Psychometric evaluation of a modified visual function questionnaire (mVFQ-25) using data from a phase III open-label randomised controlled trial in patients with *RPE65*-mediated inherited retinal dystrophy (IRD)

D Viriato,[1] C Spera,[1] N Williamson,[2] C Tolley,[2] R Arbuckle,[2] J Bennett[3]

[1]Novartis Pharma AG, Basel, Switzerland; [2]Adelphi Mill, Bollington, Cheshire, SK10 5JB, UK; [3]University of Pennsylvania, Philadelphia, Pennsylvania, USA

ISPOR EU, Copenhagen, Denmark; 2nd-6th November 2019

Background

- *RPE65*-mediated inherited retinal dystrophy (IRD) is a rare condition that causes retinal degeneration. Patients can suffer from a range of visual function symptoms including night blindness, impaired contrast sensitivity, peripheral vision and loss of visual acuity and may eventually progress to total blindness.[1]
- Patients experience significant impacts on their functional vision and wider health-related quality life (HRQoL).[1]
- A modified version of the Visual Function Questionnaire (mVFQ-25) was developed for use with pediatric populations with visual impairments, which was administered to patients with *RPE65*-mediated IRD in the phase III trial of voretigene neparvovec, the first one-time gene therapy for patients with vision loss due to *RPE65* mutations.[2,3]

Objectives

- This study aimed to evaluate the psychometric properties of the mVFQ-25 using data from the voretigene neparvovec phase III trial, pooled across treatment groups.

Methodology

Study design

- The trial was a phase III, open-label, randomised, controlled, crossover study.
- The sample comprised 31 patients, aged three years and above, randomised into either an intervention (n=21) or control group (n=10). Control patients transitioned into the intervention group one year after baseline.

Assessments and administration

- The mVFQ-25 consists of 25 items assessing activities of daily living affected by patients' vision impairment. Each item has a one month recall period and a 0-10 numerical response scale with anchors of 'always' and 'never'; higher scores indicate better functional vision.
- The mVFQ-25 was completed by patients and/or caregivers. Scores from patient self-completions are reported as 'patient reported scores', while 'caregiver/patient reported scores' comprise caregiver reports where available and patient reports where no caregiver data were available.
- The Multi-Luminance Mobility Test (MLMT) is a performance-related outcome (PerfO) assesses functional vision in different lighting conditions. MLMT change scores were calculated as the difference between scores on the lowest light level passed at follow up, compared to baseline. Change scores of ≥+1 indicated improvement, 0 stability and ≤-1 worsening.
- Additional assessments included visual acuity (VA), visual field (VF) and the full-field light sensitivity threshold (FST). Patients were grouped into the following severity groups based on VF scores; mild (89.0-100%), moderate (72.0-88.0%), severe (45.0-71.0%) and very severe (0-44.0%).[4]
- Core assessment visits were baseline, days 30, 90 and 180 and 1-year.

Analyses

- All psychometric analyses were conducted using SAS version 9.4 or higher (Table 1).
- The primary analysis population (n=38) was used for most analyses and consisted of patients in the intervention group (n=20), control patients at 1-year (n=9) and the same control patients at year 2.

Table 1. Overview of statistical analyses

Analysis		Description
Item properties	Quality of completion	• Quality of completion was assessed by examining the proportion of missing data at the item-level at baseline and 1-year follow-up.
	Item response distributions and floor/ceiling effects	• Item response distributions and floor/ceiling effects for each item were examined at baseline and 1-year follow-up to identify any skewed distributions.
Reliability	Internal consistency	• Cronbach's alpha was calculated at baseline and 1-year (alpha >0.70 considered evidence of good internal consistency).
	Test-retest reliability	• Test retest reliability (TRT) was evaluated by examining the stability of scores between days 30-90 and days 90-180 in stable patients. Stability was defined in terms of stability in VF severity and as those with a MLMT change score of 0. • An intra-class correlation (ICC) coefficient of >0.70 was considered evidence of good TRT.[5]
Construct validity	Convergent validity	• Convergent validity was evaluated by examining correlations between mVFQ-25 scores and other measures at baseline and 1-year follow-up. • Domains assessing related concepts were expected to correlate at ≥0.50 (MLMT; bilateral change score and assigned first eye, FST and VF). Measures assessing unrelated concepts (VA) were expected to correlate at ≤0.30.[5]
	Known-groups analysis	• The known-groups method evaluated mVFQ-25 scores between groups of participants defined by MLMT scores; at baseline and 1-year follow-up.[5]
Ability to detect change	Responsiveness	• mVFQ-25 scores were compared among 'improved' or 'stable/ worsened' groups defined based on changes in MLMT scores.[5]
Interpretation of score changes	Estimation of clinically meaningful change/ response definition	• Distribution-based methods were used to estimate meaningful change thresholds by calculating the 0.5 standard deviation (SD) and standard error of the mean (SEM) at baseline. • Exploratory anchor-based estimates were derived using the ability to detect change analysis.
	Additional exploratory anchor-based estimates	• Individual-level responder definitions were assessed using forest plots and distribution-based methods. • Group-level anchor-based methods, Receiver Operating Characteristic (ROC) analysis curves and empirical cumulative distribution function (eCDF) curves were also calculated. • To derive meaningful change thresholds, baseline and 1-year follow-up mVFQ-25 change scores were assessed using: the mean change score and SD of change, and; the median change score and 25th percentile change score.

Results

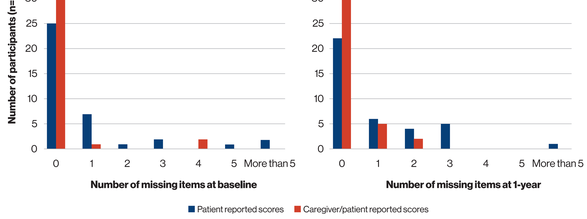
Demographic and clinical characteristics

- Patients had a mean age of 15.1 years (range: 4-44 years) and eighteen patients were female (58.0%).

Item properties

- Completion rates were high using patient and caregvier/patient reported scores (Figure 1); 65.8%-92.1% of participants had no missing data at baseline and 57.9%-81.6% of participants at 1-year.
- Item response distributions showed high rates of floor effects at baseline (range: 10.5%-60.5%) and high rates of ceiling effects at 1-year (range: 10.5%-60.5%). The pattern indicates improvements in functional vision due to treatment, providing supportive evidence that the instrument will have the ability to detect change over time.

Figure 1. Quality of completion per participant at baseline and 1-year follow up (n=38)



Internal consistency

- Cronbach's alpha for the mVFQ-25 total score using both patient and caregiver/patient reported scores exceeded 0.70 at baseline and 1-year, indicating excellent internal consistency reliability (range: 0.854-0.970).

Test-retest reliability

- The ICCs for the mVFQ-25 total score in a subgroup defined as stable exceeded 0.70 for both patient and caregiver/patient reported scores, indicating good reliability (ICC=0.889-0.937; Table 2).

Table 2. Test-retest reliability at day 30-90 and day 90-180 using MLMT and VF anchors (TRT population)

Modified VFQ-25 total score	MLMT change score				VF severity category			
	Patient reported scores		Caregiver/patient reported scores		Patient reported scores		Caregiver/patient reported scores	
	N	ICC	N	ICC	N	ICC	N	ICC
Day 30-90	32	**0.904**	32	**0.914**	27	**0.889**	27	**0.937**
Day 90-180	32	**0.912**	32	**0.916**	27	**0.900**	27	**0.937**

ICC >0.70 in bold.

Convergent validity

- Convergent validity was supported by a logical pattern of correlations with concurrent measures including moderate correlations with the FST, MLMT and VF scores post-baseline (r=-0.42 to 0.68; Table 3). No evidence of divergent validity between the mVFQ-25 and VA was found (r=-0.41 to -0.49; Table 3).

Table 3. Convergent/divergent validity correlations at baseline and 1-year follow-up (n=38)

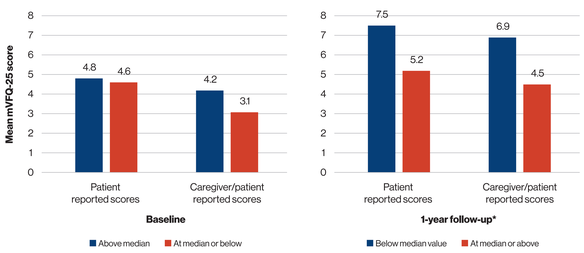
Modified VFQ-25 total score		FST	VA	MLMT*	MLMT†	VF HF	VF HM	VF G
Baseline	Patient reported scores	0.12	-0.44	0.33	0.46	0.22	0.36	0.06
	Caregiver/patient reported scores	-0.09	-0.43	0.42	0.33	0.42	0.47	0.05
1 year	Patient reported scores	-0.48	-0.49	**0.67**	**0.68**	**0.64**	**0.65**	0.41
	Caregiver/patient reported scores	-0.42	-0.41	**0.65**	**0.57**	**0.53**	**0.55**	0.46

*MLMT = Bilateral score; †MLMT = Assigned first eye score;
VF Humphrey-Foveal Sensitivity (HF); VF Humphrey-Mean Macula Threshold (HM); VF Goldmann average radius score (G)
Hypothesized correlations >0.50 in bold.

Known-groups analysis

- There were statistically significant differences in mVFQ-25 mean scores between known-groups defined by MLMT scores at 1-year for both patient and caregiver/patient reported scores (p<0.001; Figure 2).
- There were no significant differences for this comparison at baseline, likely due to the high floor effects.

Figure 2. Known-groups analysis comparing mVFQ-25 total score by MLMT group at baseline and 1- year follow-up (n=38)



*Indicates statistical significance at p<0.001.

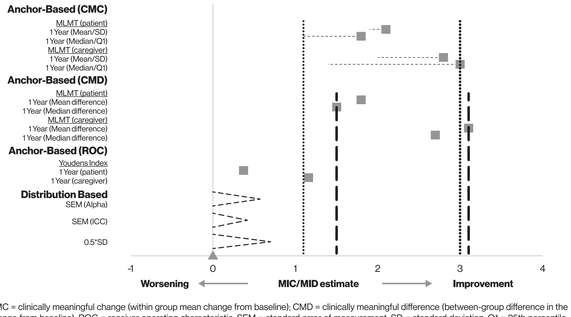
Ability to detect change

- There were statistically significant differences in the mVFQ-25 total score for both patient and caregiver/patient reported scores at baseline (p<0.001) and 1-year (p<0.05) between 'improved' and 'stable/worsened' groups defined using MLMT scores, indicating that the instrument can detect changes over time.

Distribution and exploratory anchor-based estimates of meaningful change thresholds

- Distribution-based methods using the 0.5 SD at baseline suggested that for the mVFQ-25 total score a mean of 0.73 or greater may be useful for guiding the lower threshold for clinically meaningful change, above measurement error.
- Exploratory anchor-based estimates showed that meaningful change thresholds for the mVFQ-25 total score could range from 1.5 to 3.1 using MLMT as an anchor (Figure 3).
- Individual-level responder definitions estimated using ROC curves suggested that a threshold of 1.5 was the optimal for distinguishing between improved and stable patients; 68.0%-77.0% of improved patients and 86.0%-100.0% of stable patients were correctly classified.
- Using the 1.5 threshold, the eCDF curve showed that 47.0%-71.0% more of the improved patients respectively achieved this particular score or higher in the improved group, compared to the stable group.
- In consideration of the above analyses, a meaningful change threshold of 1.5 on the mVFQ-25 total score was considered most appropriate. This threshold is at the lower end of the range of change scores and thus is more likely to reflect a minimum level of important difference.

Figure 3. Forest plot of distribution and anchor-based meaningful change estimates for mVFQ-25 total score



CMC = clinically meaningful change (within group mean change from baseline); CMD = clinically meaningful difference (between-group difference in the change from baseline); ROC = receiver operating characteristic; SEM = standard error of measurement; SD = standard deviation; Q1 = 25th percentile; med = median.

Limitations

- The sample size for this study was small, limiting the statistical power of the psychometric analyses. Therefore results should be interpreted with some degree of caution.
- Modifications to the original VFQ-25 were not based on prior qualitative research with patient input in this context of use and the mVFQ-25 lacks evidence of content validity, a key requirement for PRO development.[6]

Conclusions

- Findings support the psychometric validity of the mVFQ-25 in IRD patients with *RPE65* mutations. Results demonstrate that the mVFQ-25 total score provides a valid and reliable assessment of impacts on functional vision and is responsive to changes following treatment.

References

1. Astuti GD BM, et al. Comprehensive genotyping reveals RPE65 as the most frequently mutated gene in Leber congenital amaurosis in Denmark. *European Journal of Human Genetics*. 2016;24:1071-1079.
2. Russell S BJ, 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(10097):849-860.
3. Mahase E. NICE recommends gene therapy for inherited retinal dystrophy. *The BMJ*. 2019;366:l5394
4. Sousa MC BL, et al. Suitability of the Visual Field Index according to glaucoma severity. *Journal of Current Glaucoma Practice*. 2015;9(3):65.
5. Nunally J, Bernstein IH. Psychometric theory. *New York* 1994:248-292.
6. US Department of Health and Human Services. Food and Drug Administration. Guidance for Industry: Patient-reported outcome measures: use in medical product development to support labeling claims. 2009.

Disclosures: This study was sponsored by Novartis Pharma AG.