The Michigan Retinal Degeneration Questionnaire: A Patient-Reported Outcome Instrument for Inherited Retinal Degenerations

GABRIELLE D. LACY, MARIA FERNANDA ABALEM, CHRIS A. ANDREWS, LILIA T. POPOVA, ERIN P. SANTOS, GINA YU, HANAN Y. RAKINE, NATASHA BAIG, JOSHUA R. EHRLICH, ABIGAIL T. FAHIM, KARI H. BRANHAM, JOAN A. STELMACK, BONNIELIN K. SWENOR, GISLIN DAGNELIE, DAVID C. MUSCH, AND K. THIRAN JAYASUNDERA

• PURPOSE: To create a psychometrically validated patient-reported outcome measure for inherited retinal degenerations.

• DESIGN: Qualitative and quantitative patient-reported outcome (PROs) questionnaire development using item response theory validation.

• METHODS: One hundred twenty-eight patients with a diagnosis of an inherited retinal degeneration at the Kellogg Eye Center (University of Michigan) were recruited and administered a 166-item questionnaire comprising 7 expert-defined domains. The questionnaire was re-administered 4-16 days later to a subset of 25 participants to assess test-retest variability. Graded response models were fit by Cai's Metropolis-Hastings Robbins-Monro algorithm using the R (version 3.6.3) package mirt. Model data were fit to assess questionnaire dimensionality, to estimate item information, and to score participants. Poorly functioning items were removed, and the model was refit to create the final questionnaire.

• RESULTS: The psychometrically validated PROs measure was reduced to a 59-item questionnaire measuring 7 unidimesnional domains: central vision, color vision, contrast sensitivity, scotopic function, photopic peripheral vision, mesopic peripheral vision, and photosensitivity. A total of 39 items were removed because of poor factor loading, low item information, poor person-ability differentiation, or high item-level interdependence. This novel questionnaire produces a reliable domain score for person ability that does not show significant test-retest variability across repeated administration.

• CONCLUSIONS: The final PRO questionnaire, known as the Michigan Retinal Degeneration Questionnaire, is psychometrically validated and available for use in the evaluation of patients with inherited retinal degenerations. (Am J Ophthalmol 2021;222:60–68. © 2020 Elsevier Inc. All rights reserved.)

RESEARCH IN INHERITED RETINAL DEGENERATIONS (IRDs) has dramatically enhanced our understanding of genetic characterization, disease progression, and therapeutic possibilities. Clinical trials targeting IRD populations have focused on visual function and structural measures such as microperimetry, optical coherence tomography, and electroretinography to detect treatment efficacy.[1] However, without understanding the patient's experience of visual function improvement, we risk failing to capture the most meaningful measure of treatment efficacy.[2] Although treatment signals have been detected on clinical visual/retinal function testing,[3,4] it remains unclear if these changes are sufficient to impact vision-dependent functioning in a patient's daily life. A compelling need remains for reliable and validated patient-reported outcome (PRO) measures for IRD clinical trials to (1) capture previously unmeasured signals of treatment efficacy and (2) understand how improvement in standard clinical tests is associated with patient-experienced treatment benefits.

PRO measures are recognized as clinical trial outcome measures that capture the patient perspective.[5,6] Although well-constructed and validated ophthalmic PRO instruments, such as the National Eye Institute Visual Function Questionnaire (VFQ-25), have been applied to IRD clinical trials,[7–9] these PRO measures were not designed for capturing the unique functional, emotional, and other domains that pertain to an IRD population.[2] Guidelines for PRO generation require a high-quality PRO to be created based on qualitative insights and feedback from the intended population.[6,10] Furthermore, the PRO must be tested and statistically validated in a representative sample population.[5,6]

AJO.com Supplemental Material available at AJO.com.

Accepted for publication Aug 20, 2020.

From the Kellogg Eye Center, Department of Ophthalmology and Visual Sciences, University of Michigan Medical School, Ann Arbor, Michigan, USA (G.D.L., M.F.A., C.A.A., L.T.P., E.P.S., G.Y., H.Y.R., N.B., J.R.E., A.T.F., K.H.B., D.C.M., K.T.J.), Department of Ophthalmology and Otolaryngology, University of Sao Paulo Medical School, São Paulo, Brazil (M.F.A.), Department of Low Vision Service, Illinois Eye and Ear Infirmary, University of Illinois College of Medicine, Chicago, Illinois (J.A.S.), Johns Hopkins University Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland (B.K.S., G.D.), and Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan (D.C.M.), USA.

Inquiries to K. Thiran Jayasundera, Retinal Dystrophy Clinic, Kellogg Eye Center, University of Michigan, 1000 Wall Street, Ann Arbor, MI 48105, USA; e-mail: thiran@umich.edu

0002-9394/$36.00
https://doi.org/10.1016/j.ajo.2020.08.032

The current article presents the validation of the Michigan Retinal Degeneration Questionnaire (MRDQ), a PRO measure for use in IRD therapeutic trials. The MRDQ has been designed in accordance with U.S. Food and Drug Administration guidelines to measure visual function in the context of the daily life of patients with an IRD and to detect treatment response in IRD clinical trials.

METHODS

APPROVAL FROM THE UNIVERSITY OF MICHIGAN INSTITUtional Review Board (HUM00115127) was obtained before the study, and the research was performed in accordance with the Declaration of Helsinki.

• **PATIENTS:** Adult patients from the Kellogg Eye Center (University of Michigan) Retinal Dystrophy Clinic were recruited from December 2016 to March 2020. Participants had a clinical diagnosis of either rod-cone dystrophy, cone/cone-rod dystrophy, or macular dystrophy. Clinical diagnoses were confirmed by 2 fellowship-trained IRD specialists (K.T.J., A.T.F.) with testing including electroretinography, optical coherence tomography, Goldmann visual field testing, fundus autofluorescence, and clinical examination. Participants gave informed consent at the time of a routine clinical visit. Patients were excluded if lacking sufficient command of the English language to understand and provide informed consent. Clinical records including visual acuity, visual fields, electroretinography, gene testing, and ophthalmic medical history were collected from the electronic health record for analysis.

• **PHASE 1: ITEM GENERATION:** The preliminary MRDQ was derived through an iterative process involving qualitative analysis to demonstrate content validity.[6,11] Initial content items for the MRDQ were drafted through a previously described[12] series of expert panel focus groups, in-depth patient interviews, cognitive interviews, and pilot questionnaire administration. Patients participated in open-ended, in-depth interviews to solicit patient perspectives on symptoms they related to their visual condition. Transcribed interviews were coded and analyzed in Atlas.ti software (Version 8.1.3 (522); Atlas.ti, Berlin, Germany), and initial content items were drafted based on grounded theory principles of theme extraction.[12] An additional group of patients participated in cognitive and pilot interviews to provide feedback and refine the questionnaire.

• **PHASE 2: ADMINISTRATION:** The pilot MRDQ questionnaire items were administered orally over approximately 35-50 minutes either in-person or over the phone. Participants were interviewed by clinical research assistants, and the MRDQ was readministered via consecutive sampling 4-16 days after the initial interview to assess test-retest variability.

• **PHASE 3: PSYCHOMETRICS ANALYSIS AND ITEM REDUCTION:** *Factor and model fit analysis*. Unidimensionality and local independence of each domain was investigated by factor analysis and analysis of linear dependence of residuals. Items loading on a second dimension having high residual dependence with another item, with low information, and/or with overlapping probability traces were considered for removal. Models were refit after item reduction.

Graded response model. Graded response models were built for 7 domains identified before questionnaire administration based on in-depth patient interviews and expert knowledge: central vision, color vision, contrast sensitivity, scotopic function, photopic peripheral vision, mesopic peripheral vision, and photosensitivity. Models were fit by Cai's Metropolis-Hastings Robbins-Monro algorithm[13] implemented in the R (version 3.6.3) package mirt.[14] A graded response model has 1 parameter for each person (θ) that quantifies his or her disability on the domain. The disabilities are centered at 0. Higher person-scores indicate greater disability. A graded response model for ordinal responses has k parameters for an item with k response categories: 1 discrimination parameter (2PL model) and $k - 1$ difficulty thresholds dividing adjacent responses. For comparison, we additionally used the Method of Successive Dichotomizations to estimate the latent disability measure.[15]

Difficulty, fit, and test information. Item probability trace and item information curves were used to assess an item's suitability for a domain. Item probability trace functions identify the likelihood of each response for each person score (θ). Item information curves describe the relative contribution of each item to determining an overall person score (θ).[16–18] Test information curves, standard error functions, and marginal reliabilities[19] were used to measure each domain's performance in differentiating patient ability. Model fit was assessed using the standardized root mean square residual, root mean square error of approximation, comparative fit index, Tucker-Lewis Index, and Cai and Monro's C2.[20]

Differential item functioning was performed against 5 covariates: sex (male, female), age (<41, 41-61, >61 years), visual acuity (logMAR <0.14, 0.14-0.437, >0.438), IRD phenotype (rod-cone, cone, macular dystrophy), and Patient Health Questionnaire (PHQ-4) score (none, mild, moderate, severe)[21,22] for the remaining 59 items. The PHQ-4 is a validated screening tool for symptoms of depression and anxiety, scored on a 0-12 point scale. Each MRDQ item was considered separately for differential item functioning using the remaining items as anchors. The discrimination (uniform) and difficulty parameters (nonuniform) were allowed to differ by covariate category for each targeted item. The presence of differential item functioning was determined by P values from

likelihood ratio tests using Bonferroni-corrected alpha. In addition, graded response and method of successive dichotomizations models for each domain were compared with likelihood ratio statistics.

Domain and trait associations. Domains were assessed via simple linear models for association with the following traits: logMAR corrected visual acuity in better eye and worse eye, age, sex, and IRD phenotype. Patient covariates were summarized by counts and percentages or medians and ranges. Missing data were summarized by item and by respondent. Associations between domains were measured by Pearson correlations.

Test-retest variability. Test-retest reliability for each domain was quantified by the Pearson correlation between participants' abilities on first and second tests and by the mean and standard deviation of change between administrations.[23]

RESULTS

• **PHASE 1: ITEM GENERATION:** Fifty-five adult patients with IRDs participated in either in-depth interviews (n = 25) until item generation was exhausted, or cognitive and pilot interviews (n = 30) for further refinement.[12] There were no significant differences in visual acuity or visual fields among these groups. The resulting draft of the MRDQ contained 103 items organized into 4 conceptual domains pertaining to visual function.

• **PHASE 2: ADMINISTRATION:** One hundred twenty-eight participants with a clinically diagnosed IRD completed the questionnaire. Twenty-five participants completed the MRDQ a second time for test-retest variability measurement. Table 1 shows the demographics and characteristics of study participants.

• **PHASE 3: PSYCHOMETRICS ANALYSIS AND ITEM REDUCTION:** All items considered for analysis were answered by at least 85% of participants. No items were removed because of high levels of missingness. Table 2 shows the 39 items removed because of poor factor loading, low item information, poor person-ability differentiation, or high item-level interdependence.

Based on factor analysis, the following 7 unidimensional domains were confirmed: central vision, color vision, contrast sensitivity, scotopic function, photopic peripheral vision, mesopic peripheral vision, and photosensitivity. Within each domain, items that loaded predominantly onto a second dimension were considered for removal. A total of 20 items were removed at least in part due to high loading on a second dimension within a singular domain: central (2), color (0), contrast (2), scotopic (4), photopic peripheral (5), mesopic peripheral (5), and photosensitivity (2). Within each domain, items with high levels of item interdependence (greater than 0.3), after controlling for overall ability score, were evaluated for removal. Four items were removed at least in part due to high interdependence with another item: contrast (1), scotopic (1), photopic peripheral (1), and mesopic peripheral (1). Nine items with low person-ability discrimination on item probability trace functions were removed. Nineteen items with low contribution to domain information were removed. Graded response and method of successive dichotomization models were compared, and the more restrictive method of the successive dichotomization model was rejected for each domain ($P < .001$, all). No items were removed because of differential item functioning for covariates: sex, age, visual acuity, IRD phenotype, and PHQ-4 score.

TABLE 1. Demographic and Participant Characteristics

Total, n	128
Female, n (%)	65 (50.8)
Age (y), median (range)	49 (18-88)
IRD phenotype, n (%)	
Rod-cone	69 (53.9)
Cone/cone-rod	30 (23.4)
Macular	29 (22.7)
Conclusive genetic test result, n (%)	77 (60.2)
Race/ethnicity, n (%)	
White, non-Hispanic	107 (83.6)
Black/African American	10 (7.8)
Asian	3 (2.3)
Hispanic	5 (3.9)
Unknown	3 (2.3)
Corrected visual acuity, median (range)	
Better eye	20/42 (20/16—NLP)
Worse eye	20/60 (20/18—NLP)

IRD = inherited retinal dystrophies; NLP = no light perception.

• **DOMAIN AND TRAIT ASSOCIATIONS:** Table 3 illustrates the relationship of domain scores with logMAR corrected visual acuity, date of administration, age, sex, and IRD phenotype. Visual acuity correlated with all domains except photosensitivity. The date of administration was not significantly correlated with any domains. Age correlated with central and color vision domains, whereas sex was only associated with the photosensitivity domain. IRD phenotype correlated with photopic and mesopic peripheral vision.

• **TEST-RETEST VARIABILITY:** No significant difference was observed in the mean change in patient responses across repeat administrations obtained in a 4-16 day period. Dates of administration and readministration had no significant effect on patient scores. Table 4 shows the reliability of each domain, Pearson's correlation between the first and

TABLE 2. Original Items not Included in the Final Michigan Retinal Degeneration Questionnaire and Reasons for Exclusion

Domain	Item	Item Topic	Missing Responses (%)	Reason
Central	Q09	Screen settings	0	IPT, IIC
	Q13	Worry when reading	2.9	Factor
	Q14	Worry when distance reading	2.9	Factor
Color	Q13	Tinted filters	2.9	IPT, IIC
	Q14	Phone color features	2.9	IPT, IIC
	Q15	Worry distinguishing color	0	IPT, IIC
Contrast	Q07	Seeing steps	0	LD
	Q16	Worry identifying objects	0	Factor
	Q17	Worry recognizing faces	0	Factor
Scotopic	Q23	Scan to see	5.9	IPT, IIC
	Q32	Bump into objects	3.9	IIC, LD
	Q35	Avoid going out	2.0	IIC
	Q41	Use guide dog/cane	3.9	IIC
	Q44	Worry in familiar places	3.9	Factor
	Q47	Worry in unfamiliar places	9.8	Factor
	Q50	Worry bumping into objects	4.9	Factor
	Q53	Worry seeing steps	1.0	Factor
Photopic	Q22	Scan to see	2.0	IPT, IIC
	Q31	Bump into objects	1.0	IIC, LD
	Q34	Avoid going out	2.9	IIC
	Q40	Use guide dog/cane	2.9	IIC
	Q43	Worry in familiar places	2.0	Factor
	Q46	Worry in unfamiliar places	5.9	Factor
	Q49	Worry bumping into objects	1.0	Factor
	Q52	Worry seeing steps	1.0	Factor
	Q55	Worry seeing uneven ground	1.0	Factor
Mesopic	Q24	Scan to see	3.9	IIC, LD
	Q33	Bump into objects	2.9	IIC
	Q36	Avoid going out	3.9	IIC
	Q42	Use guide dog/cane	2.9	Factor
	Q45	Worry in familiar places	3.9	Factor
	Q48	Worry in unfamiliar places	10.8	Factor
	Q51	Worry bumping into objects	2.9	Factor
	Q54	Worry seeing steps	2.0	Factor
	Q57	Worry seeing uneven ground	2.9	IPT, IIC
Photosensitivity	Q04	Screen sensitivity	2.0	Factor
	Q08	Adjust screen	4.9	Factor, IIC
	Q10	Worry going out	1.0	IPT, IIC
	Q11	Worry about bright screens	6.9	IPT, IIC

Item did not load on main domain factor (Factor); item had low information content (IIC); item had highly overlapping probability traces (IPT); item had high linearly dependence with another item (LD). Data are the amount of missing responses for each item, n (%).

second administrations, and average change between administrations.

• **MICHIGAN RETINAL DEGENERATION QUESTIONNAIRE:** Figure 1 illustrates the distribution of participant scores across each finalized domain (after item reduction) in a person-item map. After item reduction, the final MRDQ contains 59 items pertaining to central vision (11), color vision (4), contrast sensitivity (7), scotopic function (12), photopic peripheral vision (9), mesopic peripheral vision (9), and photosensitivity (7). The total information curve and standard error identified in each domain are shown in Supplemental Figure 1. Supplemental Tables 1-4 include the correlation of domain scores, item and model fit parameters, and final MRDQ items.

DISCUSSION

IN THIS STUDY, THE MRDQ WAS PSYCHOMETRICALLY VALIdated in patients with IRDs using item response theory techniques. The MRDQ generates an ability score from

TABLE 3. Associations Between Domain Scores (θ) and Participant Characteristics Measured by Adjusted R^2 of the Linear Model and by the *P* Value of the *F* Test of no Association

	Corrected VA					
Domain (θ)	Better Eye	Worse Eye	Date of MRDQ	Age	Sex	IRD Phenotype
Central	61.8 (<.001)	48.6 (<.001)	1.5 (.214)	5.5 (.018)	0.0 (.980)	1.6 (.658)
Color	37.3 (<.001)	27.8 (<.001)	0.1 (.720)	7.4 (.006)	0.2 (.639)	4.9 (.173)
Contrast	31.3 (<.001)	28.7 (<.001)	0.1 (.728)	5.3 (.020)	0.3 (.604)	4.5 (.206)
Scotopic	12.2 (<.001)	11.5 (.001)	0.2 (.667)	0.4 (.527)	1.8 (.183)	24.7 (<.001)
Photopic peripheral	25.0 (<.001)	23.8 (<.001)	0.2 (.671)	2.2 (.140)	0.2 (.688)	17.8 (<.001)
Mesopic peripheral	14.2 (<.001)	16.4 (<.001)	0.3 (.590)	3.0 (.085)	1.6 (.213)	23.8 (<.001)
Photosensitivity	3.40 (.065)	1.5 (.228)	0.0 (.839)	1.7 (.192)	6.7 (.009)	6.1 (.104)

IRD = inherited retinal degeneration; MRDQ = Michigan Retinal Degeneration Questionnaire.
Corrected logMAR visual acuity (VA) taken at the time of the most recent clinical visit.

TABLE 4. Domain Score (θ) Test-Retest Reliability

			Test-Retest (95% CI) (n = 25)		
Domain (θ)	No. of Questions	Marginal Reliability (n = 128)	ρ Correlation	Mean Change	SD ME
Central	11	0.94	0.95 (0.88, 0.98)	−0.02 (−0.16, 0.11)	0.23 (0.13, 0.32)
Color	4	0.84	0.84 (0.66, 0.93)	0.23 (0.06, 0.40)	0.29 0.20, 0.38)
Contrast	7	0.88	0.91 (0.80, 0.96)	−0.18 (−0.37, 0.00)	0.32 (0.14, 0.50)
Scotopic function	12	0.96	0.92 (0.82, 0.96)	−0.01 (−0.16, 0.15)	0.27 (0.19, 0.35)
Photopic peripheral	9	0.87	0.91 (0.80, 0.96)	−0.05 (−0.22, 0.11)	0.29 (0.21, 0.36)
Mesopic peripheral	9	0.94	0.88 (0.75, 0.95)	0.00 (−0.19, 0.19)	0.33 (0.23, 0.42)
Photosensitivity	7	0.90	0.91 (0.79, 0.96)	−0.06 (−0.23, 0.12)	0.30 (0.22, 0.38)

SD ME = standard deviation of measurement error.
Marginal reliability estimated from original 128 participants. Three test-retest statistics (Pearson correlation, mean difference, and standard deviation of measurement error) and their 95% confidence intervals were computed from 25 pairs of tests taken approximately 2 weeks apart.

patient-reported visual function in domains representative of physiological visual function pathways. Notably, the presented 95% confidence interval for the test-retest variability of domain score (θ) enables an investigator to determine an efficacy signal in an individual patient after a therapeutic intervention.

Recent trends have caused a shift away from validation of PROs using classical test theory and toward applying item response theory analysis, namely Rasch psychometric techniques.[24,25] Although Rasch analysis has many strengths, the graded response model was determined to be the appropriate item response theory analysis for the MRDQ, given the restrictive requirements of the Rasch approach. A graded response model has the advantage of fitting a model based on the variance and distribution of patient responses, whereas a Rasch model requires uniform item discrimination.[26–28] The graded response model provides flexibility to fitting items of different levels of difficulty and person-ability discrimination within a domain measuring the same underlying trait.[16–18,29] Food and Drug Administration guidelines for creating a PRO measure do not require the use of a particular validation technique; the graded response model has been used in other ophthalmic PROs and is included within the PRO Measurement Information System initiative guidelines.[26,28,30]

In graded response model analysis, the overall person score (θ) is the metric for person-level ability of a particular domain/trait. This score aggregates the information provided by each PRO item within a unidimensional domain into a summarized score for the functional ability of participants in the measured trait.[16,17] In this analysis, θ is centered at the mean trait level of the population, and extreme values for θ (ie, −3, +3) are indicative of low or high visual dysfunction based on item responses. Each item's relative contribution to the overall domain score is an aggregate of the information provided by the item and the discrimination ability of the item. For the MRDQ, the flexibility of accommodating different domain item discrimination is a notable strength over the Rasch

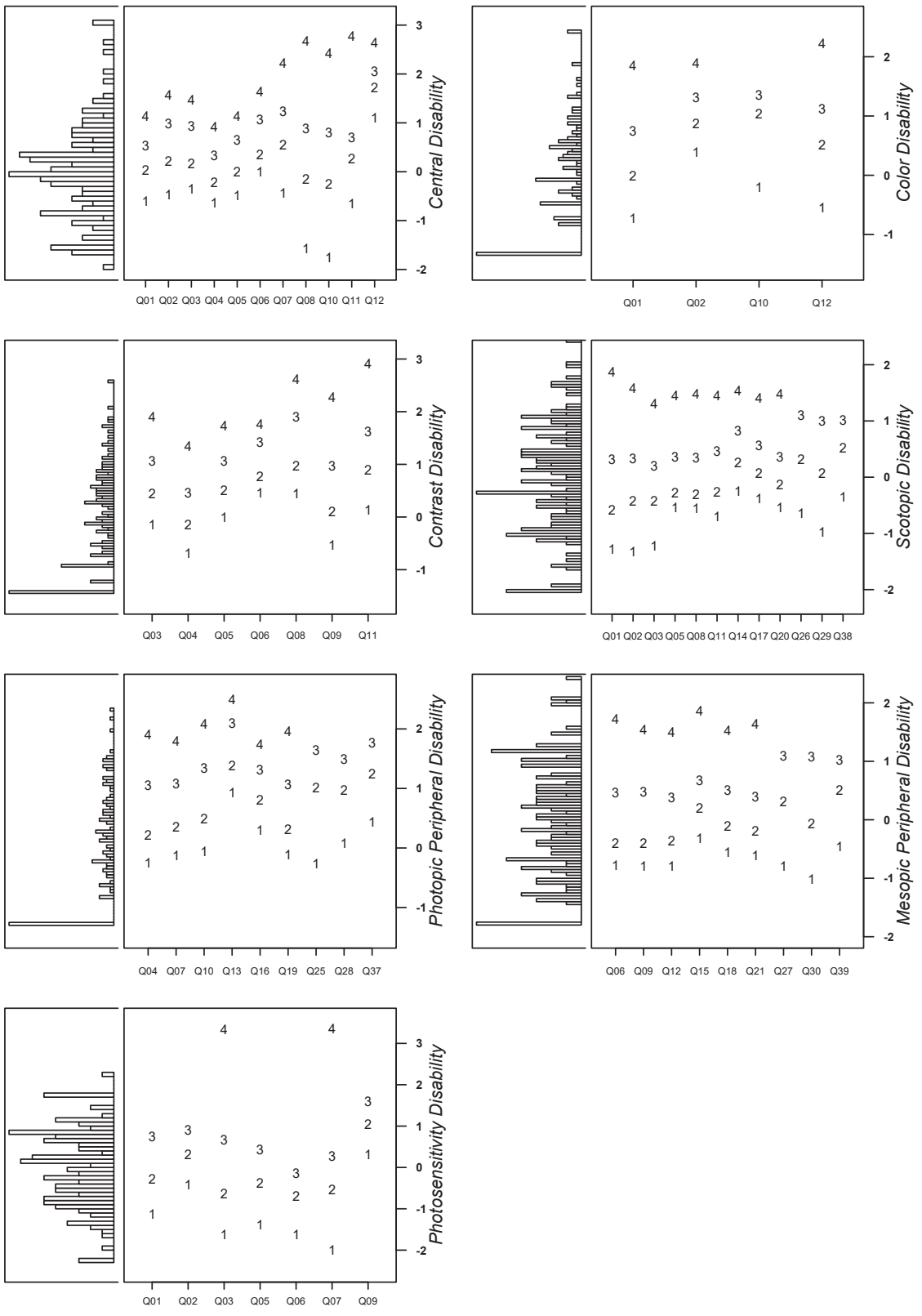

FIGURE 1. Person-item map for each domain.

approach, which requires each domain item to have equal weight in the domain score. This technique allows more informative PRO items to have a greater influence on the overall score.

Test reliability and variability of repeat measures are a significant challenge for IRD populations.[2] In addition to a domain score, the standard deviation (and 95% confidence interval) is given such that future clinical trials are able to interpret a true change in functional ability beyond test-retest variability (Table 4). Furthermore, the design of the MRDQ includes anchoring points that further aid in the reliability of the measure. Rather than a numerical response choice, the MRDQ responses are defined by patient-derived language. When asked about a level of difficulty, patients are given the following precise response choices: "None: I do not have trouble with this," "A little difficulty: I notice a problem, but I do not struggle," "Moderate difficulty: I struggle but I can still do this," "Extreme difficulty: I struggle a lot and sometimes I cannot do this," and "N/A for non-vision reasons: I do not do this." These defining statements are consistent throughout the PRO to reduce cognitive burden while still establishing a clear response rubric for patients and interviewers.

Furthermore, the MRDQ is able to measure patient-reported deficits irrespective of covariates such as age, sex, IRD phenotype, visual acuity, and PHQ-4 score. When analysis is performed at the level of each item, using differential item functioning analysis, no questions were removed because of participant characteristics that predict an individual's response after accounting for person-level disability.

Although the MRDQ has several strengths, we acknowledge that there are remaining limitations in this PRO instrument. As the MRDQ is intended to be applied as a clinical trial outcome measure, the latent traits of this measure are focused on visual function and do not directly address how patients perceive their vision to influence other domains of their quality of life. Although we acknowledge that understanding a patient's overall quality of life is valuable, other existing PRO measures[31] are equipped to capture this. In addition, as with many PRO measures, the ability to distinguish patients at the extremes of ability is a challenge. In this case, the MRDQ is limited when differentiating individuals with very low levels of disability (ie, fairly good vision). Considering that the population of patients with IRDs who are eligible for clinical trials are unlikely to be asymptomatic, the differentiability of patients on this end of the spectrum is of low concern. Given that the population in this study is representative of potential clinical trial candidates, the PRO is well equipped to differentiate patients with symptomatic IRDs. Although the MRDQ may also be relevant and appropriate for application in routine clinical care and low-vision rehabilitation, further validation may be be necessary in these settings.

The authors acknowledge that the study population was not evenly distributed among racial/ethnic groups or IRD phenotypes. The demographic of the study group represents the IRD patient population at a particular academic institution. Further investigation may be necessary when considering the application of this PRO measure to diverse cultural and language contexts. Although the study population is not an even sample of IRD phenotypes, this is reflective of the general prevalence of phenotypes.[1] Given the over 270 currently identified number of IRD causing genes,[32] it would be impractical to recruit a sufficiently large sample population of each genetic diagnosis.

With the understanding that item function may depend on the specific IRD presentation, the authors thought that it was important to create a PRO measure encompassing physiologically differentiated visual function domains. Applying an item response theory model enables this measure to quantify the wide range of disability seen for individual patients in each domain. We recommend that for the study of a specific IRD, investigators administer the PRO in a subset of subjects likely to enroll in a particular clinical trial/study and the investigators use discretion in selecting MRDQ domains for their study (IRD-specific short form) based on the distribution of abilities/scores for the sample of the target population. In addition, we recognize that the PRO method of administration may influence responses; therefore, we also recommend a consistent administration method for all participants in a clinical trial/study.

The MRDQ has undergone content and psychometric validation in IRD populations and meets the standards of a clinical trial outcome measure following the guidelines established by the Food and Drug Administration.[6] With provided instructions, the MRDQ can be administered with no interviewer training and an anticipated completion time of 25 minutes. MRDQ domains pertain to visual function pathways and thereby are able to record a patient's response to emerging therapeutics for these domains by the use of θ from item response theory.

ALL AUTHORS HAVE COMPLETED AND SUBMITTED THE ICMJE FORM FOR DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST.
 Funding/Support: This research was supported by the National Institute of Health grants TL1TR002242 (G.D.L.), K23EY027848 (J.R.E.), K12EY022299 (A.T.F.), and K23EY026985 (K.T.J.). The authors would like to thank Lindsay Godsey, Adam R. Holmes, Milan K. Patel, and Naheed W. Khan for their contributions to this work. Author Contributions: G.D.L.: Conceptualization, methodology, investigation, writing, visualization, and project administration. M.F.A.: Conceptualization, methodology, investigation, writing, supervision, and project administration. C.A.A.: Methodology, formal analysis, investigation, writing, visualization, and funding acquisition. L.T.P., E.P.S., G.Y., H.Y.R., N.B.: Conceptualization and investigation. J.R.E.: Conceptualization, validation, and writing—review and editing. A.T.F., K.H.B.: Investigation and resources. J.A.S., B.K.S., G.D.: Validation and writing—review and editing. D.C.M.: Conceptualization, methodology, valida-

tion, and writing—review and editing. K.T.J.: Conceptualization, methodology, investigation, resources, writing, supervision, project administration, and funding acquisition. Financial Disclosures: C.A.A. and D.C.M. are consultants to Applied Genetics Technologies Corporation, who wish to make use of this patient-reported outcome (or a defined subset) for their clinical trials in planning. This manuscript has not been published elsewhere and has not been submitted simultaneously for publication elsewhere. All authors attest that they meet the current ICMJE criteria for authorship.

REFERENCES

1. Thompson DA, Ali RR, Banin E, et al. Advancing therapeutic strategies for inherited retinal degeneration: recommendations from the Monaciano Symposium. *Invest Ophthalmol Vis Sci* 2015;56:918–931.
2. Lacy GD, Abalem MF, Musch DC, Jayasundera KT. Patient-reported outcome measures in inherited retinal degeneration gene therapy trials. *Ophthalmic Genet* 2020;41:1–6.
3. MacLaren RE, Groppe M, Barnard AR, et al. Retinal gene therapy in patients with choroideremia: initial findings from a phase 1/2 clinical trial. *Lancet* 2014;383:1129–1137.
4. Lam BL, Davis JL, Gregori NZ, et al. Choroideremia gene therapy phase 2 clinical trial: 24-month results. *Am J Ophthalmol* 2019;197:65–73.
5. Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health* 2007;10(Suppl 2):S125–S137.
6. U.S. Department of Health and Human Services FaDA, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. Available at ; 2009. http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf;. Accessed August 10, 2020.
7. Russell S, Bennett J, Wellman JA, et al. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet* 2017;390:849–860.
8. Dimopoulos IS, Hoang SC, Radziwon A, et al. Two-year results after AAV2-mediated gene therapy for choroideremia: the Alberta experience. *Am J Ophthalmol* 2018;193:130–142.
9. Weleber RG, Pennesi ME, Wilson DJ, et al. Results at 2 years after gene therapy for RPE65-deficient Leber congenital amaurosis and severe early-childhood-onset retinal dystrophy. *Ophthalmology* 2016;123:1606–1620.
10. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. *Value Health* 2011;14:978–988.
11. Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN study design checklist for patient-reported outcome measurement instruments 2019. Available at https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf. Accessed August 12, 2020.
12. Lacy GD, Abalem MF, Popova LT, et al. Content generation for patient-reported outcome measures for retinal degeneration therapeutic trials. *Ophthalmic Genet* 2020;41:315–324.
13. Cai L. High-dimensional exploratory item factor analysis by A Metropolis–Hastings Robbins–Monro Algorithm. *Psychometrika* 2010;75:33–57.
14. Chalmers RP. mirt: a multidimensional item response theory package for the R environment. *J Stat Softw* 2012;48:1–29.
15. Bradley C, Massof RW. Method of successive dichotomizations: an improved method for estimating measures of latent variables from rating scale data. *PLoS One* 2018;13:e0206106.
16. Hambleton RK, Swaminathan H. Item Response Theory: Principles and Applications. Hingham, MA: Kluwer-Nijhoff Pub.; 1985.
17. Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of Item Response Theory. Newbury Park, CA: Sage Publications, Inc; 1991.
18. Baker FB, Kim S-H. Item Response Theory: Parameter Estimation Techniques, 176, 2nd ed. New York, NY: Marcel Dekker; 2004.
19. Thissen D, Wainer H. Test Scoring. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2001.
20. Monroe S, Cai L. Evaluating structural equation models for categorical outcomes: a new test statistic and a practical challenge of interpretation. *Multivariate Behav Res* 2015;50:569–583.
21. Kroenke K, Spitzer RL, Williams JB, Lowe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics* 2009;50:613–621.
22. Löwe B, Wahl I, Rose M, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *J Affect Disord* 2010;122:86–95.
23. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135–160.
24. Massof RW. An interval-scaled scoring algorithm for visual function questionnaires. *Optom Vis Sci* 2007;84:E690–E705.
25. Pesudovs K. Patient-centred measurement in ophthalmology—a paradigm shift. *BMC Ophthalmol* 2006;6:25.
26. Elsman EBM, van Nispen RMA, van Rens G. Psychometric evaluation of a new proxy-instrument to assess participation in children aged 3-6 years with visual impairment: PAI-CY 3-6. *Ophthalmic Physiol Opt* 2019;39:378–391.
27. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res* 2007;16(Suppl 1):5–18.
28. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 2007;45:S22–S31.
29. Hulin CL, Drasgow F, Parsons CK. Item Response Theory: Application to Psychological Measurement. Homewood, IL: Dow Jones-Irwin; 1983.

30. Elsman EBM, van Nispen RMA, van Rens G. Psychometric evaluation of the Participation and Activity Inventory for Children and Youth (PAI-CY) 0-2 years with visual impairment. *Qual Life Res* 2020;29:775–781.
31. Prem Senthil M, Khadka J, De Roach J, et al. Development and psychometric assessment of novel item banks for hereditary retinal diseases. *Optom Vis Sci* 2019;96:27–34.
32. Daiger S, Rossiter B, Greenberg J, Christoffels A, Hide W. Data services and software for identifying genes and mutations causing retinal degeneration. *Invest Ophthalmol Vis Sci* 1998;39:S295.