3 for results with Catquest-9SF Rasch scores). Men and older patients had overall significantly lower (i.e. better visual function) scores at the start of treatment ($p < 0.001$ resp. $p < 0.001$) and did not show more improvement after surgery compared to women at mean age (additional Cohen's $d = 0.05$, $p = 0.555$; $d = 0.00$, $p = 0.929$). Patients who had surgery in both eyes had significantly worse baseline scores but showed much more improvement in Catquest-9SF scores (d = -0.59, $p < .001$).

We found no baseline differences ($p = 0.289$) between the group with and the group without ocular comorbidity. We found a nonsignificant trend that treatment was worse when ocular comorbidity was present ($p = 0.025$, Tables 2 and 3). Of the 550 patients with ocular comorbidity, we were able to classify 285 patients with ocular comorbidity affecting visual function, and 35 patients as having ocular comorbidity that did not affect visual function. We had insufficient information to classify the 311 other ocular comorbidities. The 285 patients with ocular comorbidity that affected visual function reported significantly worse baseline scores ($p = 0.004$, not in Table) compared to patients without ocular comorbidity. No significant relative treatment effect was observed in this group ($p = 0.098$). No significant differences were found at baseline between patients with ocular comorbidity that did not affect

Table 1 Patients' characteristics.

	Total
Patients, *n (%)*	870 (100)
Gender, male *n (%)*	396 (47)
Age, mean ±SD	72 ±9.9
Surgery in both eyes, *n (%)*	509 (59)
Ocular comorbidity, *n (%)*	505 (58)
Affecting vision	285 (33)
• *Retinal detachment*	34 (4)
• *Amblyopia*	30 (3)
• *Diabetic retinopathy (DRP)*	25 (3)
• *Age-related macular degeneration (ARMD)*	100 (12)
• *Glaucoma*	97 (11)
• *Macular pucker*	28 (3)
Not affecting vision	35 (4)
• *Fuchs endothelial dystrophy (FED)*	31 (4)
• *Previous corneal refractive surgery*	4 (<1)
Other	311 (36)
Peroperative complications, *n (%)*	27 (3)
Anterior capsule rupture	10 (1)
Nuclear drop	3 (<1)
Zonulolysis ≥3 hours	5 (<1)
Other	9 (1)
Postoperative complications, *n (%)*	50 (6)
Endophthalmitis	11 (1)
Cystoid macular edema (CME)	16 (2)
Corneal edema	9 (1)
Other	14 (2)
IOL type, *n (%)*	
Monofocal	651 (88)
Toric monofocal	78 (11)
(Toric) multifocal	10 (1)
Preoperative sum score Catquest, mean ±SD	-0.43 ±1.70
Postoperative sum score Catquest, mean ±SD	-3.33 ±2.16
Difference pre-post	2.90

Table 2 Multilevel regression analysis for Catquest-9SF Rasch scores

	Estimates*	[95% CI]		*p* value
Intercept	-0.49	-0.76	-0.23	<0.001
time (this means: after surgery)	-2.67	-3.01	-2.34	<0.001
Male	-0.41	-0.64	-0.19	<0.001
time * male	0.09	-0.20	0.38	0.554
age	-0.02	-0.03	-0.01	0.001
time * age	0.00	-0.01	0.02	0.856
surgery in both eyes	0.28	0.04	0.51	0.021
time * surgery in both eyes	-0.97	-1.27	-0.67	<0.001
Ocular comorbidity	0.13	-0.11	0.36	0.289
time * ocular comorbidity	0.34	0.04	0.64	0.025
peroperative complications	0.46	-0.22	1.15	0.184
time * peroperative complications	0.13	-0.74	1.00	0.775
postoperative complications	0.21	-0.29	0.70	0.417
time * postoperative complications	1.16	0.53	1.80	<0.001

* Catquest-9SF Rasch scores

Table 3 Estimates pre- and post-surgery Catquest-9SF Rasch scores

	Pre-operative	Post-operative	Change	Cohen's *d*	Additional Cohen's *d**	*p* value
Gender						
Women	-0.49	-3.17	-2.67	-1.60		<0.001
Men	-0.91	-3.49	-2.59	-1.55	0.05	0.554
Ten years older than mean age	-0.70	-3.36	-2.66	-1.59	0.01	0.856
Operated on both eyes	-0.21	-3.86	-3.65	-2.19	-0.58	<0.001
Ocular comorbidity	-0.37	-2.70	-2.34	-1.40	0.20	0.025
Peroperative complications	-0.03	-2.58	-2.55	-1.53	-0.08	0.775
Post-operative complications	-0.28	-1.80	-1.51	-0.90	0.70	<0.001

Notes: Estimates are for females at mean age, unless otherwise specified.

* compared to women at mean age.

P-values are compared to effect for females at mean age, unless otherwise specified.

visual function (n = 35) and patients without comorbidity (p = 0.983). They also had no significant treatment effect (p = 0.409). When confronted with postoperative complications, patients showed significantly less improvement (d = 0.70, p < 0.001).

Discussion

Main findings

The results of this study suggest that the Catquest-9SF is a sufficiently sensitive tool to add a broader perspective on outcome measurement in cataract care. This additional value was demonstrated in several ways. First, surgery led to a large improvement in the visual function experienced by patients. This improvement in visual functioning was particularly strong in patients who had surgery on both eyes. Second, as was to be expected, we found that postoperative complications had a negative effect on the visual function experienced by patients. More interesting is the absence of an effect of peroperative complications. A possible explanation for this could be that, when a surgeon handles the perioperative complication adequately, there may be no effect on the patient's visual function three months after surgery (the moment when the postoperative questionnaire was completed). Alternatively, only 27 (3%) cases of peroperative complications were observed. This number may be too low to find a significant effect. Comparable low rates of perioperative complications have been reported elsewhere (Eloranta & Falck 2017). The perioperative complications found in this Finnish cohort rarely led to long-term complications, which is in line with our results [15].

Third, although, on average, older patients reported a better experienced visual function before surgery, we did not observe a differential age-related effect of surgery.

Relationship to previous studies

As previously described for cataract care [8, 16], the outcomes of the Catquest-9SF are often closely related to clinical outcomes. Nevertheless, the Catquest-9SF rather adds a broader perspective on outcome measurements in cataract care. It emphasizes the importance of complementing clinical outcome measures with an additional instrument. A recent study suggests that existing instruments such as the Catquest-9SF may even be extended by including negative dysphotopsia complaints, as it seems that these complaints are underreported after cataract surgery [17].

Many of our findings were in line with those of other studies. We found that patients who experienced good preoperative visual function were more likely to experience less improvement in Catquest-9SF score by cataract surgery [8, 9]. We also found that patients who underwent second-eye surgery were more likely to have a

better Catquest-9SF score after surgery compared with those undergoing first-eye surgery [16, 18].

Ocular comorbidity tended to have a negative effect on the treatment, but this effect was not significant ($p = 0.025$), applying a Bonferroni correction. This trend is in line with the findings of Ronbeck et al. [16], in contrast with the findings of Grimfors et al. [19].

In contrast with other studies, patients in our study with peroperative complications were not more likely to have a worse postoperative Catquest- 9SF score [8, 9] compared to patients without peroperative complications. Apparent discrepancies in this respect may be due to the fact that other studies only looked into 'capsule complications', and we studied all reported peroperative complications.

Alternatively, we may not have found a significant relationship because of the small number of administered peroperative complications *(n* = 27, 3%) in our sample. Furthermore, the other studies were performed longer ago. Therefore, the surgeons in our study may have had access to newer and improved equipment. In addition, new insights into handling peroperative complications may have had an impact on the visual function outcome with a peroperative complication.

Strengths and limitations

The benefit of complementing clinical parameters with the Catquest-9SF score in our study is subject to uncertainties due to various factors including methodological constraints. Although professionals were carefully instructed about our study, ophthalmologist did not always note the visual acuity of the patients before and after the first and second cataract surgery in the patient records. Because of this missing data, it was not possible to better determine the relation with visual acuity.

Another limitation of our approach is that we have no data about the severity of the ocular comorbidities. The type of ocular comorbidities was reported in the patients' records, but not the grade of the severity. As we found that patients with ocular comorbidity more often have postoperative complications, it would be interesting to study whether this relation might depend on the severity of the ocular comorbidity. Additionally, we had limited information to distinguish between vision-related and not related comorbidity, this may have caused a reduced power to find significant effects, for example, patients with ocular comorbidity not affecting vision could have profited more from the operation.

Despite these limitations, a strength of our approach is the availability of data from five hospitals spread across the Netherlands. Thereby, the study population included is a good reflection of the Dutch population of cataract patients.

IMPLICATIONS

Our findings suggest that the Catquest- 9SF – because it reflects patients' views – is a valuable instrument for ophthalmologists who want to gain insight into patients' experiences of outcome after cataract surgery. It is known that patients undergoing elective surgery who have a better preoperative understanding and more realistic expectations report better experiences. Therefore, when informing patients about cataract surgery prior to their surgery, ophthalmologists should include information derived from the Catquest-9SF.

Besides that, the Catquest-9SF can be used to reflect the patients' view, and it can also be used to optimize the cataract care pathway. We found, for example, different starting points in Catquest-9SF score between hospitals. This information may help to answer questions such as follows: which hospitals were doing better, the ones who did surgery in an earlier stage of cataract or the ones who did surgery in a later stage? Even though we have not answered such questions in this study, it would be interesting for ophthalmologists and policymakers to discuss how to deal with such issues.

Further, the Catquest-9SF can enhance the communication between patients and ophthalmologists about the results of cataract surgery. Knowledge of the experiences of previous patients in relation to outcome can support, for example, a better understanding of ophthalmologists regarding patients' expectations, help patients to make better informed decisions and have more realistic expectations about the care they will receive. The subjective measurement by the Catquest-9SF is important to gain understanding of patients' perspectives on the effects of cataract surgery; it can give more information about patients' satisfaction than more objective clinical parameters such as postoperative visual acuity or incidence of complications.

Although the Catquest-9SF appears to be a promising instrument for use in clinical practice, several barriers remain related to its use. First, Catquest- 9SF data should be collected on regular base in all cataract patients and added to the medical files of patients. Implementation of such regular patient-reported data collection and addition to medical files seems difficult in practice.

In the Netherlands, data about outcomes for cataract surgery are collected by the Dutch ophthalmic society (NOG), a professional association of ophthalmologists. However, these data are not for public use, but for internal use by members of the NOG. The second barrier, therefore, is finding a way to use the data for public use. The Swedish National Cataract Register which collects nationwide data on cataract surgeries seems to be a successful example of how data seem to improve knowledge about trends and results [20].

Conclusion

This study emphasizes the added value of the Catquest-9SF as a measure for visual function experienced by patients. In summary, we conclude that the additional information from the Catquest- 9SF to clinical parameters as we have reported here can improve patientcentred approaches in clinical practice.

References

1. Denniston AK, Kyte D, Calvert M & Burr JM (2014). An introduction to patient-reported outcome measures in ophthalmic research. Eye, 28, 637–645.

2. Fung SSM, Luist J, Hussain B et al. (2016). Patientreported outcome measuring tools in cataract surgery: clinical comparison at a tertiary hospital. Journal of Cataract and Refractive Surgery, 42, 1759–1767.

3. McAlinden C, Gothwal VK, Khadka J et al. (2011). A head-to-head comparison of 16 cataract surgery outcome questionnaires. Ophthalmology, 118, 2374–2381.

4. Nilsson E, Orwelius L & Kristenson M (2016). Patient-reported outcomes in the Swedish National Quality Registers. Journal of Internal Medicine, 279, 141–153.

5. Lundström M & Pesudovs K (2009). Catquest-9SF patient outcomes questionnaire: nine item short-form Rasch-scaled revision of the Catquest questionnaire. Journal of Cataract and Refractive Surgery, 35, 504–513.

6. Visser MS, Dieleman M, Klijn S et al. (2017). Validation, test-retest reliability and norm scores for the Dutch Catquest-9SF. Acta Ophthalmologica, 95, 312–319.

7. Khadka J, McAlinden C & Pesudovs K (2013). Quality assessment of ophthalmic questionnaires: review and recommendations. Optometry and Vision Science, 90, 720–744.

8. Lundström M & Stenevi U (2013). Analyzing patient-reported outcomes to improve cataract care. Optometry and Vision Science, 90, 754–759.

9. Mollazadegan K & Lundström M (2015). A study of the correlation between patientreported outcomes and clinical outcomes after cataract surgery in ophthalmic clinics. Acta Ophthalmologica, 93, 293–298.

10. Lundström M, Roos P, Jensen S & Fregell G (1997). Catquest questionnaire for use in cataract surgery care: description, validity, and reliability. Journal of Cataract and Refractive Surgery, 23, 1226–1236.

11. International Consortium for Health Outcomes Measurement (2017). Cataracts: data collection reference guide. [Online]. Available at: http://www.ichom.org/download/cataracts-reference-guide/. Accessed on 28 Feb 2018.

12. Linacre JM (2018). Winsteps_ Rasch measurement computer program. Beaverton, OR: Winsteps.com.

13. Singer JD & Willett JB (2003). Applied longitudinal data analysis – modeling change and event occurrence. Oxford: Oxford University Press.

14. Cohen J (1992). A power primer. Psychology Bulletin, 112, 155–159.

15. Eloranta H & Falck A (2017). Is an ophthalmic check-up needed after uneventful cataract surgery? A large retrospective omparative cohort study of Finnish patients. Acta Ophthalmologica, 95, 665–670.

16. Ronbeck M, Lundström M & Kugelberg M (2011). Study of possible predictors associated with self-assessed visual function after cataract surgery. Ophthalmology, 118, 1732–1738.

17. Makhotkina NY, Nijkamp MD, Berendschot TTJM et al. (2017). Effect of active evaluation on the detection of negative dysphotopsia after sequential cataract surgery: discrepancy between incidences of unsolicited and solicited complaints. Acta Ophthalmologica, 96, 81–87.

18. Lundström M, Stenevi U, Thorburn W & Roos P (1998). Catquest questionnaire for use in cataract surgery care: assessment of surgical outcomes. Journal of Cataract and Refractive Surgery, 24, 968–974.

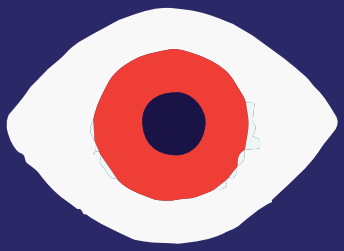
19. Grimfors M, Mollazadegan K, Lundström M& Kugelberg M (2014). Ocular comorbidity and self-assessed visual function after cataract surgery. Journal of Cataract and Refractive Surgery, 40, 1163–1169.

20. Lundström M, Stenevi U & Thorburn W (2002). The Swedish National Cataract Register: a 9-year review. Acta Ophthalmologica, 80, 248–257.

CHAPTER 7

A Very Short Version of the Visual Function Questionnaire (Vfq-3oo7) for Use as a Routinely Applied Patient-Reported Outcome Measure

Martijn S. Visser · Reinier Timman · Karlijn J. Nijmeijer · Hans G. Lemij · Emine Kilic · Jan J. van Busschbach



Acta Ophthalmologica 2020

Abstract

Background

Patient-reported outcome measures (PROMs) are valuable supplements in regular care to facilitate routine monitoring of quality of life from the patient's perspective. The 25-item National Eye Institute Visual Function Questionnaire (NEI-VFQ-25) is a widely used PROM in ophthalmology. However, the NEI-VFQ-25 is too time-consuming and cumbersome for routine evaluations in regular care. The aim of this study is to construct a 7-item questionnaire of which only 3 items are presented to the patient, by means of routing. This VFQ 3 out of 7 (VFQ-3oo7) should have a minimal loss of information compared with the NEI-VFQ-25.

Methods

An historical database including 3293 administrations of the NEIVFQ- 25 was constructed involving patients with retinal detachment, cataract, corneal diseases, glaucoma, macular degeneration, uveal melanoma and a normal population sample. The data were subjected to Rasch analyses, in particular a generalized partial credit model. Items were sorted on the latent trait and divided into seven categories. From each category, the item with the highest discriminative value was selected. Through routing, only three out of the seven remaining questions are used, where the answers navigate patients to a fitting trait level.

Results

A one-dimensional structure was considered fitting. The VFQ-3oo7 showed a small loss of information compared with the total score of the NEIVFQ- 25: correlation 0.927 and a relative precision of 0.868.

Conclusion

The very short, but valid, VFQ-3oo7 can be applied to evaluate the patient's perceived vision-related health status in routine evaluations of treatments in regular care, with a small burden for patients.

Introduction

Patient-reported outcome measures (PROMs) summarize the patients' perceived functional ability, health and well-being (Michelotti et al. 2017). Patient-reported outcome measures (PROMs) in ophthalmology are considered a valuable supplement to medical outcomes in (cost)effectiveness evaluations in clinical trials and quality improvement at a population level (Somner et al. 2012; Denniston et al. 2014). Moreover, PROMs can be valuable in the consultation room in regular care to routinely monitor the patient's perspective, consequently stimulating patient participation and shared clinical decision-making (Boyce et al. 2014; Fung et al. 2016). With such routine measurements, PROMs are also helpful in making the quality of care more transparent to patients, the government and financing bodies such as insurers. However, systematic, routine measurement of PROMs does not take place in ophthalmology (Michelotti et al. 2017). For such routine use in regular care, measurement instruments should be short, practical and useful (Somner et al. 2012). Most of the available PROMs are valid and reliable for research, but not for such routine use (Somner et al. 2012; Michelotti et al. 2017).

When patients administer PROMs and other questionnaires on a routine basis to systematically provide data on the quality of treatments in terms of treatment outcome, this is referred to as routine outcome monitoring (ROM). For example, the National Health Service introduced in 2009 the routine use of PROMs for hip surgery, knee surgery, hernia repair and the treatment of varicose veins. Over 100 000 cases are administered each year (NHS 2018). The International Consortium for Health Outcome Measurement (ICHOM) provides another example, aiming to settle an international standard for routine administration of PROMs based on the framework developed at the Harvard Business School by Michael Porter (ICHOM 2018). In The Netherlands, healthcare providers and the ministry of health organized an online platform in 2017 for the benchmarking of PROMs (Zorgladder 2018). All patients administer a set of questionnaires at predetermined points during the therapy. In all these initiatives, the aim is to make the quality of treatments more insightful.

The National Eye Institute Visual Function Questionnaire (NEI-VFQ-25) is a widely used patient-reported outcome measure (PROM) in clinical trials in Ophthalmology (Mangione et al. 2001). The NEI-VFQ-25 is applied in many eye disorders, such as cataract, age-related macular degeneration, diabetic retinopathy, retinal detachment, corneal disease, uveal melanoma, glaucoma, retinitis and patients with low vision from any cause. However, the NEI-VFQ-25 with its current length of 39 items is too time-consuming and cumbersome for patients for routine use in regular care. This could particularly be a burden for patients with eye disorders. First, problems with their eyesight may cause more difficulty in reading the

questionnaires. Second, eye disorders present themselves mainly in elderly, who often suffer from a significant comorbidity. A high comorbidity may require more follow-up visits and treatments, with more frequent PROMs as a consequence.
We targeted to reduce the administrative burden of the NEI-VFQ-25 and potentially prevent the phenomenon of respondent burden. Respondent burden occurs when respondents' motivation drops as a result of the length of a survey and the data quality begins to deteriorate. The aim of the current study was to construct a very short version, suitable for routine use in regular ophthalmic care with a minimal loss of information. So, the focus was to retain the range of the latent trait or traits as wide as possible while the scale will still be sensitive for patients with severe as well as with mild visual problems.
The current study is not the first attempt to shorten the NEI-VFQ-25, although our aims were more rigorous in reducing the number of items, our methods also differed from those in previous studies and the range of ophthalmic diseases is much broader. Fukuhara et al. (2013) constructed an 11-item short version of the NEI-VFQ-25 using item response theory also referred to as Rasch analysis from glaucoma, cataract and macular degeneration data. They intended to retain information on all domains, and therefore included at least one item on each of these domains. In another study, by Kowalski et al. (2012), six items of the NEI-VFQ-25, without the additional 14 items, were selected, based on Rasch analysis from glaucoma and macular oedema data. The item selection was based on the goodness of fit of the items. In an iterative procedure, the ill-fitting items were removed. Also, Kowalski et al. selected at least one item per domain. Rasch analysis makes the assumption of uni-dimensionality. In 2010, Pesudovs reported issues concerning the NEI-VFQ-25 caused by multidimensionality in cataract patients (Pesudovs et al. 2010). He suggested using two domains: visual functioning and a socio-emotional scale.
The operationalization of our aim was to create the 'VFQ 3 out of 7' (VFQ-3oo7), by reducing the NEIVFQ- 25 to seven items. By using a smart routing, the patient would only have to answer 3 of the 7 items, as items out of the range would not be presented to the patient. For instance, if the first item already indicated that the patient had severe visual problems, items about minor problems would not be presented.

Methods

The Medical Ethics Committee of the Erasmus Medical Centre (Rotterdam, the Netherlands) judged that according to Dutch law, this study did not require a formal approval, as the data were anonymized and had been collected in previously approved studies.

Study sample

A sample of 2383 patients was collected from archival data pertaining to various eye disorders. Several of our data sources have been described before: corneal disease (van Cleynenbreugel et al. 2014), glaucoma (Islamaj et al. 2018), macular degeneration (Lushchyk et al. 2013), uveal melanoma (van Beek et al. 2018) and retinal detachment (de Jong et al. 2017). An exception was the cataract data and some of the macular degeneration data, which were both collected in the Rotterdam Ophthalmic Institute, but that had not been published before. The number of patients and the distribution of background variables published previously differed slightly from the data currently presented, because we used different inclusion and exclusion criteria. We only excluded patients that failed to fill in the questionnaire. In order to enhance the generalizability, we collected an additional sample of 910 people from the general population, stratified for age and gender. In total, we had 3293 administrations of the NEI-VFQ-25. The background variables are described in Table 1.

NEI-VFQ-25

The NEI-VFQ-25 is a vision-specific QoL questionnaire consisting of a 25- item base set of questions and a supplement of 14 additional items measuring vision-related QoL. The NEI-VFQ- 25 can be summarized into a 'total component score' and generates the following domains: global vision rating, near activities, distance activities, limitations in social functioning, role limitations due to vision, dependency, mental health, driving, peripheral vision, colour vision and ocular pain, ranged from 0 to 100. Most NEI-VFQ- 25 items include five Likert scale answer categories. However, 19 items also include a sixth 'opting out' category, 'Stopped doing this for other reasons or not interested in doing this', which is treated as a missing value.

Item selection

In a first selection, we excluded items with more than 10% missing values. This was a strict criterion, as for the proposed 'routing procedure' (see below), missing data would have been problematic. The previously described 'opting out' category caused all of the, due to missing values, excluded items. So some items could be dropped beforehand for contextual reasons, as they only applied to subgroups of patients: such as 'car driving' or 'visiting movies, plays, or sports events' and an opting out was present. The NEIVFQ- 25 has two pairs of items which could be seen as duplicates, namely item 1 and A1, and item 2 and A2. The main difference is that the second of both deviated in answer categories compared with all other items, and was therefore excluded.

Table 1 Numbers and participant characteristics

	Population							
	Retinal detachment*	Cataract*	Corneal diseases	Glaucoma	Macular degeneration*	Uveal melanoma	Normal	Total
Baseline	191	124	81	115	336	111	910	1868
Follow-up 1	58		84	113	285	113		653
Follow-up 2			24	105		110		239
Follow-up 3			51	72		105		228
Follow-up 4				51		80		131
Follow-up 5				37		67		104
Follow-up 6						62		62
Follow-up 7						8		8
Total	249	124	240	493	621	656	910	3293
Gender								
Female, *n* (%)	53 (31)	68 (56)	52 (62)	55 (46)	212 (63)	54 (48)	456 (50)	950 (51)
Male, *n* (%)	118 (69)	53 (44)	32 (38)	65 (54)	125 (37)	59 (52)	454 (50)	906 (49)
Mean age ±SD	60.7 ±13.3	69.9 ±10.7	72.0 ± 8.3	59.9 ± 9.1	78.7 ± 8.5	60.4 ±12.7	69.2 ±11.5	69.1 ±12.4
VFQ-25 sum-score ±SD	76.8 ±17.6	72.1 ±14.0	73.6 ±13.2	84.3 ±11.7	59.2 ±19.9	81.7 ±13.3	88.4 ± 9.5	79.6 ±17.5

*For some respondents gender is missing

SD = standard deviation

Following the recommendations made by Reeve et al. (2007), we first performed a confirmatory factor analysis (CFA). In case of a poor fit, a principal component analysis (PCA) was performed to check uni-dimensionality by the following requirements: (i) the explained variance of the first component should be at least 40%, (ii) the first eigenvalue should be at least five times higher than the second one and (iii) items should load at least 0.50 on the first component. A Monte Carlo PCA for parallel analysis was performed as an additional evaluation of the eigenvalues (Watkins 2006). Cronbach's alpha was calculated for the remaining items, as well as the person separation and person reliability indices. These indices are considered better suited as a measure of reliability for Rasch analysis. The person separation index should be at least 2.0 and the person reliability at least 0.80 (Linacre 2019). In case the assumption of unidimensionality was not sufficiently met, a second dimension would be analysed, and a second short form VFQ would be constructed for this dimension. All items complying with these unidimensional requirements were analysed with a generalized partial credit model (gPCM). This is a two parameter Rasch model for ordered categories. The gPCM assumes equal differences between the answer categories over the items. This makes an ordering of the items on the latent trait possible, based on the item measure, and provides item differentiation parameters. Rasch analysis also allows to express the respondent's performance on this same latent trait, the person measure (Embretson & Reise 2013). The operationalization of our aim was to create the 'VFQ 3 out of 7' (VFQ-3oo7), by reducing the NEI-VFQ-25 to seven items. Seven items were deemed sufficient for a broad classification in an computerized administration, where routing reduced the number of presented items to three. The selection was done by classifying the latent variable into seven classes, and from each class, the best discriminating item was selected for the final version of the VFQ-3oo7 (Figure 1).

Validating VFQ-3oo7

To test the statistical validity of the VFQ-3oo7, several analyses were performed.

FIT STATISTICS

Infit and outfit measures are mean squares provided by Winsteps, to detect poorly fitted items. Mean squares greater than 1.0 indicated an underfit to the model, and mean squares less than 1.0 indicated an overfit, where values between 0.7 and 1.3 were considered acceptable (Wright et al. 1994).

DIFFERENTIAL ITEM FUNCTIONING (DIF)

DIF may occur when a test item does not have the same relationship to a latent variable across two or more groups (Embretson & Reise 2013). That means that persons from different groups who have the same position on the latent trait will have a different outcome. In this study, DIF was discerned for the different eye disorders. For large samples, the DIF t-value is unduly often significant(Tristan 2006). To compensate for this, we applied the normalizing procedure described at the Rasch Organization site, and adjusted the standard errors with √(N/100).

MULTILEVEL STRUCTURE

The patient samples included pre-treatment baseline scores and one or more follow-up measures after treatment. All these measures were included. This was not in accordance with the independence of measurement assumption. Generally, this can be overcome by performing multilevel analyses, the persons form the upper level, their repeated measures the lower level.

Unfortunately, the IRT program we applied is not capable of performing multilevel analyses. Therefore, we applied a procedure to estimate the effect of neglecting the multilevel structure (Mallinson 2011). This procedure provides a visual presentation of the deviation caused by the dependency. Additionally, we preferred a formal test for the deviation and calculated the mean absolute difference (MAD), and compared this to the standard errors of the person measures. The MAD then should be within the 95% confidence interval of the gPCM person measures, thus lower than 1.96 times the standard error. This procedure is described in more detail in Appendix S1 and Figure S1-S5.

PRECISION

In order to determine the sensitivity level of the VFQ-3oo7, we used the relative precision method (McHorney et al. 1992; Gothwal et al. 2009). Although Gothwal et al. (2009) applied F-tests for the calculation of precision, we preferred random effect models as these make more efficient use of all data. We used the data of our largest longitudinal sample and calculated the relative precision with the t-values, with the Likert-score of the VFQ-25 as reference. We excluded the items on car driving because these were excluded in the first place by Pesudovs et al. (2010), Kowalski et al. (2012) and Fukuhara et al. (2013) as well. The remaining 32 items were Likert scored and Rasch scored, just as the selection of the items made by Pesudovs, Kowalsky and Fukuhara.

DISORDER-SPECIFIC ANALYSES

For practical reasons and optimization of generalizability, one uniform VFQ- 3oo7 is preferred; however, we performed separate analyses for the individual disorders, leading to different versions of the VFQ-3oo7. We applied sensitivity analyses within the various samples in order to decide whether it is worthwhile to have different versions for each particular eye disorder.

ROUTING OF THE VFQ-3OO7

By using routing, the number of presented items was reduced to three, as items out of the range would not be presented to the patient. For instance, if the first item already indicated that the patient had severe visual problems, items about minor problems would not be presented. For administration, the first item to be filled in was in the middle of the latent trait, the second on a quarter or three quarters, depending on the answer on the first item. The routing was determined by the medians. The answer on the second item determined which of the remaining four items will be presented as the third item (Figure 1). To successfully perform routing, items with more than 10% missing values were excluded in an earlier stage, as missing data made routing problematic.

ITEM WEIGHTS FOR CALCULATING THE VFQ-3OO7 SCORE

Lastly, in an iterative procedure, weights for the VFQ-3oo7 score were determined by applying a maximum Pearson correlation with the gPCM measure as criterion. These weights were rescaled so that, after a logit transformation, the scores had a range from 0 to 100, consistent with the range of the NEI-VFQ-25.

Confirmatory factor analysis (CFA) was performed with STATA version 15.1 [StataCorp, College Station, Texas 77845 USA], and gPCM was performed with Winsteps version 4.1.0 [Linacre, J. M. (2018). Winsteps_ Rasch measurement computer program. Beaverton, Oregon: Winsteps.com]. All other analyses were performed with SPSS version 25 [IBM SPSS Statistics for Windows, Armonk, NY: IBM Corp.].

Table 2 Original NEI-VFQ-25+14 items, PCA loadings and number of missing values

Item	Text	PCA loading	Percent missing values
1	General health	0.392	0.8
2	Present eyesight	0.677	1.6
3	Worry about eyesight	0.593	0.2
4	Pain or discomfort	0.285	0.2
5	Difficulty reading ordinary print in newspapers	0.820	0.4
6	Difficulty with work or hobbies	0.798	1.4
7	Difficulty finding something on a crowded shelf	0.820	0.6
8	Difficulty reading street signs or the names of stores	0.813	0.9
9	Difficulty going down steps, stairs, or curbs in dim light or at night	0.755	1.3
10	Difficulty noticing objects off to the side while walking along	0.742	1.0
11	Difficulty seeing how people react to things you say	0.825	0.9
12	Difficulty picking out and matching clothes	0.734	1.5
13	Difficulty visiting people in their homes, at parties, or in restaurants	0.779	1.4
14	Difficulty going out to see movies, plays, or sports events		10.7
15	Are you currently driving, at least once in a while?		0.9
15a	If no: have you never driven a car or have you given up driving?		66.3 *
15b	If you gave up driving: Was that mainly because of eyesight?		86.2 *
15c	If currently driving: difficulty driving during daytime in familiar places		35.4 *
16	Difficulty driving at night		37.2 *
16a	Difficulty driving in difficult conditions		36.4 *
17	Do you accomplish less than you would like because of your vision?	-0.792	0.5
18	Limited in how long you can work or do other activities?	-0.774	1.0
19	Pain or discomfort keeps you from doing what you'd like to be doing	-0.541	0.6
20	Stay home most of the time because of eyesight	-0.722	0.3
21	Frustrated a lot of the time because of eyesight	-0.793	0.5
22	Much less control, because of eyesight	-0.840	0.4
23	Because of eyesight, I must rely too much on what other people tell	-0.835	0.4

24	I need a lot of help from others because of my eyesight	-0.854	0.6
25	I worry about doing things that will embarrass myself or others	-0.752	0.6
A1	How would you rate your overall health, on a 0-10 scale?		1.9
A2	How would you rate your eyesight now, on a 0-10 scale?		1.2
A3	Difficulty reading small print on a medicine bottle, or on legal forms	0.749	1.5
A4	Difficulty figuring out whether bills you receive are accurate	0.839	2.4
A5	Difficulty shaving, styling your hair, or putting on makeup	0.759	2.8
A6	Difficulty recognizing people from across a room	0.802	1.6
A7	Difficulty in active sports or other outdoor activities you enjoy		12.1
A8	Difficulty seeing and enjoying programs on TV	0.814	1.6
A9	Difficulty entertaining friends and family in your home	0.763	1.9
A11a	Do you have more help from others because of your vision?	-0.812	1.4
A11b	Limited in the kinds of things you can do because of your vision?	-0.861	1.4
A12	I am often irritable because of my eyesight	-0.694	2.2
A13	I don't go out of my home alone, because of my eyesight	-0.682	2.3

Shaded items are included in Rasch analysis

PCA = principal component analysis

*these high missing values are partly a result of the limited amount drivers among the respondents

Results

We excluded all items about car driving, because these items included a large number of missing values (Table 2). Item A7 'sports' had 12.1% and item 14 'movies, plays' had 10.7% missing values and were excluded. These missing values were a result of a sixth answer category 'Stopped doing this for other reasons or not interested in doing this'. The additional items A1 and A2 were similar to the regular items 1 and 2, but had a divergent number of answer categories, namely ten instead of five, and hence were also excluded. This resulted in 32 items for principal component analysis (PCA). The confirmatory factor analysis (CFA) indicated an insufficient fit: CFI (0.844), TLI (0.833), RMSEA (0.097) and SRMR (0.051). Following the recommendations by Reeve et al., we then performed an exploratory factor analysis. The explained variance was 56.50% by the first component and 5.05% by the second component.

Table 3 Items in GPCM model sorted by item measure and categorized in seven categories.

Item	Item measure	Discrimi-native value	Original domains	Pesudovs Two scale approach		Kowalski	Fukuhara
2	-2.95	1.00	general vision	V11**			**
3	-1.95	0.54	mental				
A3	-1.75	0.95	near	V15*			**
6	-1.09	1.13	near	V12**		**	**
17	-1.08	1.18	role		S12**	*	**
9	-0.95	0.89	distance	V13**			
5	-0.91	1.07	near	V14**			**
18	-0.74	1.01	role		S10**	**	
A11b	-0.49	1.33	role		S9*		*
8	-0.36	1.06	distance	V10**			**
10	-0.32	0.87	peripheral	V7*			
22	-0.29	1.11	mental		S11**		*
21	-0.12	0.94	mental				
A8	-0.04	1.11	distance	V6*			**
A11a	0.12	1.09	role		S4*		*
7	0.22	1.15	near	V9**		*	
A4	0.26	1.13	near	V4*			*
A12	0.29	0.72	mental				
19	0.37	0.25	pain				
23	0.56	1.11	dependent		S8**		*
A6	0.61	1.06	distance	V3*			**
24	0.63	1.18	dependent		S6*		**
A5	0.66	0.93	near	V2*			
11	0.72	1.16	social		S3*	**	**
25	1.12	0.98	mental		S7**	**	**
20	1.18	0.93	dependent		S5**	**	*
13	1.24	1.11	social		S1**		*
12	1.34	1.02	color	V1*		*	
A9	1.80	1.17	social				
A13	1.92	1.02	dependent		S2*		

Shaded items have the highest discriminative value in a category and are included in the final selection for the VFQ-3oo7

* First selection by Pesudovs, Fukuhara and Kowalski

** Final selection by Pesudovs, Fukuhara and Kowalski.

Kowalski selected item 14, which we did not take into account because of >10% missing values.

Two scale approach of Pesudovs: V=Visual functioning and S=Socioemotional item. Both items were ordered by severity. GPCM = generalized partial credit model

The following eigenvalues for the first four components were as follows: 18.08; 1.62; 1.30 and 1.01, respectively. The Monte Carlo PCA parallel analysis suggested that the first eigenvalue should be at least 1.187, the second 1.164, the third 1.147 and the fourth 1.13. Thus, according to this criterion a second and third factor might be present. However, the solution with two components resulted in one pain item (item 4). The three component solution included the two items on pain only in the third component (items 4 and 19), where the second component was entirely recessive. For this reason, we continued with the one-component solution, and the second and third components were not suitable. This one-component solution yielded two items with a too low component loading (<0.50); item 1 on the general health state (0.39), and item 4 on pain and discomfort around the eyes (0.29, Table 2). Therefore, 30 items were selected for gPCM analyses. These 30 items had a person separation index of 3.79, a person reliability index of 0.93 and a Cronbach's alpha of 0.975.
Ordering of the items on the basis of the latent trait, classification and selection of the most discriminating items per class, resulted in the selection of the items 2, A3, 17, A11b, 24, 11 and A9 (Table 3).

FIT STATISTICS

The item infit mean square measure for the 30 item gPCM analysis was 1.08 and the outfit measure was 0.90. For the seven-item analysis (including all scored categories), these measures were respectively 1.04 and 0.91. All these measures were well within the acceptable range of 0.70 and 1.30.

DIFFERENTIAL ITEM FUNCTIONING (DIF)

Item 2, which was the first item at the best seeing side of the latent variable, showed DIF for uveal melanoma patients (adjusted *t*-value = 2.94; p = 0.003). For these patients, this item was ranked second, while this item was far on the well seeing side for the other six populations. Note that there were 49 DIF tests applied (seven items times seven eye disorders). A Bonferroni correction would result in a corrected significance level of p = 0.001.

MULTILEVEL STRUCTURE

The mean absolute difference (MAD) for the total sample was 0.130, thus lower than 1.96 times the standard error (1.96*0.073 = 0.142). The MADs were larger for retinal detachment (0.161; CI < 0.403), glaucoma (0.172:
CI < 0.214) and uveal melanoma (0.173; CI < 0.243), and smaller for corneal disease (0.101; CI < 0.292) and macular degeneration (0.100; CI < 0.212), but the same held for the standard errors. Within each patient group, the MADs were within acceptable confidence limits.

PRECISION

We applied our data on the treatment of macular degeneration for the precision analyses, as this was our largest sample with at least two measurements. We had 336 baseline measures and 285 follow-up measures. Rasch scoring performed better than Likert scoring for all selections (Table 4). The selection of 11 items by Fukuhara et al. (2013) yielded the largest relative precision, followed by the visual function scale of Pesudovs, the VFQ-3oo7 and the selection of Kowalsky.

DISORDER-SPECIFIC ANALYSES

Rasch analyses within the data of the different eye disorders generally led to other selections of items. In the retinal detachment sample, the solution yielded a correlation of 0.928, which was 0.005 higher than the general solution in this sample (r = 0.923). In all other samples, the sample-specific solution led to a lower correlation than the overall solution.

Table 4 Precision of various selections of the NEI-VFQ-25, based on our macular degeneration sample.

Scale	Method	Number of items	Macular degeneration	
			t-value	Relative precision
VFQ-32	Likert	32	-4,539	100,0
	Rasch	32	-5,366	118,2
VFQ-3oo7 *		3	-3,940	86,8
VFQ-7 **	Likert	7	-4,806	105,9
	Rasch	7	-4,919	108,4
Fukuhara	Likert	11	-4,927	108,6
	Rasch	11	-5.585	123.1
Kowalsky	Likert	6	-2,460	54,2
	Rasch	6	-3,284	72,4
Pesudovs-visual function scale	Likert	6	-4,912	108,2
	Rasch	6	-5.098	112.3
Pesudovs-socioemotional scale	Likert	7	-2,435	53,7
	Rasch	7	-2,894	63,8

* The VFQ-3oo7 has a weighted sum score that is derived from Rasch analyses.
** These are all the 7 items applied in the VFQ-3oo7 without routing.

ROUTING OF THE VFQ-3OO7

The first item presented to every patients was item A11b 'Are you limited in the kinds of things you can do because of your vision'? as it is in the middle of the latent trait (Figure 1). Answer category 'a' led to item A3, where categories 'b' to 'e' led to item 11. Categories 'a' and 'b' on item A3 led to item 2, and categories 'c' to 'e' led to item 17. Category 'a' on item 11 led to item 24 and categories 'b' tot 'e' led to item A9.

ITEM WEIGHTS FOR CALCULATING THE VFQ-3OO7 SCORE

The optimal weights gained from the iterative procedure are presented in Figure 2. This solution resulted in a correlation of 0.924 with the person measures of the 30-item gPCM.

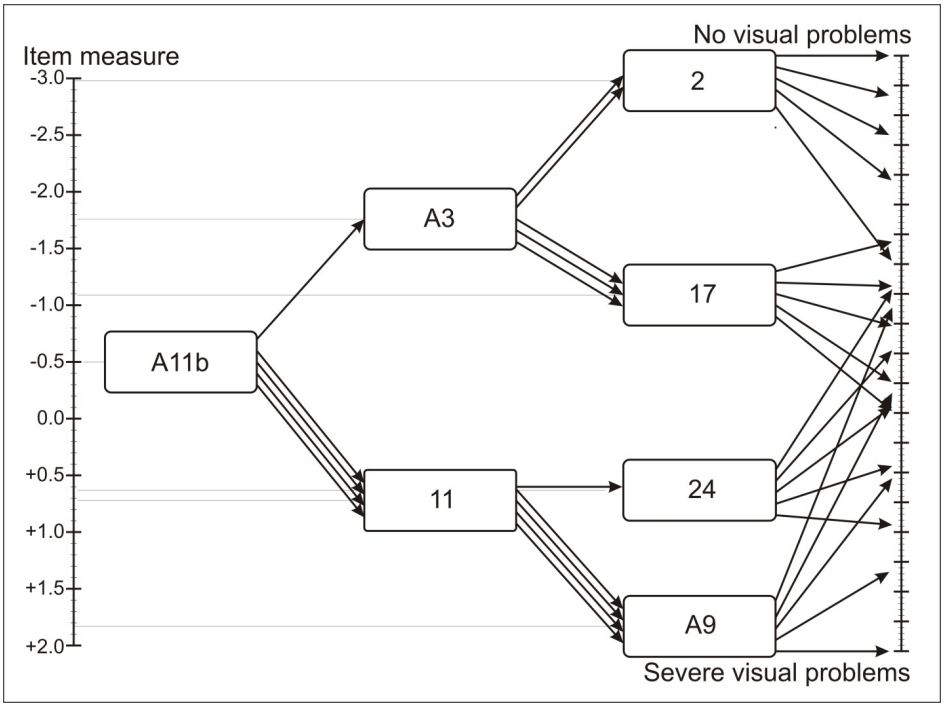


Figure 1 Schematic presentation of the VFQ-3oo7.
Rasch analysis allows expressing the respondents' performance on the same latent trait as the item measure. First item A11b, in the middle of the latent trait, is administered. Depending on the answer the respondent is routed through the questionnaire. Every arrow represents an answer category and the split is determined by the median of the item. In the end only three out of the seven items could be used, where the answers navigate patients to a fitting trait level.

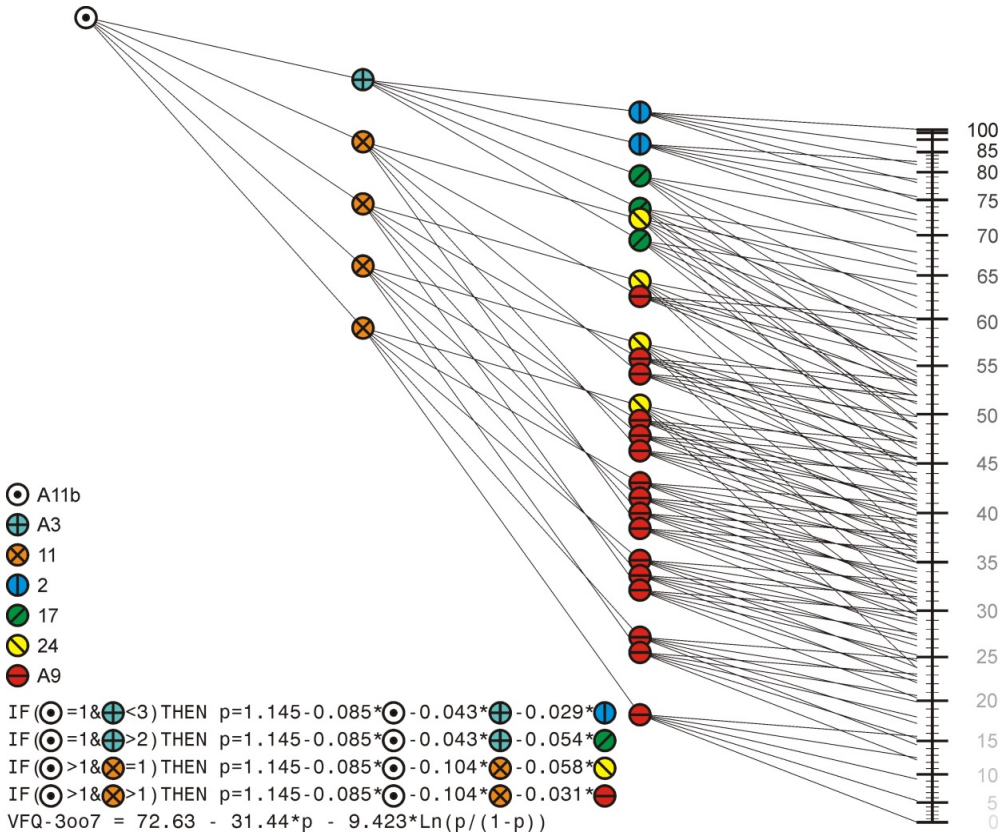


Figure 2 Visual presentation and calculation of the VFQ-3oo7.
The upper left symbol starts at 1.145. Then 0.085 times the response on the first item (A11b) is subtracted. When this first response is 1, the next question is A3, and 0.043 times the response on A3 is subtracted. The same procedure holds for the third question.
When the response on the first question is larger than 1, the second question is 11, and subsequently 0.104 times the response on this question is subtracted. Then the third question follows in the same way.
Note that the scale is not linear, but a logit scale.
The routing is based on medians, therefore question A9 is most frequent.

Discussion

Principal findings

The main goal of this study, to reduce the number of items for routine use of PROMs in regular care, is particularly important for patients with eye disorders. First, problems with eyesight may cause more difficulty in answering questionnaires. Second, they represent mainly an elderly population, having a higher comorbidity. A higher comorbidity in turn demands more follow-up visits and treatments, with more frequent PROM questionnaires as a consequence. Hence, these patients are probably more subject to respondent burden.

This study reduced the number of items of the NEI-VFQ-25, including additional items, from 39 to the predetermined three administered items out of seven. A one-dimensional structure was considered fitting. During the item selection, five items were removed due to a large number of missing values and two items were removed as a result of similarity to other items. Two other items showed too low component loadings and were therefore excluded. The last step of reducing the remaining 30 items to seven items was accomplished by Rasch analysis. By routing, only three items out of the seven need to be administered. The VFQ-3oo7 showed a small loss of information compared with the total score of the NEI-VFQ- 25: correlation 0.927 and a relative precision of 0.868.

Our principal findings in relation to the existing literature

Successfully shortened versions of the NEI-VFQ-25 have been produced earlier. However, our aims were more rigorous in reducing the number of items to as few as 3 out of 7; in addition, our methods differed from those used before. In concordance with Fukuhara et al. (Fukuhara et al. 2013), we also excluded the three driving items. They used a scree plot and eigenvalue criterion to distinguish the number of dimensions, which more or less matched our uni-dimensionality criteria 2 and 3, namely the first eigenvalue should be at least five times higher than the second eigenvalue, and items should load at least 0.50 on the first component. For item selection, a more stringent criterion was applied than in previous studies (Pesudovs et al. 2010; Fukuhara et al. 2013), which used a loading of 0.40, with an explained variance of 0.40^2 = 16%. A loading of 0.50, indicating an explained variance of 25%, will provide more reliable items.

Fukuhara et al. (2013) and Kowalski et al. (2012) intended to retain information on all domains, and therefore included at least one item on the domains general vision, near activities, distant activities, vision-specific social functioning, vision-specific mental health, visionspecific role difficulties and vision-specific dependency (Table 3). Our study did not aim to include every domain, but to retain the largest range as possible of the visual spectrum. The result is that we did not include items for the distant activity and mental health domains. Pesudovs et al. (2010) discerned two scales: a visual functioning scale and a socio-emotional scale. In our data, we did not find a socio-emotional dimension. When we evaluated the two and three component analyses, these components only included the two items on pain (items 4 and 19).

The precision analyses showed that the selection of Fukuhara et al. (2013) distinguished best between the measurements in our data of patients with macular degeneration. It must be noted that they also used the most items. Fewer items

obviously lead to a lower precision. The six items that are selected for Pesudovs' vision scale perform virtually as well as all 32 items. However, Pesudovs' social scale performs much less in this population. It may be that it better performs in other patient populations. Also, Kowalski's selection does not perform very well in this population. The VFQ-3oo7 performs less than the selections of Fukuhara and Pesudovs. This imprecision is likely to be the price for the large reduction of items, if the relative precision would also be calculated relative to the number of items, the VFQ-3oo7 would score best.

Limitations

Item 2 'present sight' showed marginally significant DIF in the uveal melanoma patients. Generally, the item at the best seeing end of the continuum was present sight. However, in the uveal melanoma population, item 3 'worry about eyesight' was the item that was most sensitive at the best seeing end. Patients with uveal melanoma experience not so much visual problems but have other worries. This may be explained by the dooming and worrisome nature of uveal melanoma and carcinoma in general, as there is a significant risk of losing an eye or even worse, die. Apparently, and logically, item 2 of the VFQ-3oo7 had a slightly different meaning for the uveal melanoma patients, as these patients experience less visual problems. This indicates that the VFQ-3oo7 may be less sensitive for uveal melanoma patients with mild visual problems.

A inevitable consequence of our aim to reduce the VFQ-25 to only three questions to be answered is that it excludes the possibility to cover all original subscales. For specific eye disorders, it can be appropriate to add one or a few relevant items considered indispensable, that is for glaucoma patients the item on pain can be added, in particular, because pain was hardly represented in the one-component solution.

Future perspectives

In this study, we focused on shortening an instrument to measure vision-related health status. However, at least for some eye disorders, additional perceived outcome domains may be relevant. For example for glaucoma, the burden/side effects of treatment have been identified as a relevant outcome domain (Somner et al. 2012). For specific eye disorders, future studies should focus on identifying additional domains and items, while preserving the requirement of a very short instrument.

The latest development in Rasch includes computer adaptive testing (CAT) in the clinical application. This creates interesting possibilities, but one needs to employ a

full CAT infrastructure. With the methodology presented here, such CAT infrastructure is not necessary and thus the VFQ3oo7 can be employed more easily. With the VFQ-3oo7, ophthalmologists have a user-friendly tool to monitor the patients' perspective on their visual functioning in regular care. The use of the VFQ-3oo7 offers possibilities to explore the patients' perspective, without the cumbersome administration of the long original NEI-VFQ-25. In our hospital, the VFQ-3oo7 is now implemented. Patients fill in the computer based questionnaire before each consultation. The outcome is directly presented in the electronic patient file and visible to the ophthalmologist. It is most appreciated in the clinical communication when clinical outcomes contradict with the outcome of the VFQ3oo7 or when the outcome of the VFQ-3oo7 shows signs of improvement or deteriorations over time. It is up to the ophthalmologist to use the outcome of the vfq3oo7. This could aid the dialogue between ophthalmologists and patients, and could help to substantiate a referral to low vision specialists or psychologists.

Conclusions

The goal of this study was to reduce the number of NEI-VFQ-25 items to seven items of which three are to be administered by the patient while retaining a high distinctive capacity, to make it suitable for routine PROM measurement in clinical practice. Correlating 0.927 with the criterion, the VFQ-3oo7 has succeeded in realizing this goal. The VFQ-3oo7 also appeared suitable for various eye disorders. The very short, but valid, VFQ-3oo7 can be applied to evaluate the patient's perceived vision-related health status in routine evaluations of treatments in regular care, with only little burden for patients.

Data availability statement

The VFQ 3oo7 can also be administered on paper, and scored by means of an Excel file that is digital available on request.

References

1. Michelotti M, de Korne DF, Weizer JS et al. (2017): Mapping standard ophthalmic outcome sets to metrics currently reported in

eight eye hospitals. BMC Ophthalmol 17: 269.

2. Somner JE, Sii F, Bourne RR, Cross V, Burr JM & Shah P (2012): Moving from PROMs to POEMs for glaucoma care: a qualitative scoping exercise. Invest Ophthalmol Vis Sci 53: 5940–5947.

3. Denniston AK, Kyte D, Calvert M & Burr JM(2014): An introduction to patient-reported outcome measures in ophthalmic research. Eye (Lond) 28: 637–645.

4. Boyce MB, Browne JP & Greenhalgh J (2014): The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. BMJ Qual Saf 23: 508–518.

5. Fung SS, Luis J, Hussain B, Bunce C, Hingorani M & Hancox J (2016): Patient-reported outcome measuring tools in cataract surgery: Clinical comparison at a tertiary hospital. J Cataract Refract Surg 42: 1759–1767.

6. NHS (2018): National Health Services -Patient reported outcome measures (PROMs).

7. ICHOM (2018): The International Consortium for Health Outcomes Measurement ICHOM.

8. Zorgladder (2018): DataplatformZorgladder. Zorgladder. Received on January 15th, 2019. Accepted on January 25th, 2020.

9. Mangione CM, Lee PP, Gutierrez PR, Spritzer K, Berry S, Hays RD & I National Eye Institute Visual Function Questionnaire Field Test (2001): Development of the 25-item national eye institute visual function questionnaire. Arch Ophthalmol 119: 1050–1058.

10. Fukuhara S, Wakita T, Yamada M, Hiratsuka Y, Green J & Oki K (2013): Development of a short version of the visual function questionnaire using item-response theory. PLoS ONE ONE [Electronic Resource] 8: e73084.

11. Kowalski JW, Rentz AM, Walt JG et al. (2012): Rasch analysis in the development of a simplified version of the National Eye Institute Visual-Function Questionnaire-25 for utility estimation. Qual Life Res 21: 323–334.

12. Pesudovs K, Gothwal VK, Wright T & Lamoureux EL (2010): Remediating serious flaws in the National Eye Institute Visual Function Questionnaire. J Cataract Refract Surg 36: 718–732.

13. van Cleynenbreugel H, Remeijer L & Hillenaar T (2014): Cataract surgery in patients with Fuchs' endothelial corneal dystrophy: when to consider a triple procedure. Ophthalmology 121: 445–453.

14. Islamaj E, Wubbels RJ & de Waard PWT (2018): Primary Baerveldt versus trabeculectomy study after one-year follow-up. Acta Ophthalmol 96: e740–e746.

15. Lushchyk T, Amarakoon S, Martinez-Ciriano JP, van den Born LI, Baarsma GS & Missotten T (2013): Bevacizumab in agerelated macular degeneration: a randomized controlled trial on the effect of injections every 4 weeks, 6 weeks and 8 weeks. Acta Ophthalmol 91: e456–461.

16. van Beek JGM, Buitendijk GHS, Timman R, Muller K, Luyten GPM & Paridaens D, Naus NC & Kilic E (2018): Quality of life: fractionated stereotactic radiotherapy versus enucleation treatment in uveal melanoma patients. Acta Ophthalmol 96: 841–848.

17. de Jong JH, Vigueras-Guillen JP, Simon TC, Timman R, Peto T, Vermeer KA & van Meurs JC (2017): preoperative posturing of patients with macula-on retinal detachment reduces progression toward the Fovea. Ophthalmology 124: 1510–1522.

18. Reeve BB, Hays RD & Bjorner JB et al. (2007): Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 45: S22–31.

19. Watkins MW (2006): Determining parallel analysis criteria. J Modern Appl Stat Methods 5: 344–346.

20. Linacre JM (2019): Reliability and separation of measures.

21. Embretson SE & Reise SP (2013): Item Response Theory. Oxfordshire: Taylor & Francis.

22. Wright BD, Linacre JM, Gustafson J & Martin-L€of P (1994): Reasonable meansquare fit values. Rasch Meas Transact 8: 370.

23. Tristan A (2006): An adjustment for sample size in DIF analysis. Rasch Meas Transact 20: 1070–1071.

24. Mallinson T (2011): Rasch Analysis of Repeated Measures. Rasch Meas Transact 25: 1317.

25. McHorney CA & Ware JE Jr, Rogers W, Raczek AE & Lu JFR (1992): The validity and relative precision of MOS short- and long-form health status scales and Dartmouth COOP charts. Results from the Medical Outcomes Study. Med Care 30: MS253-265.

26. Gothwal VK, Wright TA, Lamoureux EL & Pesudovs K (2009): Rasch analysis of the quality of life and vision function questionnaire. Optom Vis Sci 86: E836–844.

27. Linacre JM (2018): A User's Guide to Winstep Ministep Rasch-Model Computer Programs. Beaverton, OR. Available at: https://www.winsteps.com/index.htm. (Accessed on 3 Sept 2019).

Supporting Information

Additional Supporting Information may be found online:
https://doi.org/10.1111/aos.14378.

- Appendix S1. Rasch analysis of repeated measures.
- Figure S1. Mean Rasch measures of the first 10 retinal detachment patients at baseline and follow-up.
- Figure S2. Mean Rasch measures for the first 10 glaucoma patients at baseline and follow-up.
- Figure S3. Mean Rasch measures for the first 10 patients with corneal diseases at baseline and follow-up.
- Figure S4. Mean Rasch measures for the first 10 patients with macular degeneration at baseline and follow-up.
- Figure S5. Mean Rasch measures for the first 10 patients with uveal melanoma at baseline and follow-up.

CHAPTER 8

General Discussion

This thesis aimed to investigate a variety of ways of using QoL measures in ophthalmology, potentially aiding decision-making in the clinical context, policy-making, and for reimbursement purposes.

In ophthalmology, as in other parts of the healthcare sector, stakeholders strive to ensure that patients obtain the best care available with the ultimate goal of reaching better health. Ideally, this care should be non-invasive, not burdensome, and have direct results without any side effects. More realistically, trade-offs for each individual are made based on all given information. Next, a pros-and-cons list is compiled to support decision-making, where patients regain their best daily functioning during and after illness. Habitually, decisions are made by health care professionals based primarily on medical information, which could be considered incomplete. From a patient perspective, this means that patient preferences, and more specifically QoL outcomes, are not always fully included. Incorporating the patient perspective using QoL questionnaires in clinical practice would support the idea that patients should be treated with all relevant perspectives in mind, rather than simply treating the disease.

Following this line of reasoning, the thesis provides a number of applications of HRQoL measures in ophthalmology, and helps to integrate these measures into ophthalmic care. Insight is offered into the impact of a number of interventions on vision-specific QoL in ophthalmology. In addition, it presents a means by which both the BSE and the WSE can be included in an HRQoL utility framework. Ultimately, short, and thus easy-to-administer vision-specific QoL questionnaires can be introduced into daily ophthalmic practice, enabling routine outcome monitoring. Hence, this thesis is informative for several levels of healthcare stakeholders, e.g. patients, doctors, hospitals, and policymakers, who may interpret the results from their own perspectives. For example, the results can facilitate communication between patients and doctors in daily clinical practice, help set up clinical guidelines, and support reimbursement decisions. This general discussion describes possible applications of the thesis findings.

Patient-reported outcome measures in treatment evaluations and shared decision-making

Although health and patient-reported outcomes (PROs), such as QoL, are highly associated, they are not always aligned. This factor underlines the relevance of using patient-reported outcome measures (PROMs) in ophthalmology. These outcomes enrich treatment evaluations and aim to ensure appraisal is based on complete information. In addition to assessing ophthalmic treatments, PROMs can help to stimulate shared decisions. Shared decision-making implies that the

process is a well-informed dialogue between ophthalmologists and patients. The PROMs presented and investigated in this thesis are the disease-specific QoL questionnaire Catquest-9SF, the vision-specific QoL questionnaire NEI-VFQ-25, and the generic QoL questionnaire SF-6D. These measures help patients incorporate their QoL in treatment evaluations and express QoL considerations in the doctor-patient dialogue. This is essential, as no one other than the patient is better able to value his QoL.

Patients' QoL is incorporated into the treatment evaluation of intravitreal bevacizumab injections for neovascular macular degeneration (**Chapter 2**). Effectiveness in terms of QoL, visual acuity, and contrast sensitivity is assessed in low vision spectacles for distance viewing, E-Scoop (**Chapter 3**). The first example of how to take the patient perspective into account is by quantifying the impact of interventions on QoL in macular degeneration. We show that in patients with macular degeneration the injection interval of bevacizumab can be prolonged without consequences for a patient's QoL (**Chapter 2**). We find that vision-related and general QoL are not affected by lowering this injection frequency to every 8 weeks instead of every 4 weeks. The QoL outcomes align with the clinical outcomes of these injections, e.g. visual acuity and central fovea thickness. Lower frequencies of intravitreal injections have the benefit of fewer injection-related adverse events. So by halving the injection frequency from every 4 to every 8 weeks, the risk of injection-related side effects is halved. Moreover, there are logistic benefits. Lowering the frequency results in a lower hospital visit rate, which saves time for patients, their companions, and doctors. Based on these results, low injection frequencies of intravitreal bevacizumab, in particular the every-8-week regimen, should not be withheld from patients with macular degeneration. Perhaps even longer treatment intervals could be considered if these were to suit the patient's wishes and, clearly, if this were to make sense from a medical point of view. This is a delicate issue in attempting to find the ideal injection frequency when treating patients with macular degeneration, as the disease can flare up when intervals are too long, and this might cause irreversible vision loss. Hence, monitoring the disease is essential.

Another argument for considering this trade-off is that in our studies on macular degeneration, patients have an high average age ranging from 77.3 to 79.5 years. Moreover, in the general Dutch population in age group 75 to 79 years, 81.7% have multimorbidity, which increases with age. As a result, these patients are confronted with a high disease burden and thus a lower base value of QoL, apart from the effects of macular degeneration. In other words, macular degeneration is probably not the main contributor to their QoL. Rather, the range of vision-related QoL in these patients is also much narrower in relation to the entire range of HRQoL. It is

thus debatable whether treating a disease in the worse seeing eye is worthwhile for every patient. In other words, do all patients have to be exposed to all possible interventions? As patients with macular degeneration are not unfamiliar with comorbidity, it could also be expected that choices are being made in favour of QoL over and above above HRQoL.

Regardless of the choice of injection frequency, macular degeneration progresses for a number of patients, resulting in serious vision loss. These patients finally run out of treatment options. Often vision rehabilitation aids are a patient's last treatment option. In particular, for distance viewing therapeutic options are scarce for these patients. This thesis presents a study of the effectiveness of low vision spectacles for distance viewing, E-Scoop (**Chapter 3**). Despite many positive reviews from satisfied users, we found no clinically relevant evidence in terms of QoL, visual acuity, and contrast sensitivity, although analyses showed a small but nevertheless statistically significant positive effect on visual acuity. This lack of evidence from the standard QoL instruments appears to contradict the positive patient experiences. Arguably, this could be the result of some sort of placebo effect in the patient-reported experience. However, there was no clear effect on QoL improvements, while QoL was also sensitive for placebo effects. Perhaps, with hindsight, we should have included patient-reported experience such as treatment satisfaction as an outcome measure. In any case, E-Scoop is used by satisfied patients, so even if it is said that it has no effect on QoL, the patients 'reveal' that they think it helps. It could even be that the positive patient evaluations occur simply because the E-Scoop is the only treatment available for these patients. So the E-Scoop returns the 'locus of control' to the patient. What we observe here is that the patient perspective and the societal perspective deviate: the E-Scoop seems to be of value for the patients, but an increase in QoL cannot be specified in an objective way, which is a condition for reimbursement in societal health care insurance.

In conclusion, shared decision-making could benefit from an open dialogue between doctor and patient which makes health-related QoL, treatment satisfaction, and also general QoL a part of the treatment plan. This can have implications for macular degeneration treatment regimes, where a balance should be found in treating, monitoring, and adjustment.

Routine use of PROMs in clinical practice

We argued above that monitoring PROs is relevant in assessing patient treatmenr regimens. For instance, we introduced short questionnaires to help patients express their QoL in dialogue with the physician (**Chapters 5, 6, 7**). Classical test theory generates questionnaires that contain many comparable questions of the same

construct in order to improve reliability. Currently, advanced analyses following item response theory, such as Rasch, enable considerable reductions in the length of questionnaires. Brevity makes them more suitable for routine clinical use. One example of such a short questionnaire is the 9-item Catquest-9SF, used with respect to cataract surgery. Catquest-9SF has been translated into, and validated for, the Dutch situation. In addition, a challenge arose to construct a short questionnaire for ophthalmology in general. For this purpose, we used a vision-specific questionnaire, the NEI-VFQ-25, which was shortened to the 7-item VFQ-3oo7, of which the patient only had to answer three questions. VFQ-3oo7 is an example of a questionnaire suitable for use in clinical practice.

First, Catquest-9sf is used for monitoring the outcomes of cataract surgery in terms of visual functioning (**Chapters 5 and 6**). Usually, patients are asked to complete this questionnaire before and three months after surgery. The difference in the outcomes of these two measurements provides a potential indicator of surgical success in terms of visual functioning. So even if the surgery is successful in terms of visual acuity and adverse events, this does not always mean it also holds for visual functioning. In the case of cataract, the second eye is often affected within a year or even months after the surgery on the first eye. However, vision does not deteriorate further after cataract surgery in this respect. Hence, there is no need to administer the questionnaire for the patient three months after surgery.

A second approach focuses on the different disease courses that various eye diseases follow. For this reason, we created the VFQ-3oo7 instrument for various eye conditions (**Chapter 7**). This questionnaire enables patients to monitor their QoL over time and make it a part of the dialogue with the doctor during each clinical visit. As we know from mental healthcare, routine outcome monitoring would be suitable in accompanying this process. Patients are provided with information based upon their own personal health outcomes over time in consultation with their doctor. This way, a treatment plan can be adjusted based on complete information. Also, patients are supported in providing information, thus facilitating a better role for them in shared decision-making.

Both questionnaires are easy for patients to complete and easy for the doctor to interpret. This facilitates frequent usage without spending much time on filling out questionnaires. If repeatedly administered, both questionnaires can be an asset in assisting both patient and doctor in making treatment decisions. By routinely monitoring outcomes, a doctor can obtain an understanding of the patient's functioning over time and the effects of treatment. Moreover, this routine outcome monitoring can be linked to treatment objectives for specific patient groups, which allows both patient and doctor to see if these align with expectations for this patient group.

Routine outcome monitoring evidently helps to stimulate shared decision-making. Meanwhile, digital developments offer opportunities to support remote monitoring. How to give substance to these developments can be improved. Electronic patient files make it possible for patients to gain more insights into their health(care), as they can consult these files at any time and anywhere. This offers the potential for patients to achieve more control over the monitoring of their own health. Together with the growing attention to e-Health, which received a boost during the corona pandemic, the incorporation of PROs has the pontential to change healthcare. Patients can complete the questionnaires at home, and for routine medical check-ups they can attend an optometrist nearby. The latter could be effective for less complex and non-acute patients. Routine outcome monitoring can function as an alarm system that, if necessary, is triggered. This could result in fewer avoidable referrals to secondary care, thus facilitating the delivery of the right care in the right place at the right time, and also promoting the transition to remote medicine. The key aims are to prevent, relocate, and replace care, and to increase the role of the patient. The new VFQ-3oo7 fits perfectly into the role e-Health is acquiring in healthcare. In addition to VFQ-3oo7, using a small set of other short outcome measures would enable patients to monitor their health at home, thus making the distance between patient and healthcare shorter.

BSEs and WSEs in ophthalmic care

Another way the patient perspective could be a part of decision-making in ophthalmology is at the health policy level. When considering reimbursement, health economic evaluations can offer insights into the efficiency of a treatment. Efficiency is generally expressed as a ratio of costs to utilities. Utilities can be viewed as valuations for generic QoL, since health care policy is about making choices that affect all aspects of the healthcare system, not just ophthalmology. Some economic evaluations in ophthalmology have not fulfilled this aim of measuring QoL in a generic way. This is evident in analyses that have not included the compensating effects that the BSE has over the WSE. If we treat the latter, we should take into account that the increase in QoL is limited, as the BSE may have already compensated the WSE. This moderating effect is not included in most health economic analyses. Frequently, effects in the WSE are then valued as if they occured in the BSE. This results in over- and underestimation of the QoL addressed to the WSE. We found that patients' QoL mostly depended on the BSE, although 56% were being treated in the WSE. In this thesis, we created an HRQoL utility framework for patients with macular degeneration, which contains both the BSE and WSEs (**Chapter 4**). This framework could help to create a more tailored approach to the application of

treatment regimes. When a patient's BSE deteriorates, the focus will most likely be on treatment, as the biggest contributor to their QoL is at risk. However, patients who are affected solely in their WSE might prefer to wait, as they can delay treatment and thus possible side effects and other burdens associated with this treatment, while QoL is barely affected. As described above, patients with macular degeneration often have comorbidity and weigh this against other aspects of QoL, such as avoiding hospital visits. It is important to observe that this may work differently in younger patients who are confronted with eye problems which are the first health-related chronic illness they experience. Nevertheless, health choices are, regardless of a patient's age or health, the patient's business. The framework provided in this thesis enables such trade-offs to be made by patient, doctor, and society.

Hospital logistics and benchmarking

By including short QoL questionnaires in treatment protocols, this thesis supports the inclusion of the patient perspective in medical decision-making. The Catquest-9SF instrument was incorporated onto the online data platform Zorgladder, thus allowing healthcare providers to move towards outcome-oriented care. For the VFQ-3oo7 we experienced more difficulty implementing the instrument in ophthalmic care. The appropriate logistic framework was still not available at the hospital.

At another level, routine outcome measurement could be implemented to improve quality of care by using it for benchmarking. For hospitals, benchmarking could help optimise care pathways by highlighting where in the spectrum a hospital is, which hospitals are doing better, what they are doing, and how care could be adjusted in ways that match their performance. For example, when performing cataract surgery, different starting points in Catquest-9SF scores were found across five hospitals (**Chapter 6**). This resulted in different treatment outcomes. This may have depended on whether surgery was at an earlier stage or that other factors in this care path played a role. Another possibility could have been differing hospital populations. In addition, benchmarking is not without risk. The labelling of hospitals can result in a discussion concerning whether care in a particular hospital is at a sufficient level, with consequences for its funding. Moreover, the trend in healthcare is to make hospitals centres of expertise. It is encouraging that by following this approach hospitals can take healthcare to higher levels with superspecialists at their disposal. However, as mentioned earlier, from a patient perspective this will result in a higher investment of time for both patient and any companion. This contrasts with the opportunities presented above to save time. Ultimately, benchmarking makes visible how hospitals care for their patients differently and so shows them how they can improve. This thesis provides evidence to support benchmarking.

Appendix

Summary

&

Ophthalmology is the perfect place in healthcare to improve quality of life (QoL), as ophthalmic diseases are generally not life-threatening. This thesis investigated several ways of using QoL measures in ophthalmology to aid decision-making in clinical consultations, policy-making and reimbursement.

The first part of this thesis investigated the role of QoL in treatment evaluations. Two randomized controlled trials (RCT) aimed at determining the impact of interventions on disease-specific quality of life in patients with age-related macular degeneration (ARMD). In ARMD the central retinal function is affected, profoundly impairing the patient's ability to perform daily activities and their QoL. A current standard therapy for the exudative form of ARMD is a four-weekly intravitreal injection of bevacizumab. In **Chapter 2** the effect of different bevacizumab injection frequencies on quality of life in patients with neovascular ARMD is investigated in an RCT. Non-inferiority of the 6 and 8 weeks frequencies was demonstrated compared to the standard 4 weeks on vision-related and general QoL. These results are in line with previously published results of lower injections frequencies regarding visual acuity and central retinal thickness. A lower injection frequency may reduce burden, side effects, and treatment costs. In consideration of these results, 8 weeks frequency injections of intravitreal bevacizumab could be considered in patients with nARMD. Regardless of the choice of injection frequency, ARMD progresses for a number of patients resulting in serious vision loss. For these patients vision rehabilitation aids are often a patient's last treatment option. Particular for distance viewing, therapeutic options are scarce for these patients. A low vision spectacle, E-Scoop, which includes low-power prisms, 6% magnification, yellow tint, and antireflection coating, might aid in daily activities by improving distance viewing. In **Chapter 3** an RCT of the effect of E-Scoop on QoL is presented. No clinically relevant improvements of E-Scoop on QoL were found. E-Scoop showed effects that were statistically significant, although not clinically meaningful and within the typical variability, on visual measures. Combining separate features within a single spectacle lens did not result in a measurable effect in this investigation. As the burden of disease is shown to be much worse compared to an age-matched normal population sample, research in this patient group is encouraged.

Once the effectiveness of a treatment is evaluated positively, the treatment needs to be assessed for potential reimbursement. For this purpose, health economic evaluations are warranted in which generic QoL has a part. However, economic evaluations in ARMD were often hampered as generic instruments are flawed, as they are based on one eye only, mostly the better-seeing eye (BSE). Moreover, frequently chosen methods relied on patient values and/or disease specific measures, while economic evaluations preferred generic QoL measures based on

societal preferences. A possible alternative, in the form of the generic QoL utility instrument EQ-5D, has shown to be insensitive for differences in visual acuity. In the second part of this thesis, **Chapter 4** aims to provide societal utility values, using the generic SF-6D, for health states acknowledging both the better-seeing eye (BSE) and worse-seeing eye (WSE). A framework of societal utility values was provided creating visual acuity health states based on both the BSE and WSE for patients with nARMD. The range of the values was smaller than previous elicited utilities from a disease-specific measure. Besides, the utility values are placed on a more realistic position on the utility scale, and SF-6D utility values avoid the problem associated with the interpretation of disease-specific utility values.

A last application of QoL could be found in clinical practice. The third part of the thesis focused on the applicability of short QoL questionnaires in daily ophthalmic practice, enabling routine outcome monitoring (ROM). This is initially described in cataract and thereafter in ophthalmic diseases in general. The benefits in visual functioning of cataract surgery are quantified with the Catquest-9SF questionnaire, a unidimensional, reliable, valid and short patient-reported outcome measure. In **Chapter 5,** a formal Dutch translation was developed of the short patient-reported outcome measure Catquest-9SF, and the validity and reliability was tested. The Catquest-9SF proved to be a suitable measure of subjective visual functioning in the Dutch cataract population. The questionnaire was valid, reliable, unidimensional and responsive to changes after cataract surgery. Summary scores and percentiles are provided. In addition, norm scores may help in interpreting surgery outcomes and cut-off values may aid in identifying patients eligible for cataract surgery. This makes Catquest-9SF suitable for routine use in clinical practice. To ensure the Catquest-9SF is also valid compared to clinical visual and refractive measures in **Chapter 6** the Catquest-9SF is further investigated in a multicentre study. The main factors were performing the surgery in one or two eyes, ocular comorbidity and pre- and postoperative complications. This study emphasized the added value of the Catquest-9SF as a measure for visual function experienced by patients. Quality of vision improved more in patients who had surgery in both eyes and in patients who had fewer postoperative complications. We found a nonsignificant 'trend' that quality of vision was worse when ocular comorbidity was present. No significant effect of preoperative complications was observed. Our results emphasize the added value of the Catquest-9SF as a tool for visual function experienced by patients; the additional information can complement clinical parameters to improve patient-centered approaches in clinical practice. As short questionnaires helped evaluating treatment effect for cataract patients; **Chapter 7** investigated the shortening of a QoL questionnaire in ophthalmic diseases in general. The construction of a short

version of the general visual QoL questionnaire NEI-VFQ-39 resulted in the VFQ-3oo7. Of this 7-item questionnaire only 3 items need to be filled out to provide a representative score with minimal loss of information. This is established through routing online in a computerized system or just paper and pencil. A high level of generalizability was reached as a broad range of ophthalmic diseases were studied. The very short, but valid, VFQ-3oo7 can be applied to evaluate the patient's perceived vision-related health status in routine evaluations of treatments in regular care, with only little burden for patients.

In **Chapter 8** the main findings of this thesis were discussed. The general discussion focuses on the number of applications of health-related QoL measures in ophthalmology and the integration of these measures into ophthalmic care. Shared decision-making could benefit when an open dialogue between doctor and patient is created to make health-related QoL, treatment satisfaction and also general QoL a part of the treatment plan. Routine use of PROMs in clinical practice can help preventing, relocating and replacing care, as we start increasing the role of the patient. Using a small set of short outcome measures would enable patients to monitor their health in real time, right at home, making the distance between patient and healthcare smaller.

Samenvatting

Oogheelkunde is de uitgelezen plek in de gezondheidszorg om de kwaliteit van leven (KvL) te verbeteren, aangezien oogaandoeningen zelden levensbedreigend zijn. In dit proefschrift zijn verschillende manieren onderzocht om KvL-metingen in de oogheelkunde te gebruiken ter ondersteuning van de besluitvorming in de spreekkamer, het maken van beleid en vergoedingsvraagstukken.

Het eerste deel van dit proefschrift onderzocht de rol van KvL in de evaluaties van behandelingen. Twee gerandomiseerde gecontroleerde trials (RCT) hadden tot doel de impact van interventies op ziekte-specifieke kwaliteit van leven bij patiënten met leeftijdsgebonden maculadegeneratie (ARMD) te bepalen. Bij ARMD is de functie van het centrale netvlies aangetast, waardoor het vermogen van de patiënt om dagelijkse activiteiten uit te voeren en de KvL ernstig worden aangetast. Een huidige standaard therapie voor de exsudatieve vorm van ARMD is een vier wekelijkse intravitreale injectie met bevacizumab. In **hoofdstuk 2** wordt het effect van verschillende bevacizumab injectiefrequenties op de kwaliteit van leven bij patiënten met neovasculaire ARMD onderzocht. Non-inferioriteit van de 6 en 8 weken frequenties werd aangetoond ten opzichte van de standaard 4 weken op visus-gerelateerde en algemene KvL. Deze resultaten zijn in lijn met eerder gepubliceerde resultaten van lagere injectiefrequenties wat betreft gezichtsscherpte en centrale netvliesdikte. Een lagere injectiefrequentie kan de belasting, bijwerkingen en behandelingskosten verminderen. Gezien deze resultaten kan overwogen worden om intravitreale bevacizumab injecties met een frequentie van 8 weken toe te dienen bij patiënten met nARMD. Ongeacht de keuze van de injectiefrequentie, vordert ARMD bij een aantal patiënten, met ernstig verlies van gezichtsvermogen tot gevolg. Daarom zijn hulpmiddelen voor gezichtsrevalidatie vaak de laatste behandelingsoptie. Met name voor het kijken in de verte zijn de therapeutische opties voor deze patiënten schaars. Een low vision bril, E-Scoop met prisma's met lage sterkte, 6% vergroting, gele tint, en antireflectie coating, zou kunnen helpen bij dagelijkse activiteiten door het kijken op afstand te verbeteren. In **hoofdstuk 3** wordt een RCT gepresenteerd van het effect van E-Scoop op de kwaliteit van leven. In dit onderzoek werden geen klinisch relevante verbeteringen van E-Scoop op de KvL gevonden. E-Scoop toonde effecten die statistisch significant waren, hoewel niet klinisch betekenisvol en vielen binnen de typische variabiliteit van visuele metingen. Het combineren van afzonderlijke kenmerken binnen één enkel brillenglas resulteerde niet in een meetbaar effect. Aangezien is aangetoond dat de ziektelast veel ernstiger is in vergelijking met een steekproef van een normale populatie die overeenkomt qua leeftijd, wordt onderzoek in deze patiëntengroep aangemoedigd.

Zodra een behandeling positief is geëvalueerd, moeten deze worden

beoordeeld voor een mogelijke vergoeding vanuit het basispakket. Daartoe zijn gezondheidseconomische evaluaties nodig, waarin generieke KvL een rol speelt. Economische evaluaties bij ARMD werden echter bemoeilijkt omdat vaak slechts utiliteitswaarden voor één oog werden toegepast, meestal het bestziende oog (BSE). Bovendien waren de vaak gekozen methoden gebaseerd op patiëntwaarderingen en/of ziektespecifieke metingen, terwijl economische evaluaties de voorkeur gaven aan generieke KvL-metingen gebaseerd op maatschappelijke voorkeuren. De voor de hand liggende generieke KvL vragenlijst EQ-5D is ongevoelig gebleken voor verschillen in gezichtsscherpte. **Hoofdstuk 4**, het tweede deel van dit proefschrift, had als doel om daarom met behulp van de generieke SF-6D maatschappelijke utiliteitswaarden te geven voor gezondheidstoestanden die zowel het bestziende (BSE) als het slechtstziende oog (WSE) erkennen. Er werd een overzicht van maatschappelijke utiliteitswaarden opgesteld voor de gezondheidstoestand van gezichtsscherpte op basis van zowel de BSE als de WSE voor patiënten met nARMD. Het bereik van de waarden was kleiner dan eerder verkregen waarden op basis van een ziekte-specifieke maat. Bovendien zijn de utiliteitswaarden op een meer realistische plek op de utiliteitsschaal geplaatst, en vermijden SF-6D utiliteitswaarden het probleem dat geassocieerd wordt met de interpretatie van ziekte-specifieke utiliteitswaarden.

Een laatste toepassing van KvL kon gevonden worden in de spreekkamer. Het derde deel van het proefschrift richtte zich op de toepasbaarheid van korte KvL vragenlijsten in de dagelijkse oogheelkundige praktijk, waardoor routinematig monitoren mogelijk is. Dit wordt in eerste instantie beschreven aan de hand van specifiek cataract en daarna bij oogheelkundige aandoeningen in het algemeen. De resultaten in termen van visueel functioneren van cataractchirurgie kan worden gekwantificeerd met de Catquest-9SF vragenlijst, een unidimensionele, betrouwbare, valide en korte patiënt-gerapporteerde uitkomstmaat. In **hoofdstuk 5** werd een formele Nederlandse vertaling ontwikkeld van deze Catquest-9SF, en werd de validiteit en betrouwbaarheid getest. De Catquest-9SF bleek een geschikte maat te zijn voor subjectief visueel functioneren in de Nederlandse cataractpopulatie. De vragenlijst bleek valide, betrouwbaar, unidimensionaal en responsief voor veranderingen na cataractchirurgie. Overall scores en percentielen worden gepresenteerd. Bovendien kunnen de nu beschikbare normscores helpen bij het interpreteren van operatie-uitkomsten en kunnen afkapwaarden helpen bij het identificeren van patiënten die in aanmerking komen voor cataractchirurgie. Dit maakt Catquest-9SF geschikt voor routinematig gebruik in de klinische praktijk. Om er zeker van te zijn dat de Catquest-9SF ook geldig is in vergelijking met klinische visuele en refractieve maten wordt in **hoofdstuk 6** de Catquest-9SF

verder onderzocht in een studie in meerdere centra. De belangrijkste factoren die van invloed waren op de score waren het uitvoeren van de operatie in één of twee ogen, oculaire comorbiditeit en pre- en postoperatieve complicaties. Deze studie benadrukte de toegevoegde waarde van de Catquest-9SF als maat voor de door patiënten ervaren visuele functie. De kwaliteit van het gezichtsvermogen verbeterde meer bij patiënten die aan beide ogen werden geopereerd en die minder postoperatieve complicaties hadden. Wij vonden een niet-significante 'trend' dat de kwaliteit van het gezichtsvermogen slechter was wanneer er sprake was van oculaire comorbiditeit. Er werd geen significant effect van peroperatieve complicaties waargenomen. Onze resultaten benadrukken de toegevoegde waarde van de Catquest-9SF als een instrument voor de visuele functie die door patiënten wordt ervaren; de aanvullende informatie kan klinische parameters aanvullen om de patiëntgerichte aanpak in de klinische praktijk te verbeteren.

Omdat korte vragenlijsten hielpen bij het evalueren van het effect van de behandeling voor cataractpatiënten, onderzocht **hoofdstuk 7** het inkorten van een KvL vragenlijst bij oogziekten in het algemeen. De constructie van een korte versie van de algemene visuele KvL vragenlijst NEI-VFQ-39 resulteerde in de VFQ-3oo7. Van deze 7-item vragenlijst hoeven slechts 3 items te worden ingevuld om een representatieve score te krijgen met minimaal verlies van informatie. Dit wordt vastgesteld door de vragenlijst online in een geautomatiseerd systeem in te vullen of gewoon met pen en papier. Er werd een hoge mate van generaliseerbaarheid bereikt, aangezien een breed scala van oogheelkundige ziekten werd onderzocht. De zeer korte, maar valide, VFQ-3oo7 kan worden toegepast om de waargenomen visus-gerelateerde gezondheidsstatus van de patiënt te evalueren bij routinematige evaluaties van behandelingen in de reguliere zorg, met slechts weinig belasting voor de patiënt.

In **hoofdstuk 8** werden de belangrijkste bevindingen van dit proefschrift besproken. De algemene discussie richt zich op het aantal toepassingen van gezondheidsgerelateerde KvL maten in de oogheelkunde en de integratie van deze maten in de oogheelkundige zorg. Gedeelde besluitvorming kan baat hebben bij het creëren van een open dialoog tussen arts en patiënt om gezondheidsgerelateerde KvL, behandelingstevredenheid en ook algemene KvL een onderdeel te maken van het behandelplan. Routinematig gebruik van PROMs in de klinische praktijk kan helpen bij het voorkomen, verplaatsen en vervangen van zorg, omdat de rol van de patiënt kan worden vergroot. Het gebruik van een kleine set van korte uitkomstmaten geeft de mogelijkheid om gezondheid in real time te monitoren, direct thuis, waardoor de afstand tussen patiënt en gezondheidszorg kleiner wordt.

Dankwoord

&

Dit onderdeel van het proefschrift wordt door velen gezien als het meest dankbare gedeelte. Ik zou dan ook graag mijn dank uit willen spreken voor de mensen die mij linksom of rechtsom hebben geholpen bij de totstandkoming van deze thesis.
Allereerst wil ik me richten tot mijn promotor prof. dr. van Busschbach. Beste Jan, ik waardeer de onuitputtelijke mogelijkheden die ik kreeg om te gaan ontdekken, daarin mijn eigen pad te vinden en me, als ik het pad uit het zicht verloor, wel weer de wind in de zeilen te geven. Het tomeloze vertrouwen betekende voor mij dat ik inderdaad die vrijheid heb kunnen nemen, veel kansen heb aangegrepen en vele mooie mensen heb mogen ontmoeten. Congressen met jou waren voor mij dan ook een ware uitputtingsslag, aangezien er altijd wel ergens iemand was om te ontmoeten en de beste contacten leg je 's avonds. Terplekke gezond eten en hardlopen compenseerde dit dat wel weer.
Dr. Timman, beste Reinier. Hoewel jij niet vanaf het begin betrokken was bij mijn onderzoekscarriere heb jij je met alle artikelen van deze thesis ingelaten. Als zijnde een ware eindsprint hebben we afgelopen jaren zeer intensief samengewerkt. Je was een zeer prettig gezelschap en dit motiveerde tot mooie innovatieve artikelen en analyses. Het meest creatieve uit jouw pet was toch wel het toesturen van een vooroorlogse reisgids van Manchester aan een zeer gewaardeerde meelezer uit Manchester. Overigens was dit volkomen ethisch verantwoord aangezien dit pas achteraf werd bedacht.
Collega's en coauteurs van het Oogziekenhuis en Ergra Low Vision, bedankt. Een aantal zaken blijven me erg bij. Netty, als drijvende kracht achter de vliegende start van dit proefschrift. Het gebaar van Sankha, die de mens achter de data zichtbaar maakte, door een kerk af te huren om resultaten terug te koppelen aan patiënten uit de studie. Kamergenoten René en Koen, waar ik me welkom voelde. En als laatst maar zeker niet het minst, het vele datawerk van Stijn.
Collega's van andere onderzoeksvelden wil ik bedanken. Met name Elly voor de drie maanden die ik bij jou mocht werken op Woudestein. Menno, ik denk met nostalgie terug aan Maastricht en Glasgow.
Promoveren bij de MPP is bijzonder door de combinatie van onderzoekers en docenten op één afdeling. Congressen van ISPOR, LolaHESG en Euroqol werden afgestruind. Ontspanning vonden we in het verven van keukenkastjes (Annelieke & Marieke) tot het hardlopen van de marathon en de bruggenloop (o.a. Jacky & Zhihao), van ons papieren gps-systeem met punaise-hoofdjes op een wereldkaart (Hester & Sohal) tot de Scorito-stand op maandag op het whiteboard (Steef & Lothar). Hester, ik drink mijn thee nog steeds liever niet uit een rietje, bedankt voor de andere inspiraties. Casimir, Lothar, Steef en Sohal, bedankt voor het slap geouwehoer tussendoor. In het bijzonder

bedank ik Steef voor de vele open gesprekken, letterlijk tussen werk en privé. Mijn vele MPP onderwijs collega's hebben mij warm en enthousiast gemaakt én gehouden voor onderwijs. In het bijzonder wil ik Jan Willem bedanken voor het hardlopen, sporten bij Hollander en het inzicht dat body pump echt niks voor mij is. Annemerle, bedankt voor de natuurlijke samenwerking. Reinoud, fijn dat er óók mensen zijn waar het tempo wat lager ligt, ik hoop alleen niet dat je ook zo fietst. Elise, fijn dat je er bent, helaas wat ver weg. Wietske, er is inderdaad eentje bijgekomen, maar het doet me veel verdriet dat jij bent weggegaan.
Lieve vrienden. Ik heb altijd geweten dat ontspanning essentieel is voor een succesvol leven. Bedankt dat jullie vrijwel nooit over dit proefschrift begonnen. Frank en Thijs, het wekelijks hardlopen bood een fysieke en verbale uitlaatklep gedurende de laatste jaren. Marjolein, Freek, Gilian en Daniel, wat waren de weekendjes weg een welkome afwisseling. Evy, bedankt voor onze vriendschap. Johannes, jij tilde elk gesprek naar een ander niveau.
Mijn (extended) schoonfamilie, Jeannine, Paul, Huib, Paul, Daphne, Willemien, Irene, Sterre, Martin en Lisette, bedankt voor de sportieve en gezellige activiteiten en reisjes.
In bijzonder bedank ik mijn familie, lieve pap en mam, Mark, Gitty, Boris, Jelle en Lisanne. Bedankt voor jullie oprechte interesse, tomeloze geduld en steun.
Tot slot, lieve Marise. Mijn sparringpartner, mijn inspiratie. Dit proefschrift had er zonder jou niet gelegen. Ik denk dat je elke letter van dit proefschrift op enig moment wel hebt gelezen. Maar met name buiten werk hebben wij het goed. Wat heerlijk om met jou te leven en te dromen over wat is en wat gaat komen, samen met Loena en Jinte.

&

About the Author

&

Martijn Stefan Visser was born in Hilversum, the Netherlands, on the 10th of July 1985. He attended secondary school at the Stedelijk Gymnasium (VWO) in Leiden, where he graduated in 2003. He studied Psychology at Leiden University between 2003 and 2008. In 2007 he obtained his Bachelor's degree. He continued with a Master of Science in Social and Organisational Psychology and obtained his Master degrees in 2008. In March 2009 Martijn started working as a junior researcher at the Erasmus MC on the department Medical Psychology and Psychotherapy. In 2011 the research described in this thesis started. During his research, he started a Master of Science in Epidemiology at the Netherlands Institute of Health Sciences (NIHES). In 2013, he obtained his Master's degree in Health Sciences. Subsequently, he was involved in the teaching of communication skills for medical students and he was active in promoting professional development among these students, e.g. the coordinating of the coach program. Currently, he is working on the to be developed Medicine curriculum stimulating students' professional identity development.

List of Publications

&

Publications

This thesis

Visser MS, Amarakoon S, Missotten T, Timman R & Busschbach JJV. Low injection frequencies of bevacizumab are non-inferior for quality of life in neovascular age-related macular degeneration. Quality of Life Research. 2020; Jul 14.doi: 10.1007/s11136-020-02580-9

Visser MS, Timman R, Kampen-Smalbrugge J, Buis K, Polling JR & Busschbach JJV. The effect of a loupe lens on quality of life in patients with macular degeneration, an randomized controlled trial. Optometry and Vision Research. 2020; Sep

Visser MS, Amarakoon S, Missotten T, Timman R & Busschbach JJV. SF-6D utility values for the better- and worse-seeing eye for health states based on the Snellen equivalent in patients with Age-Related Macular Degeneration. PLoS One. 2017;12(2):e0169816.

Visser MS, Dieleman M, Klijn S, Timman R, Lundström M, Busschbach JJV & Reus NJ. Validation, test-retest reliability and norm scores for the Dutch Catquest-9SF. Acta Ophthalm. 2016;Oct 24. doi: 10.1111/aos.13287.

Stolk-Vos AC, **Visser MS**, Klijn S, Timman R, Lansink P, Nuijts R, Tjia K, Zijlmans B, Kranenburg LW, Busschbach JV, Reus NJ.Effects of clinical parameters on patient-reported outcome in cataract patients: a multicentre study. Acta Ophthalmol. 2018 Mar 25. doi: 10.1111/aos.13747.

Visser MS, Timman R, Nijmeijer KJ, Lemij HG, Kilic E & Busschbach JJV. The development of the Visual Function Questionnaire Short Form 3 out of 7 (VFQ-3oo7) fitted for the Routine use of Patient Reported Outcome Measures. ActaOphthalmologica. 2020 Mar 18. doi: 10.1111/aos.14378.

Remaining publications

Kouwenberg CAE, Kranenburg LW, **Visser MS**, Busschbach JJ, Mureau MAM. The validity of the EQ-5D-5L in measuring quality of life benefits of breast reconstruction. J Plast Reconstr Aesthet Surg. 2019 Jan;72(1):52-61.

Redeker S, Oppe M, **Visser MS,** Busschbach JJV, Weimar W, Massey EK, Ismail S. Cost-effectiveness of a home-based group educational programme on renal replacement therapies: a Study Protocol. BMJ Open. 2019

Wong ELY, Shah K, Cheung AWL, Wong AYK, **Visser MS**, Stolk E. Evaluation of Split Version and Feedback Module on the Improvement of Time Trade-Off Data. Value in Health. 2018 Jun;21(6):732-741.

Visser MS, Zonneveld LNL, van't Spijker A, Hunink MG & Busschbach JJV. The cost-effectiveness of cognitive-behavioral group training for patients with unexplained physical symptoms. Value in Health. 2015;18(5), 570-577.

Vogelaar L, van 't Spijker A, Vogelaar T, Busschbach JJV, **Visser MS**, Kuipers EJ & van der Woude CJ Solution focused therapy: A promising new tool in the management of fatigue in Crohn's disease patients. Journal of Crohn s and Colitis. 2011 5(6):585-91

Publiced posters

Visser MS, Amarakoon S, Missotten T & Busschbach JJV PSS27 We Treat Eyes, Not People: The Systematic Overestimations of Utility in Age-Related Macular Degeneration Models. Value in Health. November 2011. doi: 10.1016/j.jval.2011.08.1492

Visser MS, Amarakoon S, Missotten T & Busschbach JJV PSS19 Cost-effectiveness of age-related macular degeneration: a model. Value in Health. November 2010. doi: 10.1016/S1098-3015(11)72648-1

Visser MS PSS26 Expected value of partial perfect information in a markov model of infliximab and etanercept in the treatment of moderate to severe plaque type psoriasis. Value in Health 12(7) · October 2009. doi: 10.1016/S1098-3015(10)75263-3

Other projects

Expertraad cataract van Stichting Zorgladder. 2017. Handboek Cataract. https://zorgladder.nl

Translation of RetDQol and RetTSQ 2017 ism health psychology research Ltd

Van Vliet EJ , Sol JCA, **Visser MS**, Lemij HG, Busschbach JJV. An estimation of the cost-effectiveness of a large-scale mobile eye screening: An economic evaluation. 2012. Rapport Oogziekenhuis Rotterdam

&

PhD portfolio

&

Summary of PhD portfolio: training & teaching

Name PhD student	Martijn Stefan Visser
Erasmus MC Department	Psychiatry, section Medical Psychology & Psychotherapy
Research School	Health Sciences
PhD period	2009-2021
Promotor	Prof.dr. J.J. van Busschbach
Copromoter	R. Timman

1. PhD training	Year	Workload (days/ hours / ECTS)
General courses		
• BROK course (basis cursus regelgeving Klinisch Onderzoek)	2009	30 hours
• English Language, SC01	2011	1.4 ECTS
• Introduction to Medical Writing, SC02	2011	1.1 ECTS
• Working with SPSS for Windows, SC04	2011	0.15 ECTS
• BROK course update	2015, 2019	3 hours

Research master NIHES: MSc in Health Science, specialization Clinical Epidemiology

Erasmus Summer Programme 2011		
• Principles of Research in Medicine, ESP01	2011	0.7 ECTS
• Methods of Public Health Research, ESP11	2011	0.7 ECTS
• Introduction to Global Public Health, ESP41	2011	0.7 ECTS
• Methods of Health Services Research, ESP42	2011	0.7 ECTS

• Primary and Secondary Prevention Research, ESP45	2011	0.7 ECTS
• Social Epidemiology, ESP61	2011	0.7 ECTS

Core Curriculum

• Study Design	2011	4.3 ECTS
• Biostatistical, CC01 Methods I: Basic Principles, CC02	2011	5.7 ECTS
• Clinical Epidemiology, CE02	2012	5.7 ECTS
• Methodologic, Topics in Epidemiologic Research, EP02	2012	1.4 ECTS
• Biostatistical Methods II: Popular Regression Models, EP03	2011	4.3 ECTS
• Oral research presentation	2011	-
• Site Visit to Municipal Health Service Rotterdam, PU03	2011	0.3 ECTS
• Integration module, PU04	2013	0.3 ECTS
• Research proposal	2013	2.5 ECTS

Advanced Short Courses

• Repeated Measurements in Clinical Studies, CE08	2012	0.9 ECTS
• Bayesian Statistics, CE09	2011	1.1 ECTS
• Missing Values in Clinical Research, EP16	2012	0.7 ECTS
• Courses for the Quantitative Researcher, EP17	2012	1.4 ECTS
• Quality of Life Measurement, HS11	2009	0.9 ECTS
• Preventing Failed Psychological Intervention Research, EXT43	2012	1.4 ECTS

Conferences

• 1st Low lands health economic study group (Lola HESG),Berg en Terblijt	2009	2 days
• ISPOR Europe on health economics outcomes research, Paris, France	2009	4 days
• Eiland Dagen – dermatology conference, Schiermonnikoog	2009	3 days
• 2nd Lola HESG,Egmond aan Zee	2010	2 days
• ISPOR Europe, Budapest, Hungary	2010	4 days
• 3rd Lola HESG,Utrecht	2011	2 days
• ISPOR Europe, Madrid, Spain	2011	4 days

Seminars, symposia, courses and workshops

• Course and Symposium 'The appraisal process, work in progress' - NvTAG/CVZ, Diemen-Zuid	2009	16 hours
• Symposium Cost-effective interventions in health-care	2009	3 hours
• Course cost-effectiveness modeling, Maastricht	2009	6 ECTS
• Course Advanced Modelling for health economic evaluation, Glasgow,	2009	3 days
• ZonMw/NVTAG bijeenkomst HTA-methodologie, The Hague	2010	3 hours
• Bayesian Statistics course, Glasgow	2010	3 days

• Outcomes Beyond the QALY Symposium', Rotterdam	2012	3 hours
• 31st EQ Plenary Meeting, Stockholm, Sweden	2014	2 days

Other

• Course Morello Basistraining webredactie, Rotterdam	2009	12 hours

Teaching

• Teaching medical students communication skills	2009 →	1800+ hours
• Patient preference assessment in health course (iBMG, EUR)	2013-2014	-
• Coaching bachelor students Medicine	2014 →	300+ hours
• Coordinator The bad news talk	2015-2018 2020 →	-
• Coordinator The disciplinary case	2015 →	-
• Coordinator Coaching	2016-2020	-
• Chair of workgroup Professional Identity Development, ErasmusArts2030	2019 →	-

Conferences, seminars, symposia, courses and workshops

• *Culturele diversiteit. Grethe van Geffen (Seba).* Doel: inzicht krijgen in cultuurverschillen en werking van relevante mechanismen. Handvatten krijgen voor het werken met gemengde groepen en voor het in gesprek te gaan over dit onderwerp. Bevorderen van eigen culturele competentie.	2013	-

• *Intervisie Leren, Mirabelle Schaub-de Jong*. Doel: kennis opdoen van de meerwaarde, voorwaarden, uitgangspunten, en valkuilen van intervisie, en docentcompetenties voor het begeleiden. Leren toepassen logische denkniveaus Bateson en coachingsmodel 1 op 1 mentorgesprek.	2013	-
• *Coach training, Lex Linsen, Emely Spierenburg, Pleun Hermsen en Noor Wolff*. Doel: kennis, inzicht en vaardigheden verwerven die nodig zijn voor het coachen van studenten. Deze training bestond voor een groot deel uit het voeren coachgesprekken met een echte student over zijn/haar coachvraagstuk tijdens rollenspellen.	2015	-
• BKO (basis kwalificatie onderwijs)	-	-
• Congres NVMO (Nederlandse Vereniging Medisch Onderwijs)	-	-