BMJ Open Ophthalmology

Rasch analysis of the NEI-VFQ-25: vision-related quality of life in Leber hereditary optic neuropathy after lenadogene nolparvovec gene therapy

Benson S Chen ,[1,2] Stéphanie Perot,[3] Magali Taiel,[4] Patrick Yu-Wai-Man,[1,5] Mike Horton,[6] LHON Study Group

To cite: Chen BS, Perot S, Taiel M, *et al*. Rasch analysis of the NEI-VFQ-25: vision-related quality of life in Leber hereditary optic neuropathy after lenadogene nolparvovec gene therapy. *BMJ Open Ophthalmology* 2025;**10**:e002164. doi:10.1136/bmjophth-2025-002164

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjophth-2025-002164).

PY-W-M and MH are joint senior authors.

This study was presented at the 50th Annual Meeting of the North American Neuro-Ophthalmology Society, Honolulu, Hawaii, USA.

Received 18 January 2025
Accepted 2 September 2025

For numbered affiliations see end of article.

Correspondence to
Dr Benson S Chen; bsc33@cam.ac.uk

ABSTRACT

Objectives This study aimed to evaluate the suitability of the National Eye Institute Visual Function Questionnaire (NEI-VFQ-25) for measuring vision-related quality of life (VRQoL) in patients with Leber hereditary optic neuropathy receiving lenadogene nolparvovec gene therapy in three Phase III randomised controlled clinical trials.

Methods VRQoL was assessed using the NEI-VFQ-25 at baseline (n=174) and 2 years after treatment (n=152). All participants received lenadogene nolparvovec in at least one eye. The scoring structure of the original NEI-VFQ-25 was evaluated for fit to the Rasch model, and a post hoc revision was created and psychometrically reevaluated. Stacked analysis was conducted to compare Rasch-revised scores at baseline and 2 years after treatment.

Results The original NEI-VFQ-25 exhibited multiple issues including limitations in response functioning and scale dimensionality. These issues were rectified by revising the NEI-VFQ25 into two separate unidimensional scales measuring 'Vision-related Activity Limitation' (VAL) and 'Socioemotional Functioning' (SEF). Participants' mean VAL score at baseline on a Rasch-transformed 0–100 scale was 46.1 (11.7), improving to 48.4 (13.7) after treatment ($F(1, 324) = 2.67$, p=0.103). On the SEF scale, there was a significant difference 2 years after treatment, with participants improving from a mean score of 40.1 (14.1) at baseline to 49.6 (17.6) ($F(1, 324) = 29.1$, p<0.001).

Conclusions The scoring structure of the original NEI-VFQ-25 has limitations that undermine its psychometric validity as a measure of VRQoL. Using the Rasch-revised NEI-VFQ-25, we determined that improvement in VRQoL after treatment with lenadogene nolparvovec was driven predominantly by an improvement in socioemotional functioning.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The National Eye Institute Visual Function Questionnaire (NEI-VFQ-25) is a patient-reported outcome measure developed to assess vision-related quality of life (VRQoL) in people with chronic eye diseases or low vision but has never been validated for people with Leber hereditary optic neuropathy (LHON).

WHAT THIS STUDY ADDS

⇒ Using the original scoring system of the NEI-VFQ-25, participants of three Phase III clinical trials experienced a significant improvement in the Composite VFQ-25 score 2 years after receiving lenadogene nolparvovec gene therapy for the treatment of LHON. However, the scoring structure of the original NEI-VFQ-25 was found to have several psychometric shortcomings that could be rectified by revising the NEI-VFQ-25 into two separate unidimensional scales.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ When completed by people with LHON, the NEI-VFQ-25 functions better as two separate psychometrically valid scales assessing vision-related activity limitation (VAL) and socioemotional functioning. The impact of new therapies on VRQoL, such as gene therapy, can be assessed on these scales. Due to the item content, small improvements in visual function may not lead to a significant improvement on the VAL scale.

INTRODUCTION

Patient-reported vision-related quality of life (VRQoL) is an important outcome measure in studies of ophthalmic diseases.[1] In the absence of a disease-specific patient-reported outcome measure (PROM), the 25-item National Eye Institute Visual Function Questionnaire (NEI-VFQ-25) is widely used as a measure of VRQoL in clinical trials. It assesses 11 dimensions that contribute to vision-targeted health status, including emotional well-being, social functioning and tasks of daily visual functioning.[2] Although the validity of the NEI-VFQ-25 has been established,[2 3] modern psychometric methods such as Rasch measurement theory have demonstrated shortcomings in the scoring structure of the NEI-VFQ-25.[4–7]

The psychometric validity of the NEI-VFQ-25 has not been established when used by individuals with Leber hereditary optic

neuropathy (LHON). LHON is the most common cause of mitochondrial blindness, affecting ~1 in 30 000.[8] Almost all patients experience bilateral vision loss, either simultaneously with the first eye (25%) or sequentially (75%), with the second eye being affected weeks to months later. The characteristic visual field defect in LHON is a dense central or caecocentral scotoma. Visual acuity may be mildly reduced initially, but declines severely to 20/200 or less, reaching a nadir ~4–6 weeks after onset of visual symptoms. Visual prognosis is poor. In most cases, visual acuity is worse than 3/60, fulfilling the criteria for legal blindness in most countries.[9] Treatment options in LHON are limited. However, the rapid pace of technological innovation within the field of gene therapy has accelerated the therapeutic delivery pipeline for this devastating condition.[10]

The NEI-VFQ-25 has been used as an outcome measure in recent gene therapy clinical trials for the LHON-causing m.11778G>A mutation in the *ND4* gene, including one Phase I clinical trial, three Phase III clinical trials (RESCUE, REVERSE and REFLECT) and one long-term follow-up study (RESTORE).[11–15] It is unclear if the NEI-VFQ-25 exhibits adequate psychometric properties that make it a valid measure of VRQoL in individuals affected by LHON.

The aims of this study were to assess the psychometric validity of the NEI-VFQ-25 in a LHON population using Rasch measurement theory, and to evaluate changes in VRQoL in patients receiving lenadogene nolparvovec gene therapy across three Phase III clinical trials.

METHODS

Participants

REVERSE, RESCUE and REFLECT were Phase III multicentre clinical trials of lenadogene nolparvovec (rAAV2/2-*ND4*), a modified adeno-associated virus gene therapy product specifically designed to complement the defective *ND4* gene by allotopic expression of the wild-type ND4 subunit from the nucleus followed by mitochondrial import of the protein product.[16] All studies were registered in the NIH Clinical Trials database (RESCUE, *NCT02652767*; REVERSE, *NCT02652780*; and REFLECT, *NCT03293524*).

The study design and results of REVERSE, RESCUE and REFLECT have been previously reported.[12–14] Briefly, all studies were randomised, double-masked, sham-controlled (REVERSE and RESCUE) or placebo-controlled (REFLECT) studies, that enrolled symptomatic participants with LHON aged ≥15 years with a confirmed m.11778G>A mutation in *ND4*. A total of 174 patients participated (37 in REVERSE, 39 in RESCUE and 98 in REFLECT). The studies differed in the timing of onset of vision loss (from 180 to 365 days in both eyes in REVERSE; at most 180 days in the first affected eye in RESCUE; at most 365 days in the first affected eye in REFLECT) and allocation of treatment (REVERSE and RESCUE participants received an intravitreal injection of lenadogene nolparvovec in one eye and a sham injection in the other eye in a 1:1 allocation; REFLECT participants received either bilateral treatment or unilateral treatment in a 1:1 allocation).

National Eye Institute VFQ-25 Questionnaire

In its original format, the NEI-VFQ-25 has four response scales assessing quality (3 items), frequency (4 items), difficulty (13 items) and agreement (6 items). Each response format is scored on a 5-point Likert scale (or 6-point Likert scale for the *general vision* item), ranging from (1) 'none of the time' to (5) 'all of the time' on the frequency scale, and (1) 'no difficulty at all' to (5) 'stopped doing this because of your eyesight' on the difficulty scale. On the agreement scale, items are scored from (1) 'definitely true' to (5) 'definitely false', with (3) representing 'not sure'. Responses for each item are converted to an integer score between 0 and 100, with 100 representing the best possible score and 0 representing the worst. Integer scores of items belonging to the same subscale are averaged to generate a subscale score.[17] Eleven vision-related subscales and a general health subscale can be generated, representing general health, general vision, ocular pain, near vision, distance vision, driving, peripheral vision, colour vision, role limitations, dependency, social function and mental health. Finally, an overall *Composite VFQ-25* score between 0 and 100 is calculated by averaging the subscale scores, excluding the general health subscale.

The NEI-VFQ-25 was administered in REVERSE, RESCUE and REFLECT 1–2 days prior to intravitreal injection of lenadogene nolparvovec ('baseline'), and at predesignated time points in the follow-up period. Due to slight differences in the protocols of the three studies, the NEI-VFQ-25 was obtained at 24, 48, 72 and 96 weeks ('2 years') post-treatment in REVERSE and RESCUE, while in REFLECT the NEI-VFQ-25 was obtained at 24, 52, 78, 104 ('2 years'), 156, 208 and 260 weeks post-treatment. The RESCUE and REVERSE studies have been completed with a follow-up until 2 years post-treatment. The REFLECT study is still on-going with follow-up until 5 years post-treatment. We analysed the NEI-VFQ-25 scores of all 174 participants at baseline, and those of participants where data was available at '2 years' or beyond, using the original scoring system of the NEI-VFQ-25 to generate subscale scores and an overall composite score at the two time points.

Rasch measurement theory

Rasch measurement theory is a modern psychometric method that is principally concerned with how data fits the unidimensional Rasch model.[7] Rasch analysis is conducted by comparing the pattern of observed responses with the model expectation, and highlighting measurement anomalies within a set of items. When anomalies are found, this is an indication that the set of items is not delivering measurement in the way that was originally intended, with the anomalies highlighting the limitations within the item set. However, post hoc

adjustments can be made to items to account for certain issues, allowing them to be 're-engineered' to fit the Rasch model. When the assumptions of the Rasch model are satisfied, this allows for the transformation of raw ordinal scores into interval-level measurement expressed as a log odds unit (logit) on a logit continuum, thereby permitting the use of parametric statistical techniques.[5]

To facilitate Rasch analysis across the two time points (baseline and follow-up), a random sample of participants was created. Participants with data available at only one time point (4 from RESCUE and 18 from REFLECT, all at baseline) were automatically included. For the remaining 152 participants who had data at both time points, a random number generator was used to assign either their baseline or follow-up data to the sample, ensuring that all participants were represented once that the final data set was evenly balanced across time points.[18] A breakdown of time point contributions by participant and trial is available in the online supplemental material.

Rasch analysis was performed using the Partial-Credit Model (RUMM2030 Plus, RUMM Laboratory, Perth, Australia). The raw scores of 17 items (items 1–17) were reversed for Rasch analysis, so that all measures would have the same polarity (ie, better and worse were consistent across all items). Using the raw scores of participants in the random data set, the psychometric validity of the NEI-VFQ-25 was assessed against established criteria including: (1) overall fit of the observed data to the Rasch model was tested with a χ^2 based goodness of fit test (a significant χ^2 test indicating misfit to the Rasch model)[19]; (2) fit of individual items to the Rasch model was tested using standardised residuals (Z score) and χ^2 test (Z values above ±2.50 and significant χ^2 tests indicating misfit to the Rasch model)[19]; (3) scale unidimensionality was tested by grouping items according to their loading with the first residual principal component analysis factor and pairwise t-tests calculated for each person (the number of significant t-tests (ie, the lower bound of the 95% CI) should be less than 5% to indicate unidimensionality)[20]; (4) the structure of the response scales was assessed by studying the ordering of thresholds (boundaries between response category options) of each item (the location of thresholds should increase in an ordered progression, representing the move from low to high levels on the logit continuum)[21]; (5) local independency of items was assessed by analysing the Q_3 fit statistic (if the correlation between the residuals of two items is >0.2 above the average correlation, this indicates that the pairwise items are locally dependent)[22]; (6) scale-to-sample targeting was assessed by checking the distribution of person and item threshold parameters across the logit continuum[21 23]; and (7) scale reliability was assessed using the Person Separation Index (PSI) (a PSI higher than 0.7 the minimum acceptable level).[19 23]

Iterative post hoc revisions of the NEI-VFQ-25 structure were then applied and psychometrically reevaluated. Next, the technique of stacked analysis was used to determine change in participants' VRQoL between baseline and follow-up.[18] NEI-VFQ-25 responses of participants at follow-up were stacked on the responses of participants at baseline and then analysed in the Rasch model, with the item difficulties and rating scale structure of the Rasch-revised NEI-VFQ-25 anchored to the calibrations derived from the joint dataset of randomly allocated participants across both time points. Finally, interval logit scores were transformed into an interval score on a scale between 0 and 100 to aid in interpretability of Rasch-revised NEI-VFQ-25 scores.

Statistical analysis

Statistical analyses of demographic data and subscale and composite scores (using the original scoring system of the NEI-VFQ-25) were performed with SPSS Statistics (V.28.0.1.1, IBM, Armonk, New York, USA). As NEI-VFQ-25 subscale and composite scores are ordinate in nature, Mann-Whitney U tests were conducted to compare participants' scores at baseline and follow-up.

For Rasch analysis, an adequate sample size is primarily concerned with the degree of precision of the item calibration estimates for any given scale. A sample size of 150 respondents for any given item set would provide at least 95% confidence that item estimates were stable to within ±0.5 logits.[24] Rasch-revised NEI-VFQ-25 scores were assessed according to standard analysis of variance between participants at baseline and last follow-up using RUMM2030 Plus. Statistical significance was set at $p<0.05$.

Rasch analysis is robust to item-level missing data, as person measures can be estimated from the available responses without requiring complete data. Missing responses were not imputed and were simply treated as missing. In the original scoring system, missing responses were also treated as missing (not imputed), and scores are calculated by averaging available items.

Data from both time points for all participants was used to calculate a change in NEI-VFQ-25 score (original scoring system and the Rasch-revised score).

Patient and public involvement

The use of the NEI-VFQ-25 was guided by patient feedback from an earlier study.[25] The design of the clinical trials did not involve patient or public involvement.

RESULTS

Demographic and clinical characteristics of study participants

The demographic characteristics of 174 participants at baseline are summarised in table 1. Most participants were men (79.9%), with mean (SD) age of 33.4 (14.5) years at onset of vision loss and mean duration of vision loss 7.9 (3.3) months. The follow-up cohort included a total of 152 participants. Within 2 days of the baseline assessment, 112 participants received a single intravitreal injection of lenadogene nolparvovec with either placebo or sham treatment in the fellow eye, and 41 participants had received bilateral intravitreal injection of lenadogene nolparvovec. 2-year study data was available for 142

Table 1 Characteristics of study participants at baseline and follow-up

	Baseline (n=174)	Follow-up (n=152)
Sex		
Male (%)	139 (79.9)	121 (79.6)
Age at time of assessment (years)		
Mean (SD)	33.4 (14.5)	36.4 (14.8)
Median	29	31.8
Range	15.0–74.0	17.0–77.0
Duration of vision loss at assessment (months)		
Mean (SD)	7.9 (3.3)	32.9 (6.8)
Median	8.3	32.3
Range	1.7–12.8	23.4–68.2
Duration of follow-up (years)		
Mean	N/A	2
Median		2
Range		1.8–3.0

N/A, not applicable.

participants (35 from RESCUE, 37 from REVERSE and 70 from REFLECT). NEI-VFQ-25 responses from an additional 10 unique participants of REFLECT were also included: 7 participants at 156 weeks post-treatment and 3 participants at 208 weeks post-treatment. The mean duration from baseline to follow-up was 104.7 (19.2) weeks, that is, 2.0 years.

Analysis of NEI-VFQ-25 using original scoring system

At baseline, the mean *Composite VFQ-25* score was 44.5 (15.8) out of 100; improving to 50.2 (18.7) at follow-up (p<0.01) (table 2). Significant improvements were also observed on the general vision, near activities, distance activities, mental health, role difficulties and dependency subscales at follow-up. Response rates for each item of the NEI-VFQ-25 at baseline and follow-up are available as online supplemental material.

Rasch analysis of the NEI-VFQ-25

Stage 1: initial analysis

Initial fit of the NEI-VFQ-25 to the Rasch model was poor ($\chi^2(52) = 122.2$, p<0.001). When examined for individual item fit, three items showed misfit to the Rasch model: two items (*driving during the daytime* and *limited by pain*) where the χ^2 test was significant; and one item (*need a lot of help from others*) where the standardised residual exceeded the ±2.5 limit. For the 12 items which showed disordered response category thresholds, non-functioning of the middle response categories was observed. This included all six items scored using the agreement response scale. Testing for local item dependency identified 43/325 pairwise dependencies.

Regarding scale-to-sample targeting, the mean location of person ability was −0.13 (1.07) logits, with the mean location of item difficulty centralised at 0 (1.01) logits, indicating excellent targeting of persons and items. There was no evidence of floor or ceiling effects. PSI was 0.94, indicating excellent reliability. The NEI-VFQ-25 was not unidimensional, with 11.5% of paired t-tests significant at p<0.05 (95% CI: 20.5% to 26.9%).

Stage 2: post hoc revision and psychometric reevaluation

Post hoc revisions to the structure of the NEI-VFQ-25 were made in response to the anomalies identified in Stage 1. A post hoc rescore was applied across the six items that included the 'not sure' response category, by rescoring 'not sure' as missing data. The *general health* item and all three driving items were removed in response to low response rates. The two pain-related items (*pain around eyes* and *limited by pain*) were removed as they were not conceptually relevant to LHON. To address scale dimensionality, the remaining items of the NEI-VFQ-25 were split into two scales; a scale related to limitations in vision-related activities ('Vision-related Activity Limitation' scale or VAL), and a separate scale related to social, psychological and emotional functioning ('Socioemotional Functioning' scale or SEF).

For the VAL scale, 10 items were initially included. Psychometric evaluation of the initial VAL scale showed poor fit to the Rasch model ($\chi^2(20) = 41.8$, p=0.003) and violated criteria for scale unidimensionality. Two items (*reading ordinary print in newspapers* and *seeing how people react*) showed disordered response category thresholds. Only one item was misfitting (*going out to movies/plays/sports*) due to a significant χ^2 test and appeared to be an undiscriminating item. In response to these anomalies, a post hoc rescore was applied to the two items with disordered response category thresholds by merging the non-functioning response categories 'moderate difficulty' and 'a little difficulty'. The misfitting item was removed. After the post hoc adjustments, the final VAL scale (table 3A) was psychometrically reevaluated and demonstrated satisfactory fit to the Rasch model ($\chi^2(18) = 15.5$, p=0.625), no misfitting items, no disordered response category thresholds, excellent scale-to-sample targeting and acceptable PSI (0.90). Testing for local item dependency revealed 2/36 conceptually unrelated pairwise dependencies (*seeing how people react* and *going down steps in dim light;* and *visiting with people in their homes/parties/restaurants* and *picking out and matching clothes*). Importantly, the VAL scale was unidimensional with 4.6% of paired t-tests significant at p<0.05 (95% CI: 1.4% to 7.8%).

For the SEF scale, nine items were initially included. Psychometric evaluation of the initial SEF scale showed poor fit to the Rasch model ($\chi^2(18) = 31.6$, p=.024). Testing for local item dependency revealed 2/36 pairwise dependencies. A super-item was created by merging *rely too much on others* and *need a lot of help from others*; both measuring the same concept of 'Reliance'. The

Table 2 NEI-VFQ-25 subscale and composite scores at baseline and follow-up using original scoring system

	Baseline n=174	Follow-Up n=152	Mann-Whitney U test Z-score, p value
Composite VFQ-25			
Mean (SD)	44.5 (15.8)	50.2 (18.7)	Z=−2.92, p=0.003
Median (IQR)	41.6 (33.3–55.3)	50.2 (34.8–62.7)	
General health			
Mean (SD)	65.7 (26.0)	64.1 (25.1)	Z=−0.63, p=0.530
Median (IQR)	75.0 (50.0–75.0)	75.0 (50.0–75.0)	
General vision			
Mean (SD)	33.8 (17.2)	41.2 (21.4)	Z=−3.64, p<0.001
Median (IQR)	40.0 (20.0–40.0)	40.0 (20.0–60.0)	
Ocular pain			
Mean (SD)	86.5 (18.5)	85.8 (18.4)	Z=−0.02, p=0.987
Median (IQR)	100 (75.0–100)	100 (75.0–100)	
Near activities			
Mean (SD)	28.7 (19.8)	34.3 (23.2)	Z=−3.58, p<0.001
Median (IQR)	25 (16.7–41.7)	33.3 (16.7–50.0)	
Distance activities			
Mean (SD)	39.0 (20.7)	43.8 (22.4)	Z=−2.90, p=0.004
Median (IQR)	37.5 (25.0–50.0)	41.7 (25.0–58.3)	
Social functioning			
Mean (SD)	48.6 (25.2)	49.7 (25.5)	Z=−1.38, p=0.167
Median (IQR)	50.0 (25.0–62.5)	50.0 (25.0–65.6)	
Mental health			
Mean (SD)	33.9 (22.3)	49.5 (26.3)	Z=−6.01, p<0.001
Median (IQR)	31.3 (18.8–50.0)	50.0 (31.3–68.8)	
Role difficulties			
Mean (SD)	34.1 (26.9)	47.9 (28.5)	Z=−4.91, p<0.001
Median (IQR)	25.0 (12.5–50.0)	50.0 (25.0–75.0)	
Dependency			
Mean (SD)	37.3 (24.6)	49.1 (28.8)	Z=−4.21, p<0.001
Median (IQR)	33.3 (16.7–50.0)	41.7 (25.0–75.0)	
Driving			
Mean (SD)	5.2 (18.0)	6.2 (21.2)	Z=−0.04, p=0.970
Median (IQR)	0 (0–0)	0 (0–0)	
Colour vision			
Mean (SD)	71.4 (27.5)	72.5 (28.1)	Z=−0.48, p=0.628
Median (IQR)	75.0 (50.0–100)	75.0 (50.0–100)	
Peripheral vision			
Mean (SD)	61.6 (26.0)	63.2 (25.4)	Z=−0.54, p=0.592
Median (IQR)	50 (50.0–75.0)	62.5 (50.0–75.0)	

NEI-VFQ-25, 25-item National Eye Institute Visual Function Questionnaire.

second pairwise dependency (*feeling frustrated* and *time worrying about eyesight*) was left as separate items as it was felt that they measured distinct concepts. The final SEF scale (table 3B) was psychometrically reevaluated and demonstrated satisfactory fit to the Rasch model ($\chi^2(16)=20.2$, p=.211), no misfitting items, no disordered response category thresholds, excellent scale-to-sample targeting, acceptable PSI (0.87) and was unidimensional

Table 3 Item fit statistics and parameter values for items of Rasch-revised Vision-Related Activity Limitation (VAL) and Socioemotional Functioning (SEF) scales

Item	Description	Location (Logits)	SE	Fit residual	χ^2 prob
A. Rasch-revised Vision-Related Activity Limitation scale					
1	Picking out and matching clothes	−1.388	0.097	1.213	0.787
2	Noticing objects off to the side	−1.171	0.104	−0.172	0.161
3	Going down steps in dim lit	−1.060	0.100	1.944	0.765
4	Visiting with people in their homes/ parties/restaurants	−0.906	0.095	1.181	0.556
5	Finding objects on a crowded shelf	−0.050	0.108	−1.592	0.060
6	Seeing how people react	0.144	0.114	0.889	0.789
7	Seeing well up close	0.661	0.099	−1.502	0.909
8	Street signs	1.655	0.112	0.169	0.720
9	Reading ordinary print in newspapers	2.114	0.142	0.520	0.255
B. Rasch-revised Socioemotional Functioning scale					
1	Stay home most of the time	−0.948	0.113	0.327	0.781
2	Doing embarrassing things	−0.615	0.107	1.148	0.364
3	Limited in how long you can work because of vision	−0.111	0.088	1.089	0.670
4	Feel frustrated	0.185	0.119	0.214	0.333
5	'Reliance' super-item: Rely too much on what others tell me and need a lot of help from others	0.229	0.084	−0.910	0.705
6	Accomplish less because of vision	0.365	0.096	−1.711	0.119
7	Much less control	0.373	0.125	−1.246	0.033
8	Time worrying about eyesight	0.523	0.101	1.845	0.224

with 4.1% of paired t-tests significant at p<.05 (95% CI: 0.9% to 7.4%).

Finally, participants' scores were analysed on the separate psychometrically valid scales. At baseline, participants' mean VAL score on a Rasch-transformed 0–100 scale was 46.1 (11.7), improving to 48.4 (13.7) 2 years post-treatment ($F(1, 324) = 2.67$, p=0.103) (figure 1). On the SEF scale, there was a significant difference 2 years post-treatment, with participants improving from a mean score of 40.1 (14.1) at baseline to 49.6 (17.6) ($F(1, 324) = 29.1$, p<0.001) (figure 2).

DISCUSSION

Despite its popularity as a PROM in ophthalmology clinical trials, the NEI-VFQ-25 has been criticised for its violations of modern psychometrics, in particular the principle of unidimensionality and its use as a measure of a single 'vision-targeted health-related quality of life' construct, that is, the items that are used to calculate the single *Composite VFQ-25* score do not actually measure a single construct.[4–6] The results of our study are consistent with those of earlier studies involving patients with cataracts, low vision and retinal disorders.[4–6] Across all studies, the NEI-VFQ-25 offered improved psychometric performance when split into separate scales measuring conceptually distinct visual functioning and socioemotional constructs. Within each scale, individual items can be positioned along a continuum of difficulty, which allows changes in visual ability over time to be interpreted in terms of a person's movement along that continuum. While the original scoring system indicated significant improvement in some subscales at follow-up, these scores combined items with different levels of difficulty and conceptual focus, making it difficult to draw clear conclusions about actual changes in visual function. For example, the near activities subscale improved using the original scoring, while the Rasch-based VAL score did not. Improvements were likely driven by easier tasks (*finding objects on a crowded shelf*), while more challenging ones (*reading ordinary print in newspapers*) remained unchanged, giving a potentially misleading impression of improvement.

The scales constructed in our study and order of items along the logit scale are similar to those of earlier studies.[4–6] We excluded three driving items, two ocular pain items and the general health and general vision items to construct separate VAL and SEF scales, both consisting of nine items. In contrast to Marella *et al*,[4] who constructed similar scales using NEI-VFQ-25 responses from people with low vision and dichotomised items due to disordered thresholds, we retained the polytomous response structure. The segregation of items into

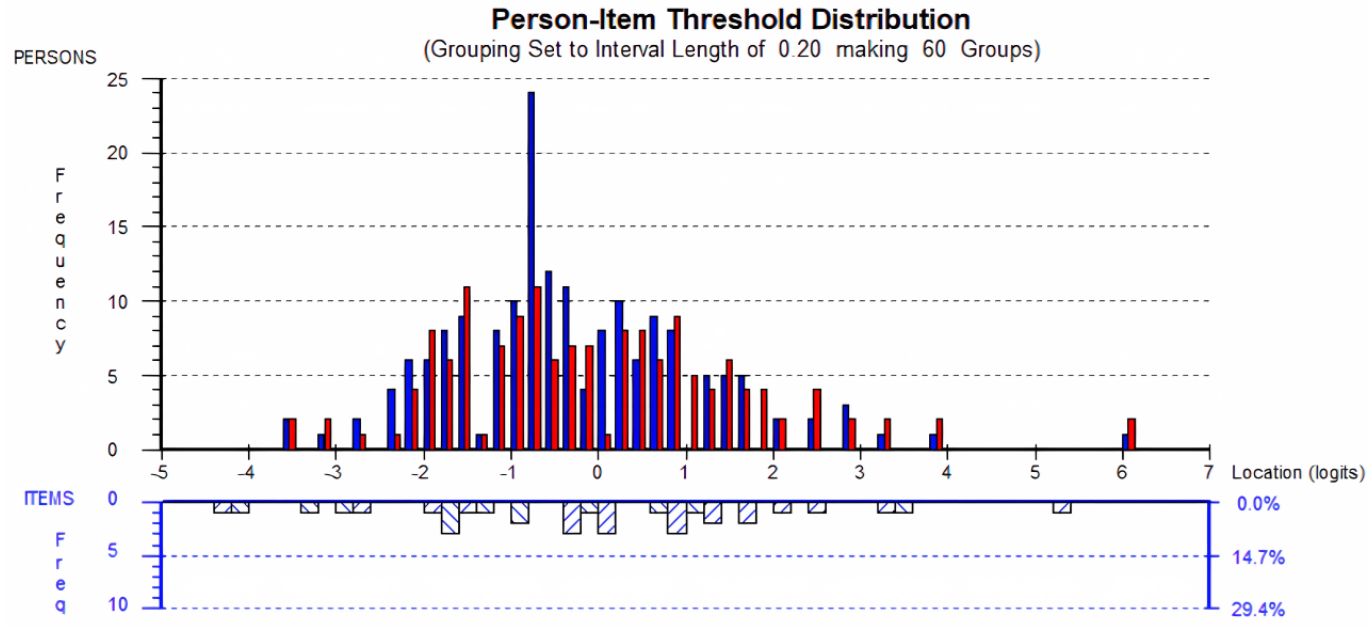

Figure 1 Person-item threshold distribution map for the Rasch-revised Vision-Related Activity Limitation (VAL) scale. The blocks in the upper part of the chart (above the x-axis) represent groups of study participants across a continuum of vision-related activity limitation, represented by the x-axis ('Location' in logits), at baseline (in blue) and at 2 years post-treatment (in red). The underlying scale represents higher levels of vision-related activity limitation (or lower levels of visual functioning) toward the left of the plot, and lower levels of vision-related activity limitation (or higher levels of visual functioning) toward the right of the plot. The blocks in the lower part of the chart (below the x-axis) represent the location of item thresholds and their distribution.

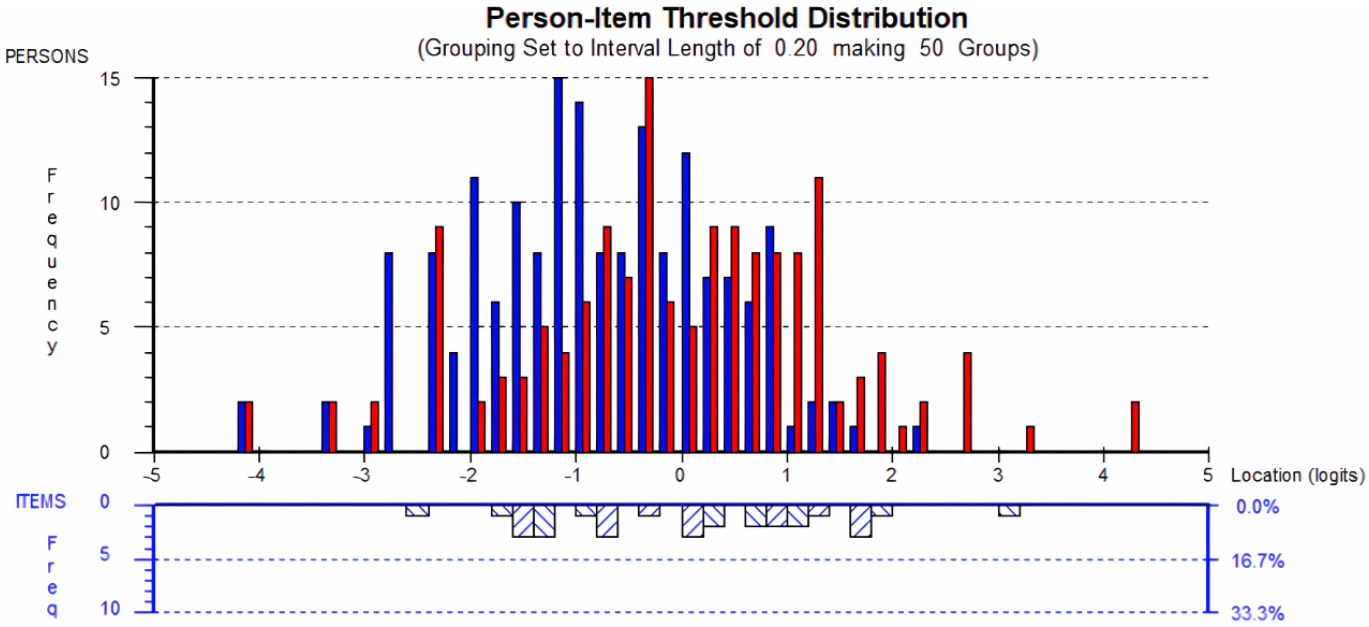

Figure 2 Person-item threshold distribution map for the Rasch-revised Socioemotional Functioning (SEF) scale. The blocks in the upper part of the chart (above the x-axis) represent groups of study participants across a continuum of socioemotional functioning, represented by the x-axis ('Location' in logits), at baseline (in blue) and at 2 years post-treatment (in red). The underlying scale represents lower levels of socioemotional functioning toward the left of the plot, and higher levels of socioemotional functioning toward the right of the plot. The blocks in the lower part of the chart (below the x-axis) represent the location of item thresholds and their distribution.

separate visual functioning and socioemotional scales was also broadly similar to studies using NEI-VFQ-25 responses from people with cataracts[5] and retinal disorders.[6]

Our study raises an important question regarding the impact of gene therapy on visual functioning and socioemotional function in patients affected by LHON. Most studies exploring the quality of life of individuals affected by LHON rely on pre-existing or novel PROMs designed to assess vision-related activity limitation.[26] Using PROMs such as the Visual Function Index (VF-14), a measure of vision-related activity limitation, patients with LHON have the lowest score compared with other ophthalmological disorders.[27] A longitudinal study of Chinese patients with the m.11778G>A mutation found that the VF-14 score improved over time, especially in younger patients.[28] Although the psychometric validity of the VF-14 has been questioned previously,[29] the study found significant improvement in easier items such as *seeing steps, curbs or stairs; doing handiwork; playing sports*, *cooking*; and *watching television.*[28] In contrast, responses to more difficult items like *reading small print; reading a newspaper or a book; reading signs*; and *writing cheques or filling out forms* were unchanged. In our study, the VAL scale was constructed with similar items representing higher visual function ability, with *reading ordinary print in newspapers* the most difficult item for participants to endorse. Although participants' visual function improved after treatment in our study, the magnitude of the improvement was not large enough to reach significance on the VAL scale. This is not surprising. Long-term follow-up of patients receiving unilateral intravitreal gene therapy for LHON showed mean best corrected visual acuity (BCVA) steadily improving from 1.57 logMAR at 12 months to 1.26 logMAR at 48 months after onset.[15]

In our study, the SEF scale included items that assessed participants' ability to work/fulfil their roles, reliance on others, and feelings of embarrassment, frustration and worry. We found that study participants experienced a significant improvement in socioemotional function after receiving gene therapy. These results must be interpreted with caution in the absence of a control group. However, from natural history studies, it is understood that LHON is characterised by chronic psychosocial losses, including loss of social/communication skills, loss of independence and freedom and loss of autonomy.[25 30] Symptoms of anxiety and depression are common in affected individuals.[25 31–33] Nearly 50% of participants in one study had at least five of the nine symptoms of major depressive disorder.[32] Another study found that 8% of men affected by LHON had mild to moderate depression using the Beck's Depression Inventory.[33] Although personal factors/attitudes and external/environmental factors, including access to emotional support and low vision aids, are important in determining the extent and duration of sadness after vision loss,[31] the negative impact of LHON on social well-being is persistent and pervasive.[25 32] Approximately 71% of affected individuals report that LHON had a negative impact on both interpersonal relationships and career goals.[32] Indeed, loss of autonomy and inability to independently perform tasks or fulfil previous roles is frequently reported by affected individuals, even with access to low vision aids and assistive technology.[25]

This study has several limitations. One limitation is the uneven distribution of participants across the three contributing trials, with a larger proportion drawn from REFLECT. However, REFLECT had comparable inclusion criteria to RESCUE and REVERSE, encompassing the same disease duration window (up to 365 days from onset of vision loss), and the follow-up interval differed by only 8 weeks. We believe these differences are unlikely to have introduced meaningful bias into the pooled results. In the absence of a control group, the change in scores on the Rasch-revised NEI-VFQ-25 scales cannot be solely attributed to the effect of gene therapy. However, the three clinical trials prohibited participants from taking idebenone for the duration of the trial and, other than the intravitreal injection of lenadogene nolparvovec, no other treatments were offered to all study participants. The change in scores on the Rasch-revised SEF scale may also be due to response shift, reflecting a change in participants' perception of the concepts being measured. It is recognised that patients might give different answers to questionnaires over time, not only because their health has changed, but also because their perception of health or what QoL means to them has changed.[34] This is especially important in LHON, where the psychological impact of the condition has been described as occurring in different phases.[25] By conducting Rasch analysis using anchor values from a random sample of participants with both time points equally represented, a possible source of time-series dependency is eliminated.[18] While this approach strengthens the internal validity of the findings, the interpretation of change remains limited without an established threshold for clinically meaningful improvement. Future work correlating Rasch-derived scores with other person-level measures, such as visual acuity, will be important for defining such thresholds, but this was beyond the scope of the present study.

Lastly, Rasch informed post hoc revisions can improve the psychometric validity of the NEI-VFQ-25, but they do not address underlying issues with the content of the PROM itself. While items contributing to the SEF scale appear to reflect socioemotional challenges experienced by individuals with LHON, important aspects such as symptoms of anxiety, depression or the economic burden of vision loss are not directly addressed.[25 26 31 32 35–37] Similarly, it remains uncertain whether the items contributing to the VAL scale are conceptually relevant or sufficiently comprehensive to capture meaningful change in visual function. Regulatory guidance, such as the Food and Drug Administration's Patient-Focused Drug Development framework,[38–41] recommends direct engagement with patients to ensure that PROMs measure concepts that are most important to those living with the condition.

Future research should focus on identifying core domains to guide the development or refinement of PROMs for use in LHON.

In conclusion, the NEI-VFQ-25 in its original structure does not possess valid psychometric properties as a single measure of VRQoL in people with LHON. The NEI-VFQ-25 functions better as two psychometrically valid scales measuring separate visual functioning and socioemotional constructs. Participants of three Phase III clinical trials experienced improvements in visual functioning and socioemotional traits 2 years after receiving intravitreal injection of lenadogene nolparvovec for the treatment of LHON. Gene therapy appears to have a significant impact on the social, psychological and emotional impact of LHON, enabling affected individuals to become more independent and self-reliant, feel less embarrassed and frustrated, and accomplish more.

Author affiliations
[1]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK
[2]Department of Medicine, The University of Auckland, Auckland, Auckland, New Zealand
[3]eXYSTAT, Malakoff, Île-de-France, France
[4]Medical, GenSight Biologics SA, Paris, Paris, France
[5]Moorfields Eye Hospital NHS Foundation Trust, London, UK
[6]Psychometric Laboratory for Health Sciences, University of Leeds, Leeds, UK

Collaborators The following individuals were part of the LHON Study Group and contributed to the data collection and/or coordination of RESCUE, REVERSE and REFLECT: Valerio Carelli, Piero Barboni, Michele Carbonelli, Lidia Di Vito, Giulia Amore, Manuela Contin, Susan Mohamed, Chiara La Morgia, Martina Romagnoli, Sara Silvestri, Pietro D'Agati, Nancy J Newman, Valérie Biousse, George Baker Hubbard, Ghazala O'Keefe, Andrew M Hendrick, Michael Dattilo, Jason H Peragallo, Eman Hawy, Lindreth DuBois, Deborah Gibbs, Alcides Fernandes Filho, Jannah Dobbs, Andre Aung, Patrick Yu-Wai-Man, Benson S Chen, James Acheson, Hayley Boston, Maria Eleftheriadou, Simona Esposti, Maria Gemenetzi, Lauren Leitch-Devlin, William R Tucker, Neringa Jurkute, Asma Burale, Shweta Anand, Muhammad A Memon, Rima Hussain, Rasha Jorany, Priyansha Sheel, Sarah Thornton, Mark L Moster, Robert C Sergott, Melissa SantaMaria, Heather Tollis, Adam A DeBusk, Julia A Haller, Maria Massini, Prem S Subramanian, Paula Pecen, Marc Mathias, Mary Preston, Steve Cho, José A Sahel, Catherine Vignal-Clermont, Jean François Girmens, Rabih Hage, Lise Plaine, Wahiba Khemliche, An-Guor Wang, Hui-Chen Cheng, Celia Chen, Jeong-Min Hwang, Chuanbin Sun, Sean Donahue, Shriji Patel, Sapna Gangaputra, Megan Barrett, Scott Ruark, Saige Wilkins, Bart P Leroy, Julie De Zaeytijd, Caroline Van Cauwenbergh, Hilde Verhauwen, Thomas Klopstock, Claudia B Catarino, Claudia Priglinger, Siegfried Priglinger, Günther Rudolph, Stephan Thurau, Bettina von Livonius, Daniel Muth, Armin Wolf, Jasmina Al-Tamami, Angelika Pressler, Cosima Schertler, Martin Hildebrandt, Michael Neuenhahn, Alfredo A Sadun, Gad Heilweil, Rustum Karanjia, Irena Tsui, Gema Rebolleda Fernández, Laia Jaumendreu Urquijo, Francisco J Muñoz Negrete, Elizabeth Fortin, Bart K Chwalisz, Dean Cestari, Rudrani Banik, Katy Tai, Lorena Castillo, Virginia Garcia, Antonio Morilla, Francesco Bandello, Piero Barboni, Maria Lucia Cascavilla, Marco Battista, Francesca Calcagno, Adelaide Pina, Stéphanie Leruez, Rod Foroozan.

Contributors BSC, PY-W-M and MH contributed to the study conception and design. MT and SP contributed to data collection and provided the dataset. BSC, SP and MH conducted the analysis. BSC and MH interpreted the study results. The first draft of the manuscript was written by BSC and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. PY-W-M is the study guarantor.

Funding This study was conducted using data obtained from clinical trials funded by GenSight Biologics, Paris, France.No funding or financial support was received for this study. BSC is supported by a Senior Clinical Research Fellowship (2433 SCF) from the Neurological Foundation (New Zealand). PYWM is supported by an Advanced Fellowship Award (NIHR301696) from the UK National Institute of Health Research (NIHR), and also receives funding from Fight for Sight (UK), the Isaac Newton Trust (UK), Moorfields Eye Charity, the Addenbrooke's Charitable Trust, the National Eye Research Centre (UK), the International Foundation for Optic Nerve Disease (IFOND), the UK National Institute of Health Research (NIHR) as part of the Rare Diseases Translational Research Collaboration, the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014), and the NIHR Biomedical Research Centre based at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology (NIHR203322). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The remaining authors have no competing interests to declare.

Competing interests MT is an employee of GenSight Biologics. PYWM is a consultant for GenSight Biologics and Stealth BioTherapeutics and has received research support from GenSight Biologics and Santhera Pharmaceuticals.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The original clinical trials (RESCUE, REVERSE and REFLECT) were approved by local independent ethics committees and adhered to principles outlined in the Declaration of Helsinki. RESCUE: Comités de protection des personnes, Ile De France V (CPP Ref: 15056); Ethics Committee of the LMU Munich (463-15 fed); London-West London and GTAC Research Ethics Committee (15/LO/1308); Comitato Etico Interaziendale Bologna-Imola (CE-BI) (946/CE); UCLA Institutional Review Board (UCLA IRB) (IRB #15-001135); Wills Eye Hospital Institutional Review Board (IRB# 15-505). REVERSE: Comités de protection des personnes, Ile De France V (CPP Ref: 15055); Ethics Committee of the LMU Munich (459-15 fed); Comitato Etico Interaziendale Bologna-Imola (CE-BI) (947/CE); Emory Institutional Review Board (IRB00084565); London-West London and GTAC Research Ethics Committee (18/LO/1309); Wills Eye Hospital Institutional Review Board (IRB# 15-506). REFLECT: Comités de protection des personnes, Nord-Ouest II (CPP Ref: 2018/15); London-West London and GTAC Research Ethics Committee (18/LO/0521); Comité de Ética de la Investigación con medicamentos de HM Hospitales (18.01.1161-GHM); Comitato etico Indipendente di Area Vasta Emilia Centro (EM1019_2019-18080-EM 2 -FARM-AUSLBO); Ghent University Medical Ethics Committee (2018/0074); UCLA Institutional Biosafety Committee (BUA-2017-258-001); Mount Sinai Institutional Biosafety Committee (2018-041); Emory Institutional Review Board (IRB00100950); Jefferson Institutional Biosafety Committee (IBC Control #15.1118); Vanderbilt Institutional Biosafety Committee for Human Subjects (IRB # 180255); Massachusetts Eye and Ear Human Studies Committee (1221964-2, IND/IDE#16538); University of Colorado Denver | Anschutz Medical Campus Institutional Biosafety Committee (IRB# 17-2196). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iD
Benson S Chen https://orcid.org/0000-0001-8214-0186

REFERENCES

1 Braithwaite T, Calvert M, Gray A, *et al*. The use of patient-reported outcome research in modern ophthalmology: impact on clinical trials and routine clinical practice. *Patient Relat Outcome Meas* 2019;10:9–24.

2 Mangione CM, Lee PP, Gutierrez PR. Development of the 25-list-item National Eye Institute Visual Function Questionnaire. *Arch Ophthalmol* 2001;119:1050.
3 Mangione CM, Lee PP, Pitts J, *et al*. Psychometric properties of the National Eye Institute Visual Function Questionnaire (NEI-VFQ). NEI-VFQ Field Test Investigators. *Arch Ophthalmol* 1998;116:1496–504.
4 Marella M, Pesudovs K, Keeffe JE, *et al*. The psychometric validity of the NEI VFQ-25 for use in a low-vision population. *Invest Ophthalmol Vis Sci* 2010;51:2878–84.
5 Pesudovs K, Gothwal VK, Wright T, *et al*. Remediating serious flaws in the National Eye Institute Visual Function Questionnaire. *J Cataract Refract Surg* 2010;36:718–32.
6 Petrillo J, Bressler NM, Lamoureux E, *et al*. Development of a new Rasch-based scoring algorithm for the National Eye Institute Visual Functioning Questionnaire to improve its interpretability. *Health Qual Life Outcomes* 2017;15:157.
7 Rasch G. *Probabilistic models for some intelligence and attainment tests*. ERIC, 1993.
8 Yu-Wai-Man P, Griffiths PG, Brown DT, *et al*. The epidemiology of Leber hereditary optic neuropathy in the North East of England. *Am J Hum Genet* 2003;72:333–9.
9 Yu-Wai-Man P, Newman NJ, Carelli V, *et al*. Natural history of patients with Leber hereditary optic neuropathy-results from the REALITY study. *Eye (Lond)* 2022;36:818–26.
10 Chen BS, Yu-Wai-Man P, Newman NJ. Developments in the Treatment of Leber Hereditary Optic Neuropathy. *Curr Neurol Neurosci Rep* 2022;22:881–92.
11 Lam BL, Feuer WJ, Porciatti V, *et al*. Leber Hereditary Optic Neuropathy Gene Therapy: Longitudinal Relationships Among Visual Function and Anatomical Measures. *Am J Ophthalmol* 2024;257:113–28.
12 Newman NJ, Yu-Wai-Man P, Carelli V, *et al*. Efficacy and Safety of Intravitreal Gene Therapy for Leber Hereditary Optic Neuropathy Treated within 6 Months of Disease Onset. *Ophthalmology* 2021;128:649–60.
13 Yu-Wai-Man P, Newman NJ, Carelli V, *et al*. Bilateral visual improvement with unilateral gene therapy injection for Leber hereditary optic neuropathy. *Sci Transl Med* 2020;12:eaaz7423.
14 Newman NJ, Yu-Wai-Man P, Subramanian PS, *et al*. Randomized trial of bilateral gene therapy injection for m.11778G>A MT-ND4 Leber optic neuropathy. *Brain (Bacau)* 2023;146:1328–41.
15 Biousse V, Newman NJ, Yu-Wai-Man P, *et al*. Long-Term Follow-Up After Unilateral Intravitreal Gene Therapy for Leber Hereditary Optic Neuropathy: The RESTORE Study. *J Neuroophthalmol* 2021;41:309–15.
16 Chen BS, Newman NJ. Clinical trials in Leber hereditary optic neuropathy: outcomes and opportunities. *Curr Opin Neurol* 2025;38:79–86.
17 Mangione CM. Version 2000. The National Eye Institute 25-Item Visual Function Questionnaire (VFQ-25) Scoring Algorithm. National Eye Institute; 2023. Available: https://www.nei.nih.gov/sites/default/files/2019-06/manual_cm2000.pdf
18 Mallinson T. Rasch Analysis of Repeated Measures. RMT 2011;251:1317.
19 Andrich D, Marais I. *A course in rasch measurement theory*. Berlin Heidelberg: Springer, 2019.
20 Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002;3:205–31.
21 Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. *Int J Nurs Stud* 2009;46:380–93.
22 Christensen KB, Makransky G, Horton M. Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Appl Psychol Meas May* 2017;41:178–94.
23 Tennant A, Küçükdeveci AA. Application of the Rasch measurement model in rehabilitation research and practice: early developments, current practice, and future challenges. *Front Rehabil Sci* 2023;4:1208670.
24 Linacre JM. Sample size and item calibration stability. *RMT* 1994;328.
25 Chen BS, Holzinger E, Taiel M, *et al*. The Impact of Leber Hereditary Optic Neuropathy on the Quality of Life of Patients and Their Relatives: A Qualitative Study. *J Neuroophthalmol* 2022;42:316–22.
26 Chen BS, Galus T, Archer S, *et al*. Capturing the experiences of patients with inherited optic neuropathies: a systematic review of patient-reported outcome measures (PROMs) and qualitative studies. *Graefes Arch Clin Exp Ophthalmol* 2022;260:2045–55.
27 Kirkman MA, Korsten A, Leonhardt M, *et al*. Quality of life in patients with leber hereditary optic neuropathy. *Invest Ophthalmol Vis Sci* 2009;50:3112–5.
28 Cui S, Jiang H, Peng J, *et al*. Evaluation of Vision-Related Quality of Life in Chinese Patients With Leber Hereditary Optic Neuropathy and the G11778A Mutation. *J Neuroophthalmol* 2019;39:56–9.
29 Chen BS, Yu-Wai-Man P, Horton M. Psychometric Validity of the Visual Function Index in Leber Hereditary Optic Neuropathy. *Transl Vis Sci Technol* 2023;12:23.
30 Ferguson J, de Abreu G. What is the lived experience for people with Leber Hereditary Optic Neuropathy? *British Journal of Visual Impairment* 2016;34:109–20.
31 Gale J, Khoshnevis M, Frousiakis SE, *et al*. An International Study of Emotional Response to Bilateral Vision Loss Using a Novel Graphical Online Assessment Tool. *Psychosomatics* 2017;58:38–45.
32 Garcia GA, Khoshnevis M, Gale J, *et al*. Profound vision loss impairs psychological well-being in young and middle-aged individuals. *Clin Ophthalmol* 2017;11:417–27.
33 Rabenstein A, Catarino CB, Rampeltshammer V, *et al*. Smoking and alcohol, health-related quality of life and psychiatric comorbidities in Leber's Hereditary Optic Neuropathy mutation carriers: a prospective cohort study. *Orphanet J Rare Dis* 2021;16:127.
34 Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med* 1999;48:1507–15.
35 Hawlina M, Kovač L, Breciková K, *et al*. Leber hereditary optic neuropathy in Slovenia: quality of life and costs from patient perspective. *Orphanet J Rare Dis* 2024;19:318.
36 Bušányová B, Vajter M, Kelifová S, *et al*. Leber hereditary optic neuropathy in Czechia and Slovakia: Quality of life and costs from patient perspective. *Heliyon* 2024;10:e32296.
37 Lawrence C, Williams E, Mumford A, *et al*. Leber hereditary optic neuropathy: utilities and carer burden from British and Irish participants. *Orphanet J Rare Dis* 2025;20:219.
38 Patient focused drug development: selecting, developing, or modifying fit-for purpose clinical outcome assessments. guidance for industry, food and drug administration staff, and other stakeholders. 2020.
39 Patient focused drug development: selecting, developing, or modifying fit-for purpose clinical outcome assessments. guidance for industry, food and drug administration staff, and other stakeholders (draft guidance). US Department of Health and Human Services; 2022.
40 Patient-focused drug development: methods to identify what is important to patients. guidance for industry, food and drug administration staff, and other stakeholders. US Department of Health and Human Services; 2022.
41 Patient-focused drug development: incorporating clinical outcome assessments into endpoints for regulatory decision-making guidance for industry, food and drug administration staff, and other stakeholders. US Department of Health and Human Services; 2023.