Title
Development of a vision-targeted health-related quality of life item measure

Journal
Quality of Life Research, 22(9)

ISSN
0962-9343

Authors
Paz, Sylvia H
Slotkin, Jerry
McKean-Cowdin, Roberta
et al.

Publication Date
2013-11-01

DOI
10.1007/s11136-013-0365-1

Peer reviewed

Qual Life Res (2013) 22:2477–2487
DOI 10.1007/s11136-013-0365-1

Development of a vision-targeted health-related quality of life item measure

Sylvia H. Paz · Jerry Slotkin · Roberta McKean-Cowdin · Paul Lee · Cynthia Owsley · Susan Vitale · Rohit Varma · Richard Gershon · Ron D. Hays

Accepted: 2 February 2013 / Published online: 9 March 2013

Abstract
Purpose To develop a vision-targeted health-related quality of life (HRQOL) measure for the NIH Toolbox for the Assessment of Neurological and Behavioral Function.
Methods We conducted a review of existing vision-targeted HRQOL surveys and identified color vision, low luminance vision, distance vision, general vision, near vision, ocular symptoms, psychosocial well-being, and role performance domains. Items in existing survey instruments were sorted into these domains. We selected non-redundant items and revised them to improve clarity and to limit the number of different response options. We conducted 10 cognitive interviews to evaluate the items. Finally, we revised the items and administered them to 819 individuals to calibrate the items and estimate the measure's reliability and validity.
Results The field test provided support for the 53-item vision-targeted HRQOL measure encompassing 6 domains: color vision, distance vision, near vision, ocular symptoms, psychosocial well-being, and role performance. The domain scores had high levels of reliability (coefficient alphas ranged from 0.848 to 0.940). Validity was supported by high correlations between National Eye Institute Visual Function Questionnaire scales and the new-vision-targeted scales (highest values were 0.771 between psychosocial well-being and mental health, and 0.729 between role performance and role difficulties), and by lower mean scores in those groups self-reporting eye disease (F statistic with $p < 0.01$ for all comparisons except cataract with ocular symptoms, psychosocial well-being, and role performance scales).
Conclusions This vision-targeted HRQOL measure provides a basis for comprehensive assessment of the impact of eye diseases and treatments on daily functioning and well-being in adults.

Keywords Vision-related quality of life · NIH Toolbox · Instrument development · Instrument psychometric evaluation and calibration

Electronic supplementary material The online version of this article (doi:10.1007/s11136-013-0365-1) contains supplementary material, which is available to authorized users.

S. H. Paz (✉) · R. D. Hays
Division of General Internal Medicine and Health Services Research, Department of Medicine, UCLA School of Medicine, 911 Broxton Avenue, Los Angeles, CA 90095-1736, USA
e-mail: shpaz@ucla.edu

J. Slotkin · R. Gershon
Department of Medical Social Sciences, School of Medicine, Northwestern University Feinberg, 625 N Michigan Ave., Suite 2700, Chicago, IL 60611, USA

R. McKean-Cowdin · R. Varma
Department of Ophthalmology, Doheny Eye Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

P. Lee
Department of Ophthalmology, Duke University, Durham, NC 27710, USA

C. Owsley
Department of Ophthalmology, School of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA

S. Vitale
Division of Epidemiology and Clinical Applications, National Eye Institute, National Institutes of Health, Bethesda, MD 20892, USA

Introduction

Vision impairment has important day-to-day impact on functioning and well-being, the major aspects of health-related quality of life (HRQOL). Valid assessment of HRQOL requires direct reports from patients rather than physicians or other parties. The importance of patient-reported outcomes is emphasized in several initiatives in the NIH Roadmap for Medical Research [1–3]. One of the initiatives involving patient-reported outcomes, the NIH Toolbox for the Assessment of Neurological and Behavioral Function (NIH Toolbox), was created to develop tools to measure motor, cognitive, sensory, and emotional function [4]. NIH Toolbox tools were designed to be short yet comprehensive measures useful for researchers engaged in epidemiological studies and other investigations, including potential comparative effectiveness analyses.

Existing vision-targeted HRQOL measures were developed for specific applications. For example, the National Eye Institute Visual Function Questionnaire assesses 12 domains targeted at individuals with eye diseases using 25 items. Similarly, the National Eye Institute Refractive Error Quality of Life Instrument assesses 13 domains targeted at refractive error using 42 items. The existing vision-targeted HRQOL measures were created prior to the movement to utilize item response theory to develop the most efficient and informative survey instruments [5]. This paper describes the development and psychometric evaluation and calibration of a vision-targeted HRQOL measure for the NIH Toolbox. We built on existing measures and used item response theory methods to provide a thorough evaluation of the psychometric properties of the new measure.

Methods

We began with a literature review of existing vision-targeted HRQOL surveys, including the following: (1) Graves' Ophthalmopathy Quality of Life Questionnaire [6–8], (2) Impact of Visual Impairment [9–13], (3) Low Vision Quality of Life Questionnaire [14, 15], (4) National Eye Institute Refractive Quality of Life-42 [15, 16], (5) National Eye Institute Visual Function Questionnaire-25 [17–19], (6) Quality of Life and Vision Function Questionnaire [20, 21], (7) Refractive Status and Vision Profile [22, 23], (8) Visual Function-14 [24, 25], (9) Impact of Dry Eye on Everyday Living [26–28], (10) Activities of Daily Vision Scale [29, 30], (11) Visual Activities Questionnaire [31, 32], and (12) Low Luminance Questionnaire [33].

According to FDA recommendations, a conceptual framework is helpful in guiding the development of a patient-reported outcomes instrument [34]. Following these recommendations, we first identified the aspects of vision-targeted HRQOL to be measured. Each of these aspects or "domains" is measured by a set of items. The multiple items assessing a domain are combined to create a domain score. In our study, we identified and included the following domains: color vision, low luminance vision, distance vision, general vision, near vision, ocular symptoms, psychosocial well-being, and role performance. We sorted the items into these domains.

Winnowing and revising items

Winnowing refers to reducing the item pool to the smallest number of items, making sure that the domain of interest will still be adequately measured. DeWalt et al. [35] stated that winnowing helps to accurately define and identify those item characteristics relevant to each domain. Items were deleted if they were considered to duplicate the information captured, or if they were deemed to be potentially confusing to respondents. We also reviewed items within each domain to be certain that all relevant aspects were captured by the items. We sought to keep a set of non-redundant items that was consistent with the domain, universally understood, and relevant to a wide range of the adult population.

Items were revised so they would be understood by the largest possible number of people. In addition, consistency was sought to the extent possible in wording, recall interval, and response options. After extensive discussion and a thorough literature review, the group decided that a 7-day recall period was optimal and would facilitate the validity of self-reports [36]. We opted to use five response options plus a tailored "not applicable" option when appropriate. The five response options to rate difficulty were: (1) *Not difficult at all*, (2) *A little difficult*, (3) *Somewhat difficult*, (4) *Very difficult*, and (5) *Unable to do because of eyesight*.

The not applicable response choice was expressed as "Unable to do for another reason (not related to vision)" or as "do not do [the activity] for another reason (not related to vision)." For example, for an item asking about difficulty driving at night ("Because of your eyesight, how difficult is it for you to drive due to glare from oncoming headlights?"), the "not applicable" response was "Unable to drive at night because of another reason," because other reasons besides eyesight might result in an individual not driving.

Response options to capture respondent-perceived quality were (1) *Poor*, (2) *Fair*, (3) *Good*, (4) *Very Good*, and (5) *Excellent*.

Response options to rate intensity of a problem were (1) *No problem at all*, (2) *A little bit of a problem*, (3) *Somewhat of a problem*, (4) *Very much of a problem*, and (5) *Unable to see*.

The "unable to see" option was used only in those items in which it was deemed necessary: For example, "How much of a problem do you have with seeing things at a distance at nighttime?"

Response options for the psychosocial well-being items were tailored specifically to what the item was asking, with standard phrasings of "Did not feel … at all," "Felt a little bit …," "Felt somewhat …," and "Felt … a lot." For example, an item that asked about feeling frustrated or upset had the following response options: (1) *Did not feel frustrated or upset at all*, (2) *Felt a little bit frustrated or upset*, (3) *Felt somewhat frustrated or upset*, and (4) *Felt frustrated or upset a lot.*

Response options for the role performance items were (1) *Not limited at all*, (2) *A little bit limited*, (3) *Somewhat limited*, (4) *A lot limited*, (5) *Unable to do because of eyesight*, and (6) *Unable or limited because of another reason*. The wording of the "unable to do because of another reason" choice varied according to what the item was asking. For example, for an item asking about limitations in daily activities ("Because of your eyesight, how limited are you in doing your daily activities?"), the "unable because of another reason" response was "Unable or limited in daily activities because of another reason," because other reasons besides eyesight might result in an individual being limited or unable to do daily activities.

To make items more universally understood, slight modifications to item stems and/or response options from their original form were made in those cases where we believed simplification was necessary. Items that were too long or confusing were shortened.

Cognitive interviews

Cognitive interviews were conducted to obtain information on all survey items and response options. Ten participants were recruited from the waiting rooms of the ophthalmology clinics at the Doheny Eye Institute, University of Southern California. The sample included 4 males and 6 females, and the age range was from 28 to 82 years. One participant had only completed middle school education, four had a high school education, two had college education, and three had graduate school education. Four participants had clinical appointments for eye disease. A trained interviewer conducted six of the cognitive interviews, and the remaining four were self-administered with oral probing by the interviewer, and a field for comments, at the end of each item.

Participants were asked to describe in their own words what each item was asking, and in some cases to compare different ways of asking the same question. These interviews were helpful because subjects could directly provide input and suggestions on every item considered for inclusion. As a result of these interviews, many items were further simplified, some words were deleted, some items were reworded, and some parts of the stems were added to response options for clarity. The revised survey was then used for field testing.

Field test data collection and sampling

Subjects were recruited through Greenfield (now Toluna), an online panel company [37]. Beyond a monetary reward, panel members experience a community approach through the use of social media technology—that is, interacting with other panelists and contributing toward research and debates across a wide range of topics. Subjects were recruited using multiple media sources and a wide array of techniques to attract different subjects. Different social media such as banners, e-mails, text messages, and word of mouth were some of the techniques initially used. Prospective subjects first completed an initial registration form with an acceptance of terms and conditions. Data were collected between February and March 2010 ($N = 819$).

Statistical analysis

We hypothesized eight multi-item domains: Color Vision, Low Luminance Vision, Distance Vision, General Vision, Near Vision, Ocular Symptoms, Psychosocial Well-being, and Role Performance. Items had between four and six response options. However, due to sparse distributions, some of the categories were collapsed so that no cell would have <5 % of the data (41 participants). We estimated item means, standard deviations, item-rest correlation (corrected for overlap), and coefficient alpha for each multi-item scale by using Stata 9 and MPlus [38, 39].

An eight-factor categorical confirmatory factor analysis model representing the original hypothesized structure and a modified six-factor model were fit using MPlus [39]. These models estimated correlations among domains (correlated factors) and did not estimate correlated errors among items. Model fit was assessed by the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), as well as the root-mean-square error of approximation (RMSEA). We considered a model to be a good model fit based on the following criteria: CFI > 0.95, TLI > 0.95, and RMSEA < 0.06 [38, 40, 41].

MULTILOG, a computer program that uses logistic models and is designed for IRT analysis for items with multiple response options, was used to estimate a two-parameter (2-PL) graded response model using marginal maximum likelihood estimation [41]. MULTILOG yields estimates that are logistic (the $D = 1.7$ scaling factor is not used). We specified 50 cycles for estimation. The graded response model has one slope parameter and $j - 1$ threshold parameters (j is the number of response options) for every item. Therefore, the model yields one slope and one threshold parameter for dichotomous items and one slope and two threshold parameters for those items

Table 1 Field test sample characteristics ($N = 819$)

Characteristic	Sample	National US (%)[a]
Male gender	50 %	49
Age, years (mean and range)	53.9 (18–85)	
≥18 and ≤30	(20 %)	41
>30 and ≤50	(20 %)	27
>50 and ≤65	187 (29 %)	19
>65 and ≤75	151 (18 %)	7
>75 and ≤85	150 (18 %)	4
Race/ethnicity		
White	724 (88 %)	72
Black or African American	48 (6 %)	13
Spanish, Hispanic or Latino	43 (5 %)	16
Asian	31 (4 %)	5
American Indian or Alaska Native	18 (2 %)	1
Native Hawaiian or other Pacific Islander	4 (1 %)	0
Other	19 (2 %)	6
Education		
<High school	161 (20 %)	13
High school or equivalent	274 (33 %)	30
One or more years >high School	177 (22 %)	29
Bachelor's degree	138 (17 %)	18
Graduate degree	69 (8 %)	10
Health conditions[b]		
All vision impairment (blindness and low vision)		3
Myopia		24
Hyperopia		10
Wear glasses or contacts for distance	552 (67 %)	NA
Wear glasses or bifocals for near vision	501 (61 %)	NA
Cataracts	238 (29 %)	17
Glaucoma	60 (7 %)	2
Macular degeneration	61 (8 %)	1
Diabetes	130 (16 %)	8
Diabetic retinopathy (% of diabetics)	22 (17 %)	5
Arthritis	300 (37 %)	26[c]
Stroke or brain hemorrhage	34 (4 %)	2[c]
High blood pressure	340 (42 %)	29[c]
Angina	54 (7 %)	4[c]
Heart attack	72 (9 %)	4[c]
Heart failure or enlarged heart	38 (5 %)	NA
Asthma	105 (13 %)	14[c]
Skin cancer	71 (9 %)	
Other cancer	89 (12 %)	
Back problems including disk or spine	247 (30 %)	
Deafness or trouble hearing	139 (17 %)	15[e]
Depression (often or always)	82 (10 %)	10[d]

[a] Census 2010 (http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf; http://www.census.gov/hhes/socdemo/education/data/cps/2011/tables.html). Ethnicity percentages include only who self-reported one race (total 97.1 % of the total population) without counting the Latino ethnicity. Diabetes (http://www.diabetes.org/diabetes-basics/diabetes-statistics/)

[b] Eye conditions based on total 2010 US Census populations and number of cases by vision problem age ≥40, except for age-related macular degeneration, age 50 and older. http://www.nei.nih.gov/eyedata/adultvision_usa.asp

[c] http://www.cdc.gov/brfss/

[d] http://www.cdc.gov/features/dsdepression/

[e] http://www.nidcd.nih.gov/health/statistics/Pages/quick.aspx

with three response choices. The slope, or discrimination parameter, provides information about how well the item discriminates among adjoining trait levels. Higher values indicate that items are better able to discriminate between adjacent categories of trait level. The threshold parameter represents the point along the latent trait at which a subject has a 50 % chance of responding in that category versus the other categories.

Table 2 Item means (SD), response options, item-rest correlations[a], and coefficient alphas, by vision-targeted HRQOL measure scale[b]

Scale—items	Mean (SD)	Response options	Item-rest correlation	Coefficient alpha
Color vision				0.8924
A.1 Match colors of clothes	1.86 (0.344)	2	0.7516	
A.2 Recognize colors	1.86 (0.345)	2	0.8292	
A.3 Distinguish between colors	1.84 (0.369)	2	0.7865	
Distance vision				0.9398
B.1 See in bright sunlight	3.12 (0.915)	4	0.5995	
B.2 See in fluorescent lighting	2.63 (0.636)	3	0.6123	
B.3 See at night	3.02 (0.921)	4	0.7555	
B.4 See street signs at night when in a car	3.01 (0.987)	4	0.7596	
B.5 See when going from lighted area into dark place	3.03 (0.923)	4	0.7145	
C.1 Go up or down steps, stairs, or curbs	2.70 (0.582)	3	0.6281	
C.2 Get around outdoors	1.87 (0.336)	2	0.6089	
C.3 See moving objects	1.84 (0.362)	2	0.6308	
C.4 See street signs	2.51 (0.666)	3	0.6846	
C.5 See television	2.71 (0.577)	3	0.6419	
C.6 Drive at night	3.06 (0.951)	4	0.7758	
C.7 Drive due to glare from oncoming headlights	2.78 (1.005)	4	0.7345	
C.8 Recognize people from across a room	2.67 (0.607)	3	0.6448	
D.1 Rate eyesight for distance	3.01 (1.176)	5	0.6239	
D.2 See things at a distance during the day	2.49 (0.695)	3	0.7020	
D.3 See things at a distance at nighttime	2.84 (0.966)	4	0.7759	
Near vision				0.9241
B.6 Read in dim light	2.72 (1.013)	4	0.6021	
D.4 Rate eyesight for close up	3.21 (1.214)	5	0.6157	
D.5 See things close up during day	2.59 (0.649)	3	0.7092	
D.6 See things close up at nighttime	3.23 (0.893)	4	0.7334	
E.1 Read newspaper or book	3.33 (0.894)	4	0.7688	
E.2 Find item on crowded shelf	2.67 (0.584)	3	0.6682	
E.3 See sign with other signs around	2.60 (0.617)	3	0.6446	
E.4 Read labels or instructions	3.22 (0.919)	4	0.7564	
E.5 Use computer	2.65 (0.592)	3	0.7109	
E.6 Operate household appliance	1.90 (0.304)	2	0.5970	
E.7 Write	1.84 (0.370)	2	0.6447	
E.8 Do work or hobbies that require to see up close	2.59 (0.637)	3	0.7337	
E.9 Recognize people when nearby	1.92 (0.272)	2	0.4953	
Ocular symptoms				0.8479
F.1 Glare from bright lights	2.37 (0.701)	3	0.6028	
F.2 Burning or stinging eyes	2.59 (0.639)	3	0.6484	
F.3 Redness in eyes	2.70 (0.562)	3	0.5807	
F.4 Headaches because of vision	1.79 (0.408)	2	0.5066	
F.5 Eyes being sensitive to light	2.44 (0.698)	3	0.6635	
F.6 Blurry vision	2.56 (0.647)	3	0.6364	
F.7 Itching in or around eyes	2.53 (0.637)	3	0.6041	
Psycho/social				0.9222
G.1 Frustrated or upset	2.58 (0.644)	3	0.7122	
G.2 Annoyed or angry	2.68 (0.576)	3	0.7617	
G.3 Felt like burden on others	1.88 (0.325)	2	0.7644	
G.4 Felt sad or depressed	2.79 (0.523)	3	0.8093	

Table 2 continued

Scale—items	Mean (SD)	Response options	Item-rest correlation	Coefficient alpha
G.5 Felt socially isolated	1.90 (0.300)	2	0.7894	
G.6 Concerned about safety at home	1.89 (0.307)	2	0.7101	
G.7 Worried	2.59 (0.625)	3	0.6525	
G.8 Rely on others	1.84 (0.367)	2	0.6137	
G.9 Visits with friends or family	1.92 (0.265)	2	0.6717	
Role performance				0.8948
H.1 Daily activities	1.84 (0.368)	2	0.7823	
H.2 Daily work inside or outside	1.83 (0.374)	2	0.7757	
H.3 Accomplish less	2.76 (0.536)	3	0.7418	
H.4 Limited in how long can do work or activities	2.74 (0.563)	3	0.7477	
H.5 Require help with work or activities	1.85 (0.357)	2	0.6512	

Items had between 4 and 6 response options. However, due to sparse distributions, some of the categories were collapsed in order to have no cell with <5 % of the data (41 participants). The resulting number of response options was a minimum of 2 and a maximum of 5. For all of these a higher score meant better health

[a] Item-rest correlations by scale—corrected for item overlap

[b] Prefixes A, B, C, etc. Correspond to the domain to which the item originally belonged (See Appendix A of ESM with the full Toolbox Vision Survey with the 8 hypothesized domains and 53 items)

We computed F statistics comparing means across major eye conditions (cataracts, glaucoma, macular degeneration, diabetic retinopathy) to evaluate construct validity. Subjects were asked if a doctor had ever told them that they had each one of these eye diseases. We hypothesized that means would be lower in persons self-reporting a history of one of these eye conditions. Finally, we examined the magnitude of the correlations between responses to the six HRQOL scale scores with the 11 NEI-VFQ scales, which were also administered to the same sample. For example, we expected to have higher correlations between the distance vision scale and the NEI-VFQ scales that measure distance vision such as the Distance Activities and Driving scales.

IRB approval

Prospective IRB review and approval were obtained by the USC Institutional Review Board (IRB) (USC Proposal # HS-06-00096) for cognitive interviews. The field test data collection phase of the study was reviewed and approved by the Northwestern University, Medical Social Sciences Institutional Review Board (IRB) (IRB Project Number: STU00013098).

Results

Table 1 summarizes demographic characteristics of the sample. The mean age was 54 years (range 18–85 years). The sample was similar to the 2010 census data in terms of gender. The sample was older, less diverse in terms of race/ethnicity, and tended to be (surprisingly) less educated than the census estimates. Sixty-seven percent ($n = 552$) wore glasses or contact lenses for distance vision; 61 % ($n = 501$) wore glasses or bifocals for near vision.

Table 2 shows item means with standard deviations and number of response options for each item, item-rest correlations, and coefficient alphas for each of the final six multi-item scales. Higher scores represent better functioning and well-being for all items. Items had between four and six response options. Because IRT requires a large number of responses for each response option to provide precise item parameter estimation, we collapsed cells with <5 % of the responses ($n = 41$ responses) to minimize estimation with sparse data. The resulting number of response options ranged from two to five.

Item-rest correlations, corrected for overlap, ranged from 0.50 for item E.9 ("Because of your eyesight, how difficult is it for you to recognize people when they are standing or sitting near you?") on the Near Vision scale, to 0.83 for item A.2 ("Because of your eyesight, how difficult is it for you to recognize colors?") on the Color Vision scale. Over 90 % of all item-rest correlations were larger than 0.60. Coefficient alpha was 0.85 or above for all six scales.

The hypothesized eight-factor model did not fit the data well enough (CFI = 0.695, TLI = 0.954, RMSEA = 0.136). Based on content considerations and modification indices from the initial model, we decided to redistribute the items in the Low Luminance and General Vision domains into the Near and Distance Vision domains. This revised six-factor confirmatory factor analytic model fit the data better

Table 3 Standardized factor loadings from six-factor categorical confirmatory factor analysis model

Factor (items)[b]	Estimated loading	SE
Factor 1—color vision		
A.1 Match colors of clothes	0.983	0.019
A.2 Recognize colors	0.998	0.015
A.3 Distinguish between colors	0.936	0.018
Factor 2—distance vision		
B.1 See in bright sunlight	0.706	0.022
B.2 See in fluorescent lighting	0.816	0.022
B.3 See at night	0.826	0.015
B.4 See street signs at night when in a car	0.846	0.013
B.5 See when going from lighted area into dark place	0.805	0.016
C.1 Go up or down steps, stairs, or curbs	0.838	0.021
C.2 Get around outdoors	0.935	0.019
C.3 See moving objects	0.939	0.018
C.4 See street signs	0.796	0.019
C.5 See television	0.853	0.021
C.6 Drive at night	0.832	0.015
C.7 Drive due to glare from oncoming headlights	0.831	0.015
C.8 Recognize people from across a room	0.810	0.022
D.1 Rate eyesight for distance	0.696	0.023
D.2 See things at a distance during the day	0.842	0.018
D.3 See things at a distance at nighttime	0.868	0.013
Factor 3—near vision		
B.6 Read in dim light	0.813	0.020
D.4 Rate eyesight for close up	0.693	0.022
D.5 See things close up during day	0.814	0.020
D.6 See things close up at nighttime	0.837	0.016
E.1 Read newspaper or book	0.837	0.016
E.2 Find item on crowded shelf	0.891	0.017
E.3 See sign with other signs around	0.899	0.016
E.4 Read labels or instructions	0.844	0.015
E.5 Use computer	0.847	0.020
E.6 Operate household appliance	0.909	0.028
E.7 Write	0.837	0.027
E.8 Do work or hobbies that require to see up close	0.838	0.019
E.9 Recognize people when nearby	0.923	0.031
Factor 4—ocular symptoms		
F.1 Glare from bright lights	0.827	0.019
F.2 Burning or stinging eyes	0.705	0.028
F.3 Redness in eyes	0.695	0.032
F.4 Headaches because of vision	0.759	0.032
F.5 Eyes being sensitive to light	0.832	0.018
F.6 Blurry vision	0.873	0.018
F.7 Itching in or around eyes	0.644	0.029
Factor 5—psychosocial		
G.1 Frustrated or upset	0.924	0.010
G.2 Annoyed or angry	0.923	0.011
G.3 Felt like burden on others	0.935	0.016
G.4 Felt sad or depressed	0.925	0.015
G.5 Felt socially isolated	0.987	0.010
G.6 Concerned about safety at home	0.926	0.021
G.7 Worried	0.837	0.018
G.8 Rely on others	0.869	0.022
G.9 Visits with friends or family	0.949	0.021
Factor 6—role performance		
H.1 Daily activities	0.967	0.010
H.2 Daily work inside or outside	0.977	0.010
H.3 Accomplish less	0.906	0.019
H.4 Limited in how long can do work or activities	0.915	0.016
H.5 Require help with work or activities	0.923	0.021

Model fit: CFI = 0.944, TLI = 0.941, RMSEA = 0.055[a]

[a] Cutoff indices for good model fit: CFI $\geq$ 0.95, TLI $\geq$ 0.95 and RMSEA $<$ 0.06

[b] Prefixes A, B, C, etc. Correspond to the domain to which the item originally belonged. (See Appendix A of ESM with the full Toolbox Vision Survey with the 8 hypothesized domains and 53 items)

than the eight-factor model: CFI = 0.944, TLI = 0.941, and RMSEA = 0.055. Factor loadings from the model are provided in Table 3. Item F.7 (“In the last 7 days, how much of a problem did you have with itching in or around your eyes?”) in the Ocular Symptoms scale had the lowest factor loading, 0.64. Factor loadings for most of the items were above 0.8. Inter-factor correlations from the six-factor categorical confirmatory factor analysis model ranged from 0.67 (Color vision with Ocular Symptoms) to 0.92 (Psychosocial Well-being with Role Performance).

Table 4 shows item parameters and standard errors for the six domains. The threshold parameters ranged from −2.75 (0.27) to 1.99 (0.14), and the slope parameters ranged from 1.13 (0.12) to 4.65 (0.78). Item G.5 (“In the past 7 days, how socially isolated did you feel because of your eyesight?”) had the highest slope. Item D.1 (“In general, how would you rate your eyesight for seeing things at a distance?”) had the largest threshold for the transition between the *Fair* category and *Poor*.

Table 4 IRT parameters: slope (discrimination) and category threshold estimates (SE)[a]

Scale (items)	Slope	Category thresholds			
Color vision					
A.1 Match colors of clothes	1.99 (0.23)	−1.44 (0.12)			
A.2 Recognize colors	1.89 (0.21)	−1.47 (0.13)			
A.3 Distinguish between colors	1.67 (0.20)	−1.40 (0.14)			
Distance vision					
B.1 See in bright sunlight	1.49 (0.13)	−2.17 (0.18)	−1.20 (0.11)	0.38 (0.08)	
B.2 See in fluorescent lighting	2.16 (0.21)	−1.75 (0.13)	−0.73 (0.07)		
B.3 See at night	2.09 (0.16)	−1.74 (0.11)	−0.91 (0.07)	0.54 (0.07)	
B.4 See street signs at night when in a car	2.10 (0.17)	−1.60 (0.11)	−0.78 (0.07)	0.40 (0.07)	
B.5 See when going from lighted area into dark place	1.94 (0.16)	−1.84 (0.12)	−0.96 (0.08)	0.52 (0.07)	
C.1 Go up or down steps, stairs, or curbs	2.13 (0.21)	−1.97 (0.14)	−0.90 (0.08)		
C.2 Get around outdoors	3.29 (0.44)	−1.24 (0.08)			
C.3 See moving objects	3.25 (0.46)	−1.13 (0.07)			
C.4 See street signs	1.97 (0.19)	−1.72 (0.12)	−0.35 (0.06)		
C.5 See television	2.48 (0.24)	−1.87 (0.13)	−0.88 (0.07)		
C.6 Drive at night	2.24 (0.18)	−1.60 (0.10)	−0.90 (0.07)	0.40 (0.06)	
C.7 Drive due to glare from oncoming headlights	2.19 (0.16)	−1.27 (0.08)	−0.50 (0.06)	0.85 (0.08)	
C.8 Recognize people from across a room	2.06 (0.21)	−1.90 (0.13)	−0.84 (0.07)		
D.1 Rate eyesight for distance	1.53 (0.11)	−1.87 (0.14)	−0.56 (0.08)	0.50 (0.08)	1.99 (0.14)
D.2 See things at a distance during the day	2.21 (0.19)	−1.52 (0.10)	−0.35 (0.06)		
D.3 See things at a distance at nighttime	2.39 (0.17)	−1.40 (0.08)	−0.58 (0.06)	0.78 (0.07)	
Near vision					
B.6 Read in dim light	2.00 (0.13)	−1.25 (0.09)	−0.47 (0.06)	0.97 (0.08)	
D.4 Rate eyesight for close up	1.26 (0.10)	−2.36 (0.20)	−0.83 (0.10)	0.31 (0.09)	1.67 (0.15)
D.5 See things close up during day	1.70 (0.18)	−1.93 (0.16)	−0.65 (0.08)		
D.6 See things close up at nighttime	1.96 (0.18)	−2.13 (0.15)	−1.13 (0.08)	0.05 (0.06)	
E.1 Read newspaper or book	1.78 (0.15)	−2.20 (0.17)	−1.39 (0.11)	−0.22 (0.07)	
E.2 Find item on crowded shelf	2.53 (0.24)	−1.89 (0.12)	−0.73 (0.06)		
E.3 See sign with other signs around	2.68 (0.22)	−1.75 (0.11)	−0.50 (0.05)		
E.4 Read labels or instructions	1.91 (0.16)	−2.00 (0.14)	−1.17 (0.09)	0.05 (0.07)	
E.5 Use computer	2.24 (0.19)	−1.95 (0.14)	−0.70 (0.06)		
E.6 Operate house-hold appliance	2.49 (0.31)	−1.53 (0.10)			
E.7 Write	1.98 (0.24)	−1.29 (0.10)			
E.8 Do work or hobbies that require to see up close	2.02 (0.18)	−1.83 (0.13)	−0.55 (0.07)		
E.9 Recognize people when nearby	2.65 (0.37)	−1.66 (0.11)			
Ocular symptoms					
F.1 Glare from bright lights	1.99 (0.18)	−1.48 (0.10)	0.03 (0.06)		
F.2 Burning or stinging eyes	1.38 (0.14)	−2.21 (0.20)	−0.66 (0.09)		
F.3 Redness in eyes	1.28 (0.14)	−2.75 (0.27)	−1.12 (0.13)		
F.4 Headaches because of vision	1.72 (0.20)	−1.12 (0.10)			
F.5 Eyes being sensitive to light	1.98 (0.18)	−1.56 (0.11)	−0.20 (0.06)		
F.6 Blurry vision	2.25 (0.19)	−1.72 (0.11)	−0.45 (0.06)		
F.7 Itching in or around eyes	1.13 (0.12)	−2.58 (0.25)	−0.44 (0.10)		
Psychosocial					
G.1 Frustrated or upset	2.87 (0.26)	−1.58 (0.09)	−0.50 (0.05)		
G.2 Annoyed or angry	2.72 (0.26)	−1.87 (0.11)	−0.75 (0.06)		
G.3 Felt like burden on others	3.42 (0.44)	−1.30 (0.08)			
G.4 Felt sad or depressed	3.41 (0.36)	−1.79 (0.10)	−1.11 (0.06)		
G.5 Felt socially isolated	4.65 (0.78)	−1.33 (0.06)			
G.6 Concerned about safety at home	3.23 (0.44)	−1.40 (0.09)			
G.7 Worried	2.25 (0.20)	−1.82 (0.13)	−0.54 (0.06)		
G.8 Rely on others	2.39 (0.25)	−1.22 (0.09)			
G.9 Visits with friends or family	3.49 (0.50)	−1.57 (0.09)			
Role performance					
H.1 Daily activities	3.41 (0.41)	−1.09 (0.06)			

Table 4 continued

Scale (items)	Slope	Category thresholds	
H.2 Daily work inside or outside	3.53 (0.41)	−1.04 (0.06)	
H.3 Accomplish le	2.60 (0.21)	−1.99 (0.13)	−1.04 (0.07)
H.4 Limited in how long can do work or activities	2.63 (0.27)	−1.83 (0.12)	−0.97 (0.07)
H.5 Require help with work or activities	2.67 (0.33)	−1.22 (0.08)	

Parameters obtained from graded response model estimated using MULTILOG (logistic metric)

[a] Prefixes A, B, C, etc. Correspond to the domain to which the item originally belonged. (See Appendix A of ESM with the full Toolbox Vision Survey with the 8 hypothesized domains and 53 items)

Table 5 Vision-targeted HRQOL scale means (SE) by validity variables: F statistic (p value)

Scale	Mean (SD)	Cataracts	Glaucoma	Macular degeneration	Diabetic retinopathy	Degrees of freedom
I. Color vision	−0.2163 (0.4788)	7.51 (0.0063)	36.77 (0.0000)	46.40 (0.0000)	38.01 (0.0000)	1 817
II. Distance vision	−0.0087 (0.9150)	8.71 (0.0033)	19.08 (0.0000)	31.36 (0.0000)	35.01 (0.0000)	1 817
III. Near vision	−0.0212 (0.8928)	10.80 (0.0011)	28.06 (0.0000)	32.00 (0.0000)	36.78 (0.0000)	1 817
IV. Ocular symptoms	−0.0708 (0.8145)	6.24 (0.0127)	18.78 (0.0000)	21.98 (0.0000)	28.77 (0.0000)	1 817
V. Psychosocial	−0.165 (0.7183)	0.00 (0.9445)	14.23 (0.0002)	27.23 (0.0000)	46.06 (0.0000)	1 817
VI. Role performance	−0.2217 (0.6383)	2.84 (0.0923)	18.71 (0.0000)	32.32 (0.0000)	44.63 (0.0000)	1 817

F statistics with p values for differences by eye disease (cataracts, glaucoma, macular degeneration, diabetic retinopathy) on the six scales are provided in Table 5. All p values were highly statistically significant ($p < 0.01$) except for Ocular Symptoms, Psychosocial Well-being, and Role Performance for cataracts. As hypothesized, means were lower in subjects who self-reported having an eye condition, for all comparisons except the ones previously mentioned.

The largest correlations between the 6 vision-targeted scales and the 11 NEI-VFQ subscales were found between the Distance Vision scale and the NEI-VFQ Driving scale, between the Near Vision and the NEI-VFQ Near Activities scale, between the Psychosocial Well-being and the NEI-VFQ Mental Health scale, and between the Role Performance and the NEI-VFQ Mental Health and Role Difficulties scales (See Table 6).

Discussion

The NIH Toolbox project is designed to produce comprehensive and parsimonious instruments to measure cognitive, emotional, motor, and sensory function. Analyses reported here provided support for the psychometric properties of an instrument measuring vision-targeted HRQOL (See Appendix A of ESM for the field test survey and Appendix B of ESM for a table listing the vision-targeted HRQOL instruments; we reviewed with domains and number of items in each domain).

This measure goes beyond static vision-targeted questionnaires by calibrating all items on the same underlying metric. Different subsets of items can be selected that are targeted to individual respondents, providing a flexible and efficient way of assessment. This makes it possible to administer the items using computer adaptive testing (CAT).

IRT methods also make it possible to evaluate equivalence of item responses between different subgroups (i.e., differential item functioning). An item shows differential item functioning if the probabilities of responding in the same categories varies according to some external variable, after controlling for what is being measured. For example, two subgroups with the same level of underlying vision function might respond differently to the same item, based on language used.

A limitation of this study is that the sample used was not representative of the US national population. Minorities were underrepresented: only 6 % were Black or African American and only 5 % were Hispanic or Latino.

Table 6 Correlations among new-vision-targeted HRQOL scores and Visual Function Questionnaire (VFQ) scales

	New Vision-targeted Scales					
	Color vision	Distance vision	Near vision	Ocular symptoms	Psychosocial	Role performance
Color vision	1.000					
Distance vision	0.490	1.000				
Near vision	0.472	0.746	1.000			
Ocular symptoms	0.447	0.739	0.680	1.000		
Psychosocial	0.478	0.712	0.696	0.730	1.000	
Role performance	0.494	0.624	0.649	0.617	0.762	1.000
General vision	0.308	0.539	0.527	0.431	0.477	0.457
Ocular pain	0.419	0.461	0.435	0.681	0.567	0.523
Near activities	0.448	0.568	0.712	0.551	0.652	0.660
Distance activities	0.470	0.666	0.579	0.563	0.666	0.643
Social functioning	0.501	0.488	0.466	0.485	0.581	0.587
Mental health	0.485	0.594	0.582	0.612	0.771	0.724
Role difficulties	0.455	0.494	0.522	0.515	0.639	0.729
Dependency	0.501	0.474	0.464	0.458	0.616	0.656
Driving	0.386	0.719	0.532	0.571	0.578	0.561
Color vision	0.587	0.324	0.319	0.339	0.401	0.398
Peripheral vision	0.499	0.555	0.522	0.527	0.605	0.619

According to the latest census figures, these minorities represent 13 and 16 % of the US population, respectively. Eighty-eight percent of the sample was of white ethnicity, whereas the US population has 72 % white ethnicity. These figures often are associated with differences in other socio-demographic characteristics (e.g., socio-economic status and education). Participation in online surveys could also be affected by these characteristics, biasing the results in some way that we were not able to evaluate. Further evaluation of the measure in different ethnic groups is needed to ensure that the NIH Toolbox items work well across race/ethnicity groups.

In addition, different ethnic groups have different prevalence rates and risk indicators for ocular diseases. For example, Latinos have higher rates of vision affected by severe diabetic retinopathy than do non-Hispanic Whites. Rates of glaucoma are also significantly higher in Latinos than in non-Hispanic Whites [42]. In the online sample, we used for analysis, and the number of respondents with glaucoma and macular degeneration was small. Even though the prevalence of these diseases in the general population is small, the number of subjects in the online panel with these diseases does not allow for subgroup analyses. Furthermore, severity of eye disease was not captured. Another limitation is that we were unable to measure visual acuity for our participants and therefore were not able to conduct construct analyses relevant to the variable that is likely of greatest interest for vision studies. Future studies should evaluate the responsiveness of the measure. However, because the items in the measure were adapted closely from existing instruments, and those existing instruments have been shown to be responsive to changes in visual acuity, the measure proposed here would likely possess similar responsiveness.

Future phases of development with this instrument could include development of items that target HRQOL for specific vision problems. A Spanish-language version of the instrument has been created and will be administered to a Spanish speaking sample in the future. Finally, readability analyses of all items would be useful to further simplify wording, increasing instrument utility in populations with lower levels of literacy.

Acknowledgments This study was supported by the Blueprint for Neuroscience Research, National Institutes of Health under Contract No. HHS-N-260-2006-00007-C. Drs. Hays and Paz were also supported in part by NIA grant P30AG021684. Dr. Hays was additionally supported in part by NIA grant P30-AG028748 and the NCMHD P20MD000182.

Conflict of interest No conflicting relationship exists for any author.

References

1. Ader, D. N. (2007). Developing the patient-reported outcomes measurement information system (PROMIS). *Medical Care, 45*, S1–S2. doi:10.1097/01.mlr.0000260537.45076.74.
2. http://nihroadmap.nih.gov/clinicalresearch/overview-dynamicoutcomes.asp.
3. http://www.nihtoolbox.org/.

4. Gershon, R. C., Cella, D., Fox, N. A., et al. (2010). Assessment of neurological and behavioural function: The NIH Toolbox. *Lancet Neurology, 9*(2), 138–139. doi:10.1016/S1474-4422(09)70335-7.
5. Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st Century. *Medical Care, 38*(9 Suppl), II28–II42.
6. Terwee, C. B., Gerding, M. N., Dekker, F. W., et al. (1998). Development of a disease-specific quality of life questionnaire for patients with Graves' ophthalmopathy: The GO-QOL. *British Journal of Ophthalmology, 82*, 773–779.
7. Terwee, C. B., Gerding, M. N., Dekker, F. W., et al. (1999). Test-retest reliability of the GO-QOL: A disease-specific quality of life questionnaire for patients with Graves' ophthalmopathy. *Journal of Clinical Epidemiology, 52*, 875–884.
8. Terwee, C. B., Dekker, F. W., Mourits, M. P., et al. (2001). Interpretation and validity of changes in scores on the Graves' Ophthalmopathy quality of life questionnaire (GO-QOL) after different treatments. *Clinical Endocrinology, 54*, 391–398.
9. Lamoureux, E. L., Pallant, J. F., Pesudovs, K., et al. (2007). The impact of vision impairment questionnaire: An assessment of its domain structure using confirmatory factor analysis and Rasch analysis. *Investigative Ophthalmology and Visual Science, 48*, 1001–1006.
10. Weih, L. M., Hassell, J. B., & Keeffe, J. E. (2002). Assessment of the impact of vision impairment. *Investigative Ophthalmology and Visual Science, 43*, 927–935.
11. Lamoureux, E. L., Hassell, J. B., & Keeffe, J. E. (2004). The determinants of participation in activities of daily living in people with impaired vision. *American Journal of Ophthalmology, 137*, 265–270.
12. Lamoureux, E. L., Hassell, J. B., & Keeffe, J. E. (2004). The impact of diabetic retinopathy on participation in daily living. *Archives of Ophthalmology, 122*, 84–88.
13. Hassell, J. B., Lamoureux, E. L., & Keeffe, J. E. (2006). Impact of age-related macular degeneration on quality of life. *British Journal of Ophthalmology, 90*, 593–596.
14. Wolffsohn, J. S., & Cochrane, A. L. (2000). Design of the low vision quality of life questionnaire (LVQOL) and measuring the outcome of low-vision rehabilitation. *American Journal of Ophthalmology, 130*(6), 793–802.
15. De Boer, M. R., de Vet, H. C. W., et al. (2005). Changes to the subscales of two vision-related quality of life questionnaires are proposed. *Journal of Clinical Epidemiology, 58*, 1260–1268.
16. Hays, R., Mangione, C., Ellwein, L., NEW-RQL Research Group, et al. (2003). Psychomeric properties of the National Eye Institute—refractive error quality of life instrument. *Ophthalmology, 110*, 2292–2301.
17. Mangione, C. M., Berry, S., Spritzer, K., et al. (1998). Identifying the content area for the 51-item National Eye Institute Visual Function Questionnaire: Results from focus groups with visually impaired persons. *Archives of Ophthalmology, 116*, 227–233.
18. Mangione, C. M., Lee, P. P., Gutierrez, P. R., et al. (2001). Development of the 25-item National Eye Institute Visual Function Questionnaire. *Archives of Ophthalmology, 119*, 1050–1058.
19. Mangione, C. M., Lee, P. P., Pitts, J., et al. (1998). Psychometric properties of the National Eye Institute Visual Function Questionnaire (NEI-VFQ). NEI-VFQ Field Test Investigators. *Archives of Ophthalmology, 116*, 1496–1504.
20. Carta, A., Braccio, L., Belpoliti, M., et al. (1998). Self-assessment of the quality of vision: Association of questionnaire score with objective clinical tests. *Current Eye Research, 17*, 506–511.
21. Gothwal, V. K., Wright, T. A., Lamoureux, E. L., et al. (2009). Rasch analysis of the quality of life and vision function questionnaire. *Optometry and Vision Science, 86*, E836–E844.
22. Vitale, S., Schein, O. D., Meinert, C., et al. (2000). The refractive status and vision profile: A questionnaire to measure vision-related quality of life in persons with refractive error. *Ophthalmology, 107*(8), 1529–1539.
23. Schein, O. D., Vitale, S., Cassard, S. D., et al. (2001). Patient outcomes of refractive surgery. The refractive status and vision profile. *Journal of Cataract and Refractive Surgery, 27*, 665–673.
24. Steinberg, E. P., Tielsch, J. M., Schein, O. D., et al. (1994). The VF-14: An index of functional impairment in patients with cataract. *Archives of Ophthalmology, 112*, 630–638.
25. Linder, M., Chang, T. S., Scott, I. U., et al. (1999). Validity of the visual function index (VF-14) in patients with retinal disease. *Archives of Ophthalmology, 117*, 1611–1616.
26. Fairchild, C. J., Chalmers, R. L., & Begley, C. G. (2008). Clinically important difference in dry eye: Change in IDEEL-symptom bother. *Optometry and Vision Science, 85*(8), E699–E707.
27. Rajagopalan, K., Abetz, L., Mertzanis, P., et al. (2005). Comparing the discriminative validity of two generic and one disease-specific health-related quality of life measures in a sample of patients with dry eye. *Value Health, 8*, 168–174.
28. Abetz, L., Venkataraman, K., Mertzanis, P, et al. (2003). The development, reliability and validity of a questionnaire to assess the impact of dry eyes on everyday life (IDEEL). *Investigative Ophthalmology and Visual Science, 44*: 2477 (E-Abstract).
29. Mangione, C. M., Phillips, R. S., Seddon, J. M., et al. (1992). Development of the 'Activities of Daily Vision Scale'. A measure of visual functional status. *Medical Care, 30*, 1111–1126.
30. Pesudovs, K., Garamendi, E., Keeves, P., et al. (2003). The activities of daily vision scale for cataract surgery outcomes: Re-evaluating validity with Rasch analysis. *Investigative Ophthalmology and Visual Science, 44*(7).
31. Sloane, M. E., Ball, K., Owsley, C., et al. (1992). The visual activities questionnaire: Developing an instrument for assessing problems in everyday visual tasks. *Technical Digest, Noninvasive Assessment of the Visual System, 1*, 26–29.
32. Gothwal, V. K., Wright, T. A., Lamoureux, E. L., & Pesudos, K. (2009). Visual activities questionnaire: Assessment of subscale validity for cataract surgery outcomes. *Journal of Cataract and Refractive Surgery, 35*(11), 1961–1969.
33. Owsley, C., Mcgwin, G., Scilley, K., et al. (2006). Development of a questionnaire to assess vision problems under low luminance in age-related maculopathy. *Investigative Ophthalmology and Visual Science, 47*, 528–535.
34. http://www.fda.gov/downloads/AboutFDA/CentersOffices/CDER/ucm118795.pdf.
35. DeWalt, D., Rothrock, N., Yount, S., PROMIS Cooperative Group, et al. (2007). Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care, 45*, S12–S21.
36. Cella, D., Riley, W., Stone, A., et al. (2010). the patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology, 63*, 1179–1194.
37. www.greenfieldonline.com.
38. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.
39. MPlus: Muthen & Muthen. www.statmodel.com/.
40. Morales, L. S., Flowers, C., Gutierrez, P., et al. (2006). Item and scale differential functioning of the mini-mental state exam assessed using the differential item and test functioning (DFIT) framework. *Medical Care, 44*, S143–S151.
41. Du Toit, M. (2003). IRT from Scientific Software International. SSI, Inc.
42. Paz, S. H. (2012). *Vision and eye care. Encyclopedia of Immigrant Health*. Berlin: Springer Science + Business Media, LLC.