Development of the 25-Item National Eye Institute Visual Function Questionnaire

Carol M. Mangione, MD, MSPH; Paul P. Lee, MD, JD; Peter R. Gutierrez, MA; Karen Spritzer, BA; Sandra Berry, MS; Ron D. Hays, PhD; for the National Eye Institute Visual Function Questionnaire Field Test Investigators

Objective: To develop and test the psychometric properties of a 25-item version of the National Eye Institute Visual Function Questionnaire (NEI VFQ-25).

Design: Prospective observational cohort study of persons with 1 of 5 chronic eye diseases or low vision who were scheduled for nonurgent visits in ophthalmology practices and a reference sample of persons without eye disease.

Setting: Eleven university-based ophthalmology practices and the NEI Clinical Center.

Patients: Eligible participants had to have 1 of the following eye conditions: age-related cataracts, age-related macular degeneration, diabetic retinopathy, primary open-angle glaucoma, cytomegalovirus retinitis, or low vision from any cause. Seven of the 12 sites also enrolled persons in a reference sample. Reference sample participants had no evidence of underlying eye disease but were scheduled for either screening eye examinations or correction of refractive error. All eligible persons had to be 21 years or older, English speaking, and cognitively able to give informed consent and participate in a health status interview.

Measurements and Main Results: To provide the data needed to create the NEI VFQ-25, all subjects completed an interview that included the 51-item NEI VFQ. Estimates of internal consistency indicate that the subscales of the NEI VFQ-25 are reliable. The validity of the NEI VFQ-25 is supported by high correlations between the short- and long-form versions of the measure, observed between-group differences in scores for persons with different eye diseases of varying severity, and the moderate-to-high correlations between the NEI VFQ-25 subscales that have the most to do with central vision and measured visual acuity.

Conclusions: The reliability and validity of the NEI VFQ-25 are comparable to those of the 51-item NEI VFQ field test version of the survey. This shorter version will be more feasible in settings such as clinical trials where interview length is a critical consideration. In addition, preliminary analyses indicate that the psychometric properties of the NEI VFQ-25 are robust for the eye conditions studied; this suggests that the measure will provide reproducible and valid data when used across multiple conditions of varying severity.

Arch Ophthalmol. 2001;119:1050-1058

From the Department of Medicine, UCLA School of Medicine, Los Angeles, Calif (Drs Mangione and Hays, Mr Gutierrez, and Ms Spritzer); RAND Health Program, RAND, Santa Monica, Calif (Drs Mangione, Lee, and Hays and Ms Berry); and Department of Ophthalmology, Duke University Medical Center, Durham, NC (Dr Lee). A list of the NEI VFQ Field Test Investigators can be found on page 1058.

HEALTH-RELATED quality of life (HRQOL) measures functioning and well-being in physical, mental, and social health realms of life and reflects the influence of a broad range of health conditions simultaneously. Because of the response burden and impact on participation rates, it is imperative that HRQOL measures be as short as feasible. Recognition of the important role that survey length plays in both data quality and costs has led to the creation of short-form versions of health surveys such as the 18-item version of the Patient Satisfaction Questionnaire,[1] the 5-item version of the Mental Health Inventory,[2] and the 36-Item Short-Form Health Survey[3] and 12-Item Short-Form Health Survey.[4] There is even greater need for short-form versions of vision-targeted surveys such as the National Eye Institute Visual Function Questionnaire (NEI VFQ),[5] because to comprehensively evaluate HRQOL requires the use of vision-targeted measures in combination with generic measures.

A number of reliable and valid short questionnaires assess difficulty with activities that require vision or assess symptoms from eye diseases and their treatments.[6-11] These vary in length from 14 to 31 questions. However, most of these surveys only capture one dimension of vision-targeted HRQOL. The 51-item field test version of the NEI VFQ[5,12] was designed to capture the influence of vision on multiple dimensions of HRQOL, such as emotional well-being and social functioning. Early feedback from users indicated that

SUBJECTS AND METHODS

STUDY DESIGN AND POPULATION

Two separate samples of visually impaired persons who completed the 51-item NEI VFQ were pooled for analyses. The first sample consisted of 262 persons from 5 academic centers who participated in the 1994 pilot test of the NEI VFQ (pilot study), and the second sample consisted of 597 persons from the 1996 NEI VFQ Psychometric Field Test (field test study). These 2 data sets were combined to provide a broader spectrum of disease severity than either data set represents alone.

A description of the sample-specific enrollment criteria is provided elsewhere.[5,12] Briefly, in both samples eligible participants had to be at least 21 years old, had to be English speaking, and had to pass a cognitive test based on an abbreviated version of the Folstein Mini-Mental State Examination,[13] where participants were only asked to complete the whole measure if they made an error on the initial orientation, short-term memory, or attention questions. Both the pilot study and field test protocols were approved by all institutional review boards, and all participants provided written informed consent.

Pilot study participants had 1 or more of the following eye conditions: age-related cataracts, age-related macular degeneration (ARMD), diabetic retinopathy, primary open-angle glaucoma, or cytomegalovirus retinitis. To be eligible for the field test, participants had 1 ocular condition only, whereas pilot study participants with multiple conditions were eligible. For these reasons, all tests of validity between the NEI VFQ-25 and clinical variables were performed solely with field test study data. The field test also enrolled persons with low vision from any cause and a reference group with no underlying eye disease. Condition-specific eligibility is described elsewhere.[5,12]

DATA COLLECTION

Demographic and Medical Characteristics

Each participant completed a 16-item medical comorbidity checklist adapted from the Medical Outcomes Study[14] and reported demographic characteristics. In the pilot study, currently corrected Snellen acuity and all diagnoses of eye diseases were abstracted from the medical records. As part of a research protocol,[15] field test study participants completed a dilated examination that included an assessment of current eye diseases and previous ophthalmic surgical procedures. These participants were tested using binocular and monocular Early Treatment Diabetic Retinopathy Study (ETDRS) visual acuity while wearing their current, or "walking about," correction.[16] Patients with visual acuity so poor that they could not read any of the largest letters at 4 m were tested at 1 m and for light perception. The presence of cataracts was graded during a slitlamp examination using the Age-Related Eye Diseases Study reference standards.[17] Each participant had a complete fundus examination to rule out significant additional ocular pathologic conditions and to grade the severity of diabetic retinopathy, ARMD, primary open-angle glaucoma, and cytomegalovirus retinitis. Field test study ophthalmologic examinations were performed prospectively by trained examiners who followed the protocol in the NEI VFQ manual of procedures.[15]

NEI VFQ

The 51-item NEI VFQ is a vision-targeted survey that assesses the influence of visual impairment on HRQOL. The content of the NEI VFQ was derived from multicondition focus groups.[12] The 51-item NEI VFQ takes 15 minutes on average to administer as an interview and includes the following: multi-item scales to rate overall health (2 items), overall vision (2 items), difficulty with near vision (7 items), difficulty with distance vision (7 items), limitations in social functioning due to vision (4 items), role limitations due to vision (5 items), dependency on others due to vision (5 items), mental health symptoms due to vision (8 items), future expectations for vision (3 items), driving difficulties (4 items), and pain and discomfort around the eyes (2 items) and single items to assess peripheral and color vision. Each subscale is scored so that 0 represents the lowest and 100 the best possible score. A full description of the 51-item field test NEI VFQ has been published.[5] All field test participants completed the 51-item NEI VFQ as part of a larger interview that contained multiple surveys. Pilot study participants completed the 96-item NEI VFQ, which included the 51 items that were eventually retained in the field test version. Only the 51 field test items were candidates for the item reduction analysis described herein.

Guiding Principles for Item Reduction

The following qualitative criteria were used to identify candidate items for the NEI VFQ-25:

- Retained items should have low missing data rates. The inclusion of items that are most likely to be answered by most persons will maximize the available information from each participant.
- To maintain breadth of content, the intent is to have all 51-item NEI VFQ constructs represented in the shorter survey, thereby remaining faithful to the range of topic areas mentioned by participants in the original focus groups.[12] To this end, the single-item color and peripheral vision questions are retained. Similarly, although the questions on driving were only answered by approximately 60% of the field test sample, these questions are retained because driving is highly valued and because difficulty with driving may motivate persons to seek eye care.
- Priority is placed on retaining items with approximately normal distributions of responses over those with skewed distributions (large ceiling or floor effects). Normally distributed responses are likely to reflect improvements or declines when used longitudinally, which should maximize the ability of the questionnaire to discriminate between persons with clinical deterioration over time and between eye conditions of varying severity in cross-sectional studies.

Continued on next page

a shorter version was needed for both research and clinical settings.

The goal of this study was to develop a short-form version of the NEI VFQ that preserves the multidimensional content, reliability, and validity of the full-length survey and also could be completed in approximately 5 minutes. This report describes the results of analyses designed to identify the best questions for inclusion in the 25-item version of the NEI VFQ (NEI VFQ-25) and examines the reliability and validity of the short-form version.

• Once these 3 qualitative criteria are taken into consideration, the items that explain the greatest portion of variance for each of the original 51-item NEI VFQ subscales in linear regression models are retained in the NEI VFQ-25.

STATISTICAL ANALYSES

Descriptive Statistics

The distributions of chronic ophthalmologic and medical conditions and demographic characteristics are displayed by sample. The NEI VFQ subscales are not strictly ordinal or equal interval measures, but because they approximate interval-level measures, parametric statistics are computed. Parametric statistics are robust to minor deviations from assumptions and more powerful than nonparametric statistical methods.[18] The item-subscale correlations are calculated, and the proportion of persons with NEI VFQ item-level scores at the ceilings and floors for each question is also presented for the combined sample.

Item Reduction Models

To create a derivation sample to construct linear regression models and a validation sample to assess the accuracy of the item selection results, the combined pilot study (n=262) and field test samples (n=597) were randomly split into halves. In each linear regression model, the dependent variable is a multi-item subscale from the 51-item NEI VFQ, and the independent variables are the individual items that constitute each subscale. A stepwise regression procedure is used to maximize the proportion of variance in the subscale score explained by a smaller set of items from that specific subscale.[19] For example, if 2 items, such as reading ordinary print in newspapers and reading the fine print on medication bottles, were both part of the same near vision subscale, the item with the higher item-subscale correlation in the linear regression model is selected for inclusion in the shorter version. This procedure is designed to select the fewest items that account for the greatest proportion of variance in the long-form subscale. Rather than applying an absolute cutoff value of the inclusion, items were selected that maximized the proportion of variance explained. To determine whether the results are similar when the field test and the pilot study samples are analyzed separately, sample-specific models were constructed. These models identified 24 of the same 25 items reported for the combined sample and are not described further in this report.

To address potential biases in item selection introduced by missing data, 2 sets of linear regression models were constructed for each of the NEI VFQ subscales. The first set did not substitute values for item-level missing data, and the second set assigned the mean item-level score from participants who answered the item to items that were missed. Final item selection is based on consensus of authors after comparing the proportion of variance explained and items selected by both sets of models. All statistical tests are considered to be significant at $P \leq .05$ (2-tailed), and exact P values are reported.

Reliability

To estimate the internal consistency of the NEI VFQ-25, we calculated the Cronbach coefficient α[20] for each of the multi-item scales. Because coefficient α increases with the number of scale items and the size of their covariances, it is important to compare the values from the short-form subscales to results obtained from the original long-form version.

Validity

To estimate the validity of the NEI VFQ-25 scores across eye conditions, a series of comparisons were conducted that follow the same a priori between-condition comparisons reported for the 51-item NEI VFQ.[5] The first set of comparisons assesses the magnitude of correlation between each 51-item subscale score, representing a "gold standard" measure of each construct and the 25-item subscale score that represents the same construct.

The second comparisons assess whether participants with poorer vision have statistically significantly lower scores on the NEI VFQ-25 than those with better vision. For example, participants in the reference group should have the best average scores, those in the low-vision group should have the poorest scores, and depending on the severity of the underlying condition, the remainder of the groups should have scores between these 2 extreme groups. These statistical comparisons of mean scores include the following: (1) scores from persons with low vision vs the reference group; (2) scores from persons with visually significant cataract vs the reference group; and (3) comparisons of selected subscales, such as the near vision scale for persons with ARMD vs the reference group, and peripheral vision scores for persons with glaucoma vs the reference group. In parallel with previously reported analyses for the 51-item NEI VFQ, linear regression models were used to adjust for between-group differences in age, sex, race, and a summary count of medical comorbidities, since these characteristics may independently influence HRQOL. Adjusted means were compared for the different subgroups.

The third set of comparisons assesses the magnitude of correlations between responses on the NEI VFQ-25 and performance-based measures that are part of the ophthalmologic examination. For example, patients with poorer ETDRS visual acuity across all conditions should have lower scores on the NEI VFQ-25. These correlations should be greatest for the subscales that capture difficulty with activities that require central visual acuity, such as the near vision scale and the driving scale, and lowest for subscales that capture other aspects of vision-targeted HRQOL, such as ocular pain. To test the significance of these associations, Pearson correlation coefficients were calculated between NEI VFQ-25 scores and visual acuity for all field test study participants and between visual field scores in the better and worse eyes for persons from the field test sample with glaucoma.

RESULTS

STUDY SAMPLE

Across the 2 samples, 859 persons contributed data for the item reduction analyses. By design, the field test study included fewer whites (63% vs 81%) and oversampled African Americans (29% vs 11%; χ^2 across race categories, $P=.01$) compared with the pilot study. Women were recruited for both studies in similar proportions (pilot study, 54%; field test, 59%; χ^2, $P=.15$) and similar proportions were working (pilot, 40%; field test, 36%; χ^2,

Table 1. Clinical Characteristics

Characteristics	1994 Pilot Study (n = 262)	1996 Field Test (n = 597)	P
Visual acuity*			
Better eye, median (range)	20/30 (20/15 to 20/400)	20/40 (20/13 to no light perception)	.46†
Worse eye, median (range)	20/50 (20/20 to 20/400)	20/50 (20/13 to no light perception)	.63†
Chronic eye diseases, No. (%)‡			
Age-related cataract	81 (31)	93 (16)	
Primary open-angle glaucoma	72 (27)	77 (13)	
Diabetic retinopathy	58 (22)	123 (21)	
Age-related macular degeneration	56 (21)	108 (18)	
Cytomegalovirus retinitis	21 (8)	37 (6)	
Low vision§	Not applicable	90 (15)	
Normal reference	Not applicable	122 (20)	
Medical comorbidities, No. (%)			
0	30 (12)	74 (12)	
1	43 (16)	110 (18)	
2	49 (19)	115 (19)	.19‖
3	43 (16)	113 (19)	
≥4	97 (37)	185 (31)	

*For the pilot study, Snellen visual acuity was abstracted from the medical records for 220 of 262 participants, whereas in the field test study, Early Treatment Diabetic Retinopathy Study acuity was measured prospectively on all participants.
†Wilcoxon signed rank test.
‡Number of diseases for pilot study sample totals to more than 262 because some participants had more than 1 condition.
§Low vision from any cause. Also, low-vision participants share diagnoses with other causes; hence, the total was more than 597 in the field test sample.
‖Mantel-Haenszel χ^2 test.

$P=.25$). Participants in the field test study were older than those in the pilot study (64 vs 61 years; t test, $P=.02$). Visual acuity between the 2 samples was similar when the normal reference group was included in the field test sample ($P=.46$) (**Table 1**), but was significantly poorer for the field test study when only persons with ocular disease were considered ($P=.01$).

Except for the general vision subscale, each of the remaining 8 mean subscale scores for the NEI VFQ-25 were significantly lower than scores for the same 51-item subscales. This was primarily due to the deletion of questions with ceiling effects.

The percentage of item-level responses that were missing or at the extremes of the distributions for the combined sample is shown in **Table 2**. These data helped identify which items to delete when constructing the NEI VFQ-25. For example, 21% of persons surveyed do not participate in games or card playing, and among those who do, 47% reported no difficulty with this activity. Therefore, by 2 criteria (high missing rate and skewed distribution), this item was not retained in the short-form version of the survey.

ITEM-LEVEL REDUCTION

A number of qualitative item reduction decisions were made before constructing the linear regression models. For example, based on reliability and content validity considerations, both ocular pain questions were retained in the NEI VFQ-25. Also, the 3 items that constitute the expectations for future vision subscale were not retained in the NEI VFQ-25, because responses to formal cognitive debriefing questions indicated that participants found these items difficult to answer, as evidenced by the low internal consistency of this subscale (coefficient $\alpha=.66$). The cognitive debriefing questions were asked of field test participants after the administration of the 51-item NEI VFQ. These questions are described in the NEI VFQ manual of operations.[15]

Regression results (**Table 3**) show that for all of the multi-item subscales examined, more than 85% of the variance in the long-form scores is explained by the items retained in the short-form versions, with most exceeding 90%. A comparison of results for models that did not substitute mean values for missing data vs those that did (not shown) indicates that the items selected and the proportion of variance explained were similar for both methods. Both random halves of the data set identified the same subset of items for retention in the short-form version of the survey. Residuals from these models were found to be normally distributed. Application of the qualitative item reduction criteria described herein and the results of the regression analyses led to the identification of 25 items for inclusion in the short-form version. NEI VFQ 51-item subscale scores can be estimated from NEI VFQ-25 scores with the β coefficients from the regression equations displayed in Table 3.

In summary, decisions regarding 16 of the 25 items dropped were based on the results of the linear regression models. The reasons for dropping the remaining 9 items can be found in Table 2.

NEI VFQ-25 ITEMS AND SUBSCALE SCORES

The items identified for inclusion in the NEI VFQ-25 are shown in boldface type in Table 2. The reduction in subscale items is as follows: the general health and general vision subscales were shortened to single-item scales; the near vision, distance vision, and driving subscales were decreased to 3 items each; the social functioning subscales was halved to 2 items; the mental

Table 2. Item Subscale Correlations and the Number and Percentage of Item Responses at the Ceiling, Floor, or Missing for the 51-Item National Eye Institute Visual Function Questionnaire (n = 859)*

Subscale and Item	Item-Subscale Correlation†	Missing, No. (%)	Floor, No. (%)	Ceiling, No. (%)
General health				
5-Level health rating	0.95	0	36 (4)	132 (15)
0-10 Health rating	0.89	2 (<1)	3 (<1)	86 (10)
General vision				
6-Level general vision	0.92	1 (<1)	4 (<1)	88 (10)
0-10 Vision rating	0.91	5 (1)	5 (1)	50 (6)
Near vision				
Reading normal newsprint	0.90	5 (1)	137 (16)	305 (36)
Reading small print	0.88	2 (<1)	145 (17)	229 (27)
Seeing well up close	0.84	8 (1)	51 (6)	309 (36)
Seeing for games/cards	0.88	180 (21)	77 (9)	403 (47)
Finding objects on crowded shelf	0.84	3 (<1)	13 (2)	468 (55)
Reading mail/bills accurately	0.88	16 (2)	70 (8)	517 (61)
Shaving/styling hair/makeup	0.74	7 (1)	25 (3)	575 (67)
Distance vision				
Going out to movies/plays	0.86	119 (14)	74 (9)	423 (49)
Going down stairs at night	0.78	33 (4)	33 (4)	297 (35)
Reading street signs	0.84	3 (<1)	63 (7)	374 (44)
Going down stairs in daylight	0.80	8 (1)	5 (1)	474 (55)
Recognizing faces in room	0.85	1 (<1)	35 (4)	428 (50)
Seeing television programs	0.83	7 (1)	16 (2)	512 (60)
Participating in sports/outdoors	0.78	90 (11)	41 (5)	515 (60)
Driving				
Daylight familiar places	0.65	137 (16)	166 (19)	480 (56)
Daylight unfamiliar places	0.84	311 (36)	22 (3)	329 (38)
Nighttime familiar places	0.85	309 (36)	80 (9)	143 (17)
Difficult conditions	0.87	319 (37)	32 (4)	247 (29)
Peripheral vision‡				
Seeing objects off to side	1.0	3 (1)	7 (1)	347 (58)
Color vision				
Difficulty matching clothes	1.0	4 (<1)	6 (1)	641 (75)
Ocular pain				
Amount pain	0.92	1 (<1)	5 (1)	490 (57)
Amount time: pain	0.86	0	3 (<1)	616 (72)
Role limitations				
Accomplish less	0.87	3 (<1)	60 (7)	393 (46)
Limited in things can do	0.89	4 (<1)	53 (6)	420 (49)
Have more help	0.85	3 (<1)	27 (3)	481 (56)
Let others do more	0.80	5 (1)	13 (2)	579 (68)
Limited in endurance	0.82	6 (1)	33 (4)	456 (53)
Dependency				
Need much help from others	0.86	3 (<1)	0	543 (63)
Stay home most of time	0.83	2 (<1)	44 (5)	559 (65)
Others know personal business	0.75	3 (<1)	22 (3)	602 (70)
Do not leave home alone	0.82	3 (<1)	53 (6)	616 (72)
Rely too much on others' word	0.86	4 (<1)	37 (4)	569 (66)
Social function				
Seeing how people react	0.84	17 (2)	29 (3)	550 (64)
Normal social activities	0.89	5 (1)	9 (1)	613 (71)
Visiting others	0.89	35 (4)	18 (2)	549 (64)
Entertaining at home	0.82	28 (3)	26 (3)	642 (75)
Mental health				
Amount true: frustrated	0.85	5 (1)	88 (10)	413 (48)
Amount true: irritable	0.76	3 (<1)	34 (4)	527 (61)
Amount true: embarrassment	0.67	0	67 (8)	477 (56)
Amount true: no control	0.77	2 (<1)	75 (9)	428 (50)
Amount time: thinking	0.73	3 (<1)	75 (9)	145 (17)
Amount time: worry	0.73	0	57 (7)	282 (33)
Amount time: frustrated	0.79	2 (<1)	64 (8)	262 (30)
Amount time: irritable	0.62	4 (<1)	5 (1)	647 (75)
Expectations				
Expectations next year	0.80	14 (2)	10 (1)	78 (9)
Expect better eyesight	0.83	3 (<1)	218 (25)	91 (11)
Expect worse eyesight	0.82	1 (<1)	96 (11)	164 (19)

*Items in boldface type indicate items retained in the short-form version. The general health rating question is a "stand alone" item that is scored separately from the 25-item National Eye Institute Visual Function Questionnaire composite.

†Pearson correlation with the 51-item subscale score.

‡The peripheral vision item was developed after the pilot study was conducted; hence, the percentages represent values for the field test sample only.

Table 3. Summary of the Item-Level Scale Reduction Analysis*

Subscale	Retained Items	Proportion of Variance (R^2) in Full-Length Scale Explained by Retained Items		Intercept	β
		Derivation	Confirmation		
General health	1 of 2	0.90	0.90	21.8	0.75
General vision	1 of 2	0.85	0.86	10.8	0.83
Near vision	3 of 7	0.94	0.94	5.8	0.94
Distance vision	3 of 7	0.95	0.94	16.2	0.85
Driving	3 of 4	0.96	0.96	5.2	0.95
Peripheral vision	1 of 1	NA	NA	. . .	1.0†
Color vision	1 of 1	NA	NA	. . .	1.0†
Ocular pain‡	2 of 2	NA	NA	. . .	1.0†
Vision specific					
Role limitations	2 of 5	0.88	0.88	14.9	0.84
Dependency	3 of 5	0.93	0.95	12.2	0.87
Social function	2 of 4	0.92	0.89	19.1	0.80
Mental health	4 of 8	0.94	0.94	14.8	0.82
Expectations for future vision§	0 of 3	NA	NA	. . .	NA

*Derivation set was the first random half (n = 437) and confirmation set was the second random half (n = 422). NA indicates not applicable.
†The 51-item and 25-item versions of these 3 scales are identical.
‡Both ocular pain items were retained based on content validity considerations.
§The expectations for future vision subscale was not retained in the short-form version because based on the cognitive debriefing questions the items were upsetting to some participants and because the scale had a high correlation with the mental health subscale.

health subscale was condensed from 8 to 4 items; the dependency subscale was decreased from 5 to 3 items; and the role functioning subscale was shortened from 5 to 2 items. The remaining ocular pain (2 items), color vision (1 item), and peripheral vision (1 item) subscales are unchanged. A prepublication version of the NEI VFQ-25 has a 2-item, rather than the current 3-item, driving scale. We recommend that the 3-item driving scale described herein be used to increase the internal consistency of this subscale. Mean scale scores for the NEI VFQ-25 are displayed by ocular condition for the field test sample in **Table 4**. The NEI VFQ-25 subscale scores are an average of the items in the subscale transformed to a 0 to 100 scale, where 100 represents the best possible score on the measure and 0 represents the worst. The composite NEI VFQ-25 score is an unweighted average of the responses to all items except for the general health rating question. The general health question is treated as a stand-alone item, because it is a robust marker of overall health status in many population-based studies and provides a comparative benchmark for groups of persons who complete the NEI VFQ-25. A copy of the NEI VFQ-25 and the scoring algorithm can be obtained from the NEI or the RAND Health Web site (http://www.rand.org/health_surveys/vfq25/).

RELIABILITY

Internal consistency estimates for the NEI VFQ-25 subscales ranged from 0.71 to 0.85. Among persons with eye diseases, all of the 8 multi-item subscales had internal consistency estimates greater than or equal to 0.70, indicating that the measure has acceptable reliability for group-level comparisons.[21]

VALIDITY OF THE NEI VFQ-25

Correlations between the NEI VFQ-25 versions of each subscale and their respective long-form version were greater than 0.90 (Table 3). For the 9 subscales that were shortened for the NEI VFQ-25, the adjusted mean scores for persons in the reference group were significantly higher compared with those with either low vision or visually significant cataract (**Figure**). Adjusted mean (SEM) scores for the NEI VFQ-25 near and distance vision scores for participants in the reference group vs those with ARMD were 90 (2) and 57 (2) and 91 (2) and 58 (2), respectively. Finally, adjusted mean (SEM) scores for the NEI VFQ-25 peripheral vision question for reference group members vs those with glaucoma were 97 (2) and 76 (2). All of these differences in mean scores were statistically significant. These selected comparisons provide evidence of between-group validity for the NEI VFQ-25.

Correlations between responses on the NEI VFQ-25 and ETDRS visual acuity were in the range of 0.65 to 0.70 for subscales that reflected degree of difficulty with visual activities related to general vision, near vision, and distance vision (**Table 5**). The remaining subscales showed lower correlations ranging from 0.39 to 0.69, with the exception of the ocular pain subscale, which showed the lowest correlations (between 0.06 and 0.11). Correlations between each of the subscales and visual acuity in the better and worse eyes were similar in magnitude. Among those with glaucoma, the Advanced Glaucoma Intervention Study[22] visual field loss scores had moderate statistically significant correlations with the NEI VFQ-25 composite score, general vision, distance vision, near vision, peripheral vision, social functioning, dependency, and mental health subscales (Table 5). Additional evidence for the validity of the NEI VFQ-25 is

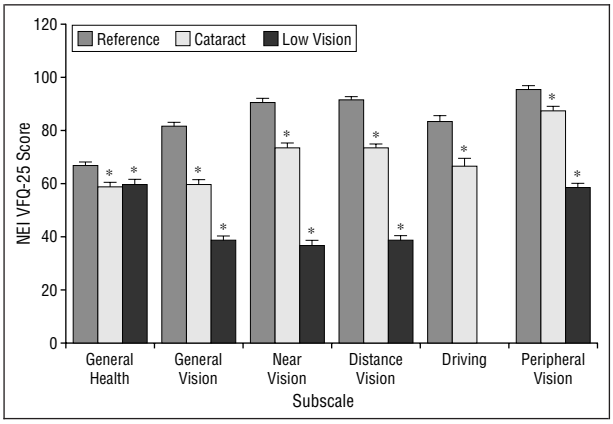
Table 4. Twenty-five–Item National Eye Institute Visual Function Questionnaire Subscale Scores by Condition for National Eye Institute Visual Function Questionnaire Field Test Sample (n = 597)*

Subscales	Diabetic Retinopathy (n = 123)	Age-Related Macular Degeneration (n = 108)	Glaucoma (n = 77)	Cataract (n = 93)	Cytomegalovirus Retinitis (n = 37)	Low Vision (n = 90)	Reference (n = 122)
General health	46 ± 25	65 ± 25†‡	62 ± 25‡	55 ± 25	45 ± 24	57 ± 27	69 ± 24
General vision	62 ± 21	53 ± 20	71 ± 17	60 ± 17	76 ± 14‡	38 ± 18	83 ± 15
Near vision	63 ± 30	54 ± 27	79 ± 23	73 ± 21	84 ± 20‡	36 ± 23	92 ± 13
Distance vision	66 ± 30	56 ± 29	77 ± 25	73 ± 22	84 ± 18	38 ± 26	93 ± 11
Driving	55 ± 40	39 ± 36	75 ± 28	63 ± 30	80 ± 28†‡	10 ± 23	87 ± 18
Peripheral vision	78 ± 29	77 ± 27	76 ± 27	87 ± 21	78 ± 21	59 ± 32	97 ± 10
Color vision	90 ± 22	85 ± 25	93 ± 17	90 ± 20	98 ± 9†‡	71 ± 31	98 ± 8
Ocular pain	88 ± 17†‡	87 ± 16†	89 ± 14†‡	86 ± 19‡	90 ± 16†‡	85 ± 20‡	90 ± 15
Vision specific							
Role difficulties	69 ± 31	61 ± 31	84 ± 23	76 ± 22	78 ± 24	44 ± 29	93 ± 13
Dependency	77 ± 30	72 ± 30	92 ± 19	88 ± 20	89 ± 12	51 ± 31	99 ± 6
Social functioning	81 ± 26	73 ± 29	89 ± 20	87 ± 19	96 ± 9	50 ± 31	99 ± 3
Mental health	66 ± 29	58 ± 27	81 ± 20	77 ± 22	74 ± 22	46 ± 27	92 ± 12

*Data are presented as mean ± SD. All pairwise comparisons between each disease group and the reference group were statistically significant at $P \leq .05$ unless otherwise specified. Fifty-three low-vision participants share diagnoses with other causes; hence, the total of the columns is more than 597 in the field test sample.

†Unadjusted *t* test comparison with reference group participants was nonsignificant ($P < .05$).

‡Linear regression result for 2-group comparison with the reference group, adjusted for age, sex, race, and medical comorbidities, was nonsignificant ($P < .05$).



Comparison of low-vision (n=90) and cataract (n=93) patients with reference patients (n=122) on mean 25-item version of the National Eye Institute Visual Function Questionnaire (NEI VFQ-25) subscale scores (field test sample only). Linear regression results for 2-group comparisons with the reference group, adjusted for age, sex, race, and medical comorbidities. The comparison of the low-vision group with the reference group on the driving subscale is not included because of a sample size of 12 for the low-vision group. Also, the sample size for the cataract group for the driving scale is 68. Asterisk indicates that all comparisons with the reference group were statistically significant at $P < .001$, except for general health and cataract ($P < .04$), general health and low vision ($P < .03$), and peripheral vision and cataract ($P < .002$); error bars, SEM.

Table 5. Pearson Correlations Between the 25-Item National Eye Institute Visual Function Questionnaire and Clinical Indicators of Visual Function (Field Test Sample Only)*

Subscale	Visual Acuity (n = 597)		AGIS Visual Field (n = 60)	
	Better Eye	Worse Eye	Better Eye	Worse Eye
General health	**0.11**	**0.14**	–0.13†	0.002
General vision	**0.67**	**0.65**	**–0.29**	–0.22
Near vision	**0.69**	**0.68**	**–0.40**	**–0.36**
Distance vision	**0.66**	**0.66**	**–0.42**	**–0.40**
Driving	**0.69**	**0.70**	–0.24	–0.26
Peripheral vision	**0.47**	**0.45**	**–0.25**	–0.23
Color vision	**0.51**	**0.39**	–0.15	–0.23
Ocular pain	0.11	0.06	–0.06	–0.05
Vision specific				
Role limitations	**0.56**	**0.56**	–0.23	**–0.25**
Dependency	**0.63**	**0.57**	**–0.51**	**–0.35**
Social functioning	**0.69**	**0.60**	**–0.47**	**–0.33**
Mental health	**0.52**	**0.56**	**–0.36**	**–0.29**
25-Item composite	**0.72**	**0.69**	**–0.41**	**–0.37**

*All values in boldface type are significant at $P \leq .05$. AGIS indicates Advanced Glaucoma Intervention Study.

†Correlations are negative because higher AGIS scores represent greater field loss.[21]

shown in Table 4 with comparisons of condition-specific mean scores using the field test sample.

POWER ESTIMATES

Power calculations based on observed between-eye condition differences in NEI VFQ-25 scores from the cross-sectional data collected in the field test study are presented in **Tables 6**, **7**, and **8** to assist with planning future research studies with sample size estimation.[23] Tables 6 and 7 should be used for designing randomized, experimental studies to maximize drawing strong causal inferences. Table 8 (self-selected groups) should be used to estimate statistical power for situations when it is not possible to randomize people to groups. To avoid a type I error due to multiple testing across the subscales, we advise researchers to develop hypotheses for which dimensions of vision are most likely to be affected by a given intervention and to conduct preplanned tests of associations for those specific subscales.

COMMENT

In this report we present the methods used to create the NEI VFQ-25 and provide evidence supporting its reli-

Table 6. Sample Sizes Needed per Group to Detect Differences in Changes Over Time Between 2 Experimental Groups for the 25-Item National Eye Institute Visual Function Questionnaire (NEI VFQ-25) Repeated-Measures Design*

		No. of Points Difference			
NEI VFQ-25 Subscale	SD	2	5	10	20
General health	26	1696	271	68	17
General vision	21	1106	177	44	11
Ocular pain	17	725	116	29	7
Near activities	29	2110	338	84	21
Distance activities	29	2110	338	84	21
Social functioning	27	1829	293	73	18
Mental health	27	1829	293	73	18
Role difficulties	29	2110	338	84	21
Dependency	28	1967	315	79	20
Driving	35	3073	492	123	31
Color vision	23	1327	212	53	13
Peripheral vision	27	1829	293	73	18
NEI VFQ-25 composite	20	1004	161	40	10

*Subscales are all scored on 0 to 100 possible range. Estimates assume $\alpha = .05$, 2-tailed *t* test, power of 80%, and an intertemporal correlation between scores of 0.60.

Table 7. Sample Sizes Needed per Group to Detect Differences Between 2 Experimental Groups for the 25-Item National Eye Institute Visual Function Questionnaire (NEI VFQ-25) Postintervention Measures Only*

		No. of Points Difference			
NEI VFQ-25 Subscale	SD	2	5	10	20
General health	26	2650	424	106	26
General vision	21	1729	277	69	17
Ocular pain	17	1133	181	45	11
Near activities	29	3297	527	132	33
Distance activities	29	3297	527	132	33
Social functioning	27	2858	457	114	29
Mental health	27	2858	457	114	29
Role difficulties	29	3297	527	132	33
Dependency	28	3073	492	123	31
Driving	35	4802	768	192	48
Color vision	23	2074	332	83	21
Peripheral vision	27	2858	457	114	29
NEI VFQ-25 composite	20	1568	251	63	16

*Subscales are all scored on 0 to 100 possible range. Estimates assume $\alpha = .05$, 2-tailed *t* test, and power of 80%.

Table 8. Sample Sizes Needed per Group to Detect Differences Between 2 Self-selected Groups for the 25-Item National Eye Institute Visual Function Questionnaire (NEI VFQ-25) Repeated-Measures Design*

		No. of Points Difference			
NEI VFQ-25 Subscale	SD	2	5	10	20
General health	26	2120	339	85	21
General vision	21	1383	221	55	14
Ocular pain	17	906	145	36	9
Near activities	29	2637	422	105	26
Distance activities	29	2637	422	105	26
Social functioning	27	2286	366	91	23
Mental health	27	2286	366	91	23
Role difficulties	29	2637	422	105	26
Dependency	28	2459	393	98	25
Driving	35	3842	615	154	38
Color vision	23	1659	265	66	17
Peripheral vision	27	2286	366	91	23
NEI VFQ-25 composite	20	1254	201	50	13

*Subscales are all scored on 0 to 100 possible range. Estimates assume $\alpha = .05$, 2-tailed *t* test, power of 80%, and an intertemporal correlation between scores of 0.60.

ability and validity across multiple chronic eye conditions. Considerations in selecting items for the NEI VFQ-25 included missing data rates, the distribution of the item-level responses, content validity considerations, and linear regressions that identified items that accounted for the maximum variability in the long-form subscale scores.

Short-form versions of any survey trade details about content and measurement precision for increased efficiency. To minimize the loss of content, we used methods that were likely to eliminate redundant or incomplete information rather than discard questions that provided unique information about vision-targeted HRQOL. The reported tests of reliability and validity indicate that although the NEI VFQ-25 is half as long as the 51-item version of the NEI VFQ, its psychometric properties are similar. In addition, the large proportion of variance in the long-form subscales explained by the smaller set of items indicates that much of the original content is retained in the shorter measure. Results of the combined sample, random split-half analysis of the data set provide evidence for the robustness of the selected items over a wide range of patients with 1 or more chronic eye conditions.

The evidence for validity of the NEI VFQ-25 includes comparisons of subscale-specific scores from the short- and long-form versions of the measure, between-group differences in NEI VFQ-25 scores for persons with different eye diseases of varying severity, and the moderate-to-high correlations between the NEI VFQ-25 subscales that have the most to do with central vision and measured visual acuity. Specifically, this article demonstrates that the NEI VFQ-25 is sensitive to the influence of age-related cataract, macular degeneration, glaucomatous field loss, and low vision from any cause. The correlations with clinical markers of disease severity provide evidence of clinical validity for the measure.

There are many surveys in the literature that assess difficulties with visual activities and/or condition-specific symptoms for persons with a wide variety of chronic eye diseases.[6-8,10,11,24] Although one may trade off sensitivity in a narrow domain uniquely affected by a specific condition, there are a number of reasons why one might select a measure such as the NEI VFQ-25, which is specific for persons with vision problems but not designed for one specific condition. These reasons include the empiric patient-driven basis for the content of the NEI VFQ-25[5]; the value added to any disease-specific study of being able to compare the relative burden of one condition with another on the same scale; the multidimensional nature of the NEI VFQ-25 subscales, which are designed to capture the impact of visual problems on

physical functioning, emotional well-being, and social functioning; and finally the rigorous multicondition evaluation of the reliability and validity of the NEI VFQ-25.

When interpreting this report, it is important to consider the following limitations. First, although persons across a large number of conditions and geographic regions were recruited for this study, to minimize the possibility of enrolling persons with false-positive diagnoses, the condition-specific enrollment criteria selected persons with moderate-to-severe disease. For this reason, we do not know whether the NEI VFQ-25 will be sensitive to the visual disability that is associated with earlier and milder forms of these or other ocular conditions. In addition, the NEI VFQ-25 estimates reported herein were derived from persons who completed the 51-item or the 96-item version of the measure. Although we suspect that the performance of the NEI VFQ-25 will be similar when these 25 items are administered alone, studies are currently under way to determine whether this is actually the case. It is important to also note that the estimates of statistical power reported in Tables 6, 7, and 8 are based on cross-sectional rather than longitudinal data. Further investigations are needed to establish the responsiveness of the NEI VFQ-25 in longitudinal studies.

Incorporating vision-targeted HRQOL measures into clinical studies will help form a more comprehensive understanding of treatment outcomes, not only in terms of benefits but also in terms of possible adverse effects. The need for a brief multidimensional vision-targeted HRQOL measure is illustrated by the fact that widespread use of the NEI VFQ-25 has preceded this report. To our knowledge, the NEI VFQ-25 has been translated into 8 languages and is currently being used in at least 7 federally funded research studies that are examining a range of ocular conditions. The evidence presented in this report illustrates the potentially useful information one might expect to gain with a relatively brief survey. Researchers are encouraged to use the NEI VFQ-25 to examine the influence various eye diseases and interventions have on a patient's day-to-day functioning and well-being.

Accepted for publication November 6, 2000.

This work was supported by the NEI, Bethesda, Md (contract C950424), and the Research Division of Merck Pharmaceuticals, Whitehouse Station, NJ. Dr Mangione is a recipient of a Clinical Investigator Award (1K08-AG00605) from the National Institute on Aging, Bethesda, Md, and a Generalist Faculty Scholar's Award from the Robert Wood Johnson Foundation, Princeton, NJ (029250). Dr Lee is a recipient of a Lew Wasserman Merit Award from Research to Prevent Blindness, New York, NY.

We would like to acknowledge and thank Leon Ellwein, PhD, Frederick Ferris III, MD, Argye Hillis, PhD, and Donald Patrick, PhD, for their methodological guidance and participation on the Scientific Advisory Panel for the NEI VFQ Psychometric Field Test.

NEI VFQ Field Test Investigators are as follows: Jonathan C. Javitt, MD, Department of Ophthalmology, The Johns Hopkins University, Baltimore, Md; Fang Wang, MD, PhD, Pfizer, Inc, Cambridge, Mass; Julian Nussbaum, MD, Rhett Schiffman, MD, Henry Ford Health Systems, Detroit, Mich; Emily Chew, MD, NEI, National Institutes of Health, Bethesda, Md; Janet DeBerry Steinberg, OD, Scheie Eye Institute, University of Pennsylvania, Philadelphia; Cynthia Owsley, PhD, Department of Ophthalmology, University of Alabama at Birmingham; Nancy K. Janz, PhD, School of Public Health, University of Michigan, Ann Arbor; Ron Klein, MD, Department of Ophthalmology, University of Wisconsin, Madison; Amy Chomsky, MD, Denis O'Day, MD, Department of Ophthalmology, Vanderbilt School of Medicine, Nashville, Tenn; Mae Gordon, PhD, Anthony Lubniewski, MD, Department of Ophthalmology, Washington University Medical School, St Louis, Mo.

Corresponding author: Carol M. Mangione, MD, MSPH, Division of General Internal Medicine and Health Services Research, Department of Medicine, UCLA, 911 Broxton Plaza, Box 951736, Los Angeles, CA 90095-1736.

REFERENCES

1. Marshall GN, Hays RD. *The Patient Satisfaction Questionnaire Short-Form (PSQ-18).* Santa Monica, Calif: RAND; 1994. P-7865-RC.
2. McHorney CA, Ware JE. Construction and validation of an alternate form general mental health scale for the Medical Outcomes Study Short-Form 36-Item Health Survey. *Med Care.* 1995;33:15-28.
3. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. *Health Econ.* 1993;2:217-227.
4. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996; 34:220-233.
5. Mangione CM, Lee PP, Pitts J, et al. Psychometric properties of the National Eye Institute Visual Function Questionnaire (NEI VFQ). *Arch Ophthalmol.* 1998;116: 1496-1504.
6. Bernth-Petersen P. Visual functioning in cataract patients: methods for measuring and results. *Acta Ophthalmol (Copenh).* 1981;59:198-205.
7. Mangione CM, Phillips RS, Seddon JM, et al. Development of the "Activities of Daily Vision Scale": a measure of visual functional status. *Med Care.* 1992;30: 1111-1126.
8. Lundstrom M, Fregell G, Sjoblom A. Vision related daily life problems in patients waiting for cataract extraction. *Br J Ophthalmol.* 1994;78:608-611.
9. Javitt JC, Brenner MH, Curbow B, et al. Outcomes of cataract surgery improvement in visual acuity and subjective visual function after surgery in the first, second, and both eyes. *Arch Ophthalmol.* 1993;111:686-691.
10. Steinberg EP, Tielsch JM, Schein OD, et al. The VF-14: an index of functional impairment in patients with cataract. *Arch Ophthalmol.* 1994;112:630-638.
11. Sloane ME, Ball K, Owsley C, et al. The Visual Activities Questionnaire: developing an instrument for assessing problems in everyday visual tasks. *Tech Dig Noninvasive Assess Vis Sys.* 1992;1:26-29.
12. Mangione CM, Berry S, Lee PP, et al. Identifying the content area for the National Eye Institute Vision Function Questionnaire (NEI VFQ): results from focus groups with visually impaired persons. *Arch Ophthalmol.* 1998;116:227-238.
13. Folstein MF, Folstein SE, McHugh PR. The Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975; 12:189-198.
14. Stewart AL, Greenfield S, Hays RD, et al. Functional status and well-being of patients with chronic conditions: results from the Medical Outcomes Study. *JAMA.* 1989;262:907-913.
15. *National Eye Institute Visual Function Questionnaire (NEI VFQ) Study: Phase II Field Test: Manual of Procedures.* Los Angeles, Calif: NEI VFQ Field Test Coordinating Center, University of California, Los Angeles (UCLA); 2000. Available from: National Technical Information Service, 5285 Port Royal Rd, Springfield, VA 22161; Accession No. PB2000-102119.
16. Ferris FL, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. *Am J Ophthalmol.* 1982;94:91-96.
17. EMMES Corp. *Phase II Manual of Operations.* Potomac, Md: EMMES Corp; December 11, 1997.
18. Sbeskin D. *Handbook of Parametric and Nonparametric Statistical Procedures.* Boca Raton, Fla: CRC Press; 1996:20.
19. Hocking RR. The analysis and selection of variables in linear regression. *Biometrics.* 1976;32:1-50.
20. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297-334.
21. Nunnally JC. *Psychometric Theory.* New York, NY: McGraw-Hill Inc; 1978:190-255.
22. Advanced Glaucoma Intervention Study Investigators. Advanced Glaucoma Intervention Study 2: visual field test scoring and reliability. *Ophthalmology.* 1994; 101:1589-1595.
23. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates Inc; 1988.
24. Lee BL, Gutierrez P, Wilson MR, et al. The Glaucoma Symptom Scale (GSS): a brief index of glaucoma-specific symptoms. *Arch Ophthalmol.* 1998;116:861-866.