Test–Retest Intervisit Variability of Functional and Structural Parameters in X-Linked Retinoschisis

Brett G. Jeffrey*, Catherine A. Cukras*, Susan Vitale, Amy Turriff, Kristin Bowles†, and Paul A. Sieving

National Eye Institute, National Institutes of Health, Bethesda, MD
† Current address: College of Optometry, University of Houston, TX

Correspondence: Brett G. Jeffrey, Ophthalmic Genetics and Visual Function Branch, National Eye Institute, Bethesda, MD, USA. e-mail: jeffreybg@nei.nih.gov

Received: 28 March 2014
Accepted: 30 June 2014
Published: 3 October 2014

Keywords: XLRS; retinoschisis; electroretinogram; OCT; MP1


Purpose: To examine the variability of four outcome measures that could be used to address safety and efficacy in therapeutic trials with X-linked juvenile retinoschisis.

Methods: Seven men with confirmed mutations in the RS1 gene were evaluated over four visits spanning 6 months. Assessments included visual acuity, full-field electroretinograms (ERG), microperimetric macular sensitivity, and retinal thickness measured by optical coherence tomography (OCT). Eyes were separated into Better or Worse Eye groups based on acuity at baseline. Repeatability coefficients were calculated for each parameter and jackknife resampling used to derive 95% confidence intervals (CIs).

Results: The threshold for statistically significant change in visual acuity ranged from three to eight letters. For ERG a-wave, an amplitude reduction greater than 56% would be considered significant. For other parameters, variabilities were lower in the Worse Eye group, likely a result of floor effects due to collapse of the schisis pockets and/or retinal atrophy. The criteria for significant change (Better/Worse Eye) for three important parameters were: ERG b/a-wave ratio (0.44/0.23), point wise sensitivity (10.4/7.0 dB), and central retinal thickness (31%/18%).

Conclusions: The 95% CI range for visual acuity, ERG, retinal sensitivity, and central retinal thickness relative to baseline are described for this cohort of participants with X-linked juvenile retinoschisis (XLRS).

Translational Relevance: A quantitative understanding of the variability of outcome measures is vital to establishing the safety and efficacy limits for therapeutic trials of XLRS patients.

Introduction

X-linked juvenile retinoschisis (XLRS) is an inherited, early onset, and slowly progressive degeneration of the retina and macula with an estimated prevalence of 1:5000 to 1:25,000 males.1 XLRS is caused by mutations in the RS1 gene (RS1) at Xp22.2,2 which encodes the 24-kDa extracellular protein, retinoschisin (RS) that is secreted from retinal cells as a disulphide linked octamer. Retinoschisin binds both photoreceptor inner segments and bipolar cells and helps maintain the structural integrity of the retina.3 RS is present at the photoreceptor synapse and affects function.4

XLRS typically involves the macula with onset in the first decade of life, resulting in subnormal acuity.5–7 XLRS presents clinically with schisis of the macula in a spoke wheel pattern of cystic spaces in the plexiform and nuclear layers of the retina, which are readily imaged with optical coherence tomography (OCT).8 Schisis cavities may extend into the peripheral retina.9,10 The disease progresses slowly, with a decline in visual acuity over the first decades of life.6,7 The schisis collapses and progressive macular atrophy is observed in midlife, notably after the 4th decade.9,10

In addition to the loss of visual acuity, other functional changes include reduced central retinal sensitivity,11 and a variable but selective reduction of the ERG b-wave in the presence of a normal/near-normal ERG a-wave.10,12,13 Although strict genotype phenotype correlations have proven difficult in...
XLRS, missense and in-frame mutations in RS1 are generally associated with a milder phenotype, while mutations predicted to cause major structural change or eliminate the production of retinoschisin protein are associated with a more severe phenotype.10,13,14 Currently there is no United States Food and Drug Administration (FDA)-approved treatment for XLRS. Small, nonrandomized studies have reported that carbonic anhydrase inhibitors reduce intraretinal fluid schisis cavities in some individuals.15–21 While retinal thickness and macular cysts were reduced in some subjects, the response is not consistent, and in others showed no improvement or even an increase in retina thickness. For the most part, changes in cystic volume and retinal thickness have not been correlated with visual acuity.15,18,21

Preclinical studies in transgenic mouse models of XLRS recapitulate the phenotype seen in humans, including retinoschisis cavities, and a selective reduction of the ERG b-wave.22–24 Thereby, these transgenic mice provide a model to investigate potential treatments for XLRS. Mice deficient in Rs1h, the murine orthologue of the human RS1 gene, have been successfully rescued by delivery of adeno-associated virus (AAV)-RS1 constructs and demonstrate recovery of the ERG b-wave,4,23,25–27 resolution of schisis cystic structures and long term preservation of photoreceptors and retinal architecture.25–27 Park et al.28 achieved functional and structural rescue of the RS1-deficient mouse using a novel vector with AAV8 administered by intravitreal injection. In contrast, a study using the topical carbonic anhydrase inhibitor, Dorzolamide, on this mouse model of XLRS failed to show any benefit of either retina structure (OCT) or function (ERG).29

Multiple Phase I/II clinical trials involving RPE65 Leber’s congenital amaurosis patients have demonstrated the clinical feasibility and safety of AAV-mediated gene therapy for retinal disease in human subjects.30–32 Preparation for treatment trials of human XLRS patients requires the selection of suitable outcome measures with parameters to define both measures of safety and efficacy. Visual acuity, OCT measurement of central retinal thickness, and the ERG have all proven useful for describing disease severity in human XLRS.8,33,34 Retinal-guided perimetry enables measurement of central retinal sensitivity in maculopathies, including XLRS,11,13,36 even in patients with eccentric fixation or poor fixation stability. Perhaps because of the rarity of inherited retinal diseases and lack of available treatments, there is a paucity of data on the intersession variability of visual acuity, retinal-guided perimetry, and retinal thickness measurements in inherited retinal diseases.36–39 Intersession variability of the ERG is better documented but only for healthy subjects and those with retinitis pigmentosa.40–44 To our knowledge, the variabilities of these four outcome measures have not been investigated systematically in the XLRS population. Here, we define the repeatability coefficients (RCs) and associated 95% confidence intervals (CIs) for visual acuity, ERG, retinal sensitivity, and central retinal thickness in a cohort of participants with XLRS.

Materials and Methods

Subjects

Seven men aged 19 to 49 years with confirmed mutations in RS1 (see Results), were enrolled from the Ophthalmic Genetics clinic at the National Eye Institute (NEI) over a 9-month period. After an initial baseline visit, each patient returned three times: at 1, 3, and 6 months. This study adhered to the tenets of the Declaration of Helsinki for research involving human subjects and was approved by NEI institutional review board. Each subject gave written informed consent after an explanation of the nature and possible consequences of the study.

Visual Acuity

Best corrected visual acuity (BCVA) was measured using the Early Treatment of Diabetic Retinopathy (ETDRS) chart and recorded as the number of letters read.45

MP1 Retinal Sensitivity

Following pupil dilation, retinal sensitivity was measured from both eyes with a microperimeter (MP-1; Navis Software, ver. 1.7.6; Nidek Technologies, Padova, Italy). Sensitivity was measured across a 68-point grid covering a 20° field centered on the fovea (similar to the Humphrey visual analyzer 10-2 protocol). With the infrared-guided fundus viewing system of the MP1, we were not able to clearly visualize the fovea in these patients. Therefore, the fovea was taken to be the point approximately 2-disk diameters temporal to and 1/3-disk diameter below the center of the optic disk.46 At each point in the grid, sensitivity was measured for a white stimulus 0.43° in diameter (Goldmann size III) presented for 200 msec against a mesopic background (1.27 cd/m²).
Threshold at each point was determined by using a 4-2 staircase. The infrared automatic eye tracking system of the MP1 was used to compensate for eye movements and ensure correct localization of stimulus presentation during the exam. The “follow-up” feature of the MP1 was used to enable sensitivity measurements at the same retinal locations across all four visits.

**Retinal Sensitivity Analysis**

All sensitivity data were exported and analyzed in Excel (Microsoft Corp., Redmond, WA). As many XLRS subjects have substantially reduced sensitivity in the central macula, the method of Chen et al. was used to examine the variability of mean sensitivity over three different retinal areas, and the following parameters were computed: (1) mean macular sensitivity (MS), the mean of all 68 loci, (2) mean central macular sensitivity (CMS), the mean sensitivity of the central 16 loci (4 × 4 grid) enclosed by a circle with a 10° diameter, and (3) mean paracentral macular sensitivity (PMS), the mean sensitivity of the remaining 52 loci outside the central 10°. The variability of individual points, or point-wise sensitivity (PWS), was also examined.

A known limitation of the MP1 is its relatively narrow luminance range (1.27–127 cd/m²) and both floor (0 dB, not seen) and ceiling (20 dB) effects were observed. Since we were interested in variability of the MP1, points that reached the floor or ceiling limits during any of the four visits were excluded from the analysis of variability.

**Electroretinography**

Following MP1 testing, subjects were dark-adapted for 40 minutes prior to the start of ERG recording. International Society for Clinical Electrophysiology of Vision (ISCEV) standard full-field flash ERGs were recorded from corneal bipolar Burian-Allen electrodes using a commercial electrophysiology system (LKC, Gaithersburg, MD). An Ag/AgCl electrode placed on the forehead served as ground.

**Optical Coherence Tomography**

Spectral-domain optical coherence tomography (SD-OCT; Cirrus HD-OCT; Carl Zeiss Meditec, Dublin, CA) captured a 512 × 128 scan pattern with the center of the 6 × 6-mm scanning area positioned at the center of the macula. Quantitative longitudinal analysis of OCT scans was performed by first aligning the scans spatially using functions provided within the OCT instrument software and then checking for accuracy. The accuracy of automated delineations of the inner and outer retinal boundaries was also verified manually. OCT retinal thickness measurements in the macula were analyzed by using a circular ETDRS-type grid positioned on the center of the fovea. Mean thickness measurements were calculated for the central subfield (central circle of diameter 1 mm) and for the four “inner” quadrants (circumscribed by a circle 3 mm in diameter, concentric to the central region and divided into superior, inferior, nasal, and temporal quadrants). For the analysis of variability, retinal thickness was log transformed (logOCT) by taking the log base 10 of the ratio of retinal thickness divided by 200.

**Quantifying Repeatability**

A quantitative estimate of repeatability was made for each parameter based on the methods outlined by Fleiss and by Bland and Altman. We tested subjects four times over a 6-month period and therefore, we used a one-way random effects model. Repeatability was quantified from the four replicated measurements by using a one-way ANOVA, with subject as the factor. From first principles, the within-subject standard deviation, $s_{\omega}$ is defined by:

$$s_{\omega} = \sqrt{\frac{\sum_{i=1}^{N} (k_i - N)^2}{\sum_{i=1}^{N} k_i - N}}$$  \hspace{1cm} (1)$$

where the variance, $s_i^2$ for the $i$th subject is calculated from $k$ replicate measures; $N$ is the total number of subjects tested. Thus, the total within-subject variance ($s_{\omega}^2$ in Equation 1) is the average of the individual variances weighted by the number of measurements for each subject.

For an actual single measurement, $x$ from a subject, the 95% CI for a true single measurement is given by:

$$x \pm 1.96 s_{\omega}$$  \hspace{1cm} (2)$$

where $s_{\omega}$ is the within-subject SD calculated in Equation 1. Since the mean difference between replicates is expected to be zero, the 95% CI for the true difference between 2 measurements on a subject is then:

$$0 \pm 1.96 \sqrt{2} s_{\omega} \text{ or } 0 \pm 2.77 s_{\omega}$$  \hspace{1cm} (3)$$

where $\sqrt{2}$ factor accounts for the compounding of the
95% CIs (Equation 2) across two measurements.\(^{48,49}\) Bland and Altman\(^{49}\) refer to the range defined by Equation 3 as the repeatability coefficient (RC). Simply stated, if the second measurement increases/ decreases by more than 2.77 \(s_{\text{w}}\) from the first measurement, then significant change has occurred for a given subject (with a 5% error rate).

**Determining the Robustness of the Repeatability Coefficient**

The method described above provides no estimate of the certainty we can place in the derived RCs, calculated from the variability data of seven subjects. The 95% CI for each RC was calculated using a jackknife resampling procedure.\(^{50,51}\) Norcia et al.\(^{52}\) described this procedure for small data sets in detail. Briefly, a new sampling distribution was defined using subsets of the original data from which the 95% CI of the RC was calculated.

The first step involved calculation of within subject SD, \(s_{\text{w}}\) (Equation 1) on the entire data set. Subsets of data were then created, with one subject excluded from each subset. Within-subject SD was calculated for each of the seven subsets (\(\tilde{s}_{\text{w}}\)), where the subscript \((i)\) denotes the subject excluded. Finally, a set of \(n\) “pseudo within-subject SDs” was calculated as follows:

\[
\tilde{s}_{\text{w}} = n s_{\text{w}} - (n - 1) \tilde{s}_{\text{w}}
\]  
(Equation 4)

where \(\tilde{s}_{\text{w}}\) was the \(i^{th}\) pseudo within-subject SD, \(n = \) total number of subjects and \(\tilde{s}_{\text{w}}\) and \(s_{\text{w}}\) are as defined above. The pseudo within-subject SDs (\(\tilde{s}_{\text{w}}\)) are normally distributed.\(^{50}\) The mean (\(\bar{s}_{\text{w}}\)) of the pseudo values calculated in Equation 4 will be similar to the within-subject SD for all subjects, \(s_{\text{w}}\) (Equation 1). The standard error (\(se_{\bar{s}_{\text{w}}}\)) of the pseudo values is used to define the 95% CI of a RC, that is

\[
\pm 1.96se_{\bar{s}_{\text{w}}}
\]  
(Equation 5)

In summary, the RC (Equation 3) defines the range by which two measurements on a subject can differ due to random variability. Equation 5 then defines 95% CIs for this range.

**Examination of Parameter Stability Over Time**

Fundamental to the calculation of the RCs (above) is the hypothesis that visual function and central retinal thickness did not systematically change in these subjects during the 6-month period of the study. To determine whether there were any trends in parameters over time, we used a series of repeated measures models to account for the correlation among visits for a patient (version 9.2, SAS Proc Mixed, compound symmetry option; SAS Institute, Cary, NC). Both a linear and a polynomial model (to allow for non-linearity over visits) were used, and subsequently we examined whether the common slope estimate for all subjects combined (ensemble fit) was significantly different from zero. Analyses were done separately for the two eyes which were strongly correlated for all parameters measured (0.7–0.9).

**Results**

**Clinical Details**

The clinical details of the seven subjects are given in Table 1. All subjects were unrelated with the exception of subjects four and five who are maternal first cousins. Subjects one through six all had missense mutations. Subject seven had a duplication of nucleotides 306 to 308 (GCT), which creates an insertion of an additional amino acid leucine at position 103.

Visual acuity was different between the two eyes of all seven subjects. Each eye from a given participant was considered separately, and two analysis groups were created based on the eyes with “better” or “worse” visual acuity at baseline. The RCs were determined separately for the two groups.

**Investigation of Change in Parameters With Time**

No significant relationship with time was observed for any parameter, irrespective of the model used (Table 2). Based on these analyses we cannot reject the null hypothesis that visual function and central retinal thickness did not change in these seven subjects during the 6-month period of the trial.

**Variability of Visual Acuity**

Figure 1 shows the replicate visual acuity measurements from the four visits for each subject. Within-subject variability is relatively small compared with the differences in visual acuity between subjects. Figure 1C and 1D highlight that change in acuity relative to baseline remains constant across the three follow-up visits, consistent with the hypothesis that visual acuity did not significantly change over the 6-month duration of the study.
Calculation of Repeatability Coefficient

Table 3 provides a detailed example of the calculation of the RC for visual acuity from the Worse Eye group. For each subject, raw visual acuity data are shown along with the variance ($s^2_i$) and number of measurements. The within-subject SD ($s_x$) of 2.0 letters was calculated from the weighted average of individual variances (Equation 1). From Equation 3, the RC for Worse Eye letter acuity was 5.7 letters. The practical implication of this result is that an increase or decrease in visual acuity of 6 letters or more at follow-up would be considered significantly different from baseline. For the Better Eye group, the RC for visual acuity was 3.7 letters (Table 4). The dashed lines in Figure 1C and 1D show the ranges defined by the RCs and illustrate the greater variability in acuity measurements from the Worse Eye group.

Table 3 also details the jackknife resampling procedure for calculating the 95% CI for the RC (see Methods). The first step involved calculation of within-subject SD for each subset, with one subject excluded (Table 3, column 5). For example, $\hat{s}_x^1$ was calculated using the data from subjects two through seven, $\hat{s}_x^2$ was calculated for subjects one, and three through seven, and so on. The pseudo values, $\hat{s}_x^i$ (Table 3, column 6) and corresponding RCs (column 7) for each of these subsets were then calculated using Equations 4 and 3, respectively. For these RC pseudo values, mean $\hat{s}_x$ was 5.9 letters with a standard error ($se_{\hat{s}_x}$) of 1.1 letters (Table 3). By definition, the 95% CI for $\hat{s}_x$ is 2.1 letters (1.96 $se_{\hat{s}_x}$). These results indicate that the true repeatability coefficient for Worse Eye visual acuity lies in the range of 3.8 to 8.0 letters (i.e., 5.9 $\pm$ 2.1 letters). For the Better Eye group, the true RC lies in the range between 2.8 to 4.6 letters (Table 4). The dotted lines in Figure 1C and 1D show the 95% CIs (dotted lines) for each RC. As expected, the 95% CIs overlap for the two groups.

Variability of the Electroretinogram

Consistent with previous reports, ERG a-wave amplitudes of all seven subjects were within the normal limits and remained relatively constant across the 6-month test period (data not shown). The RCs and associated CIs were similar for the Better and
Worse Eye groups (Table 4). Figure 2 shows scotopic mixed rod-cone ERGs from our seven XLRS subjects at their baseline visit (A: Better eyes, B: Worse eyes). Two important characteristics of the ERG b-wave in XLRS subjects are observed in Figure 2. First, while four subjects exhibited a classic electronegative ERG in which b-wave amplitude was decreased below that of the a-wave; the other three subjects had substantial

Table 3. Jackknife Estimation of Repeatability Coefficients for Worse Eye Visual Acuity

<table>
<thead>
<tr>
<th>Subject (i)</th>
<th>Visual Acuity by Visit (Letters)</th>
<th>Repeatability</th>
<th>Jackknife Resampling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$k_i$</td>
<td>$s^2_i$ (Letters)</td>
</tr>
<tr>
<td>1</td>
<td>$-, 48, 49, 48$</td>
<td>3</td>
<td>0.33</td>
</tr>
<tr>
<td>2</td>
<td>$30, 23, 25, 26$</td>
<td>4</td>
<td>8.67</td>
</tr>
<tr>
<td>3</td>
<td>$55, 57, 59, 58$</td>
<td>4</td>
<td>2.92</td>
</tr>
<tr>
<td>4</td>
<td>$55, -, 54, 54$</td>
<td>3</td>
<td>0.33</td>
</tr>
<tr>
<td>5</td>
<td>$68, 75, 71, 70$</td>
<td>4</td>
<td>8.67</td>
</tr>
<tr>
<td>6</td>
<td>$61, 59, 61, 63$</td>
<td>4</td>
<td>2.67</td>
</tr>
<tr>
<td>7</td>
<td>$35, 34, 35, 38$</td>
<td>4</td>
<td>3.00</td>
</tr>
</tbody>
</table>
### Table 4. Repeatability Coefficients

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Better Eye</th>
<th>Worse Eye</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RC ± 95% CI</td>
<td>RC % Change (% Range)</td>
</tr>
<tr>
<td>Visual acuity letters</td>
<td>3.7 ± 0.9</td>
<td>2.8–4.6</td>
</tr>
<tr>
<td>Scotopic ERG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a-wave amplitude (log µV)</td>
<td>0.241 ± 0.116</td>
<td>43% (25–56)</td>
</tr>
<tr>
<td>b-wave/a-wave ratio*</td>
<td>0.24 ± 0.20</td>
<td></td>
</tr>
<tr>
<td>Photopic ERG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 Hz amplitude, log µV</td>
<td>0.285 ± 0.093</td>
<td>48% (36–58)</td>
</tr>
<tr>
<td>Retinal Sensitivity MP1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macular sensitivity, dB</td>
<td>2.2 ± 0.61</td>
<td>40% (31–48)</td>
</tr>
<tr>
<td>Central macular sensitivity, dB</td>
<td>3.3 ± 0.75</td>
<td>53% (44–61)</td>
</tr>
<tr>
<td>Paracentral macular sensitivity, dB</td>
<td>2.1 ± 0.73</td>
<td>39% (27–48)</td>
</tr>
<tr>
<td>Point–wise sensitivity, dB</td>
<td>6.8 ± 3.6</td>
<td>78% (38–92)</td>
</tr>
<tr>
<td>Central retinal thickness, logOCT</td>
<td>0.107 ± 0.057</td>
<td>22% (11–31)</td>
</tr>
</tbody>
</table>

* Linear parameter, percentage change will depend on absolute value of b/a ratio.

---

**Figure 2.** ERGs. Scotopic full-field ERGs recorded at baseline to an ISCEV standard flash from the Better (A) and Worse Eye (B) groups. *Red traces* show the ERG response from a representative healthy subject. The *blue traces* are the average ERG from two subjects with genetically confirmed congenital stationary night blindness type 1 (CSNB1). *Vertical dotted lines* at 47 msec indicate mean b-wave implicit time for healthy subjects with our ERG system.

http://tvstjournal.org/doi/full/10.1167/tvst.3.5.5
b-waves, indicating some level of signaling from the photoreceptors to the second order neurons. Second, for all seven subjects, the ERG b-waves were broad and the peaks poorly defined. For this reason we measured b-wave amplitude at the mean b-wave implicit time for healthy subjects with our ERG system (47 msec post flash; vertical dotted lines).

The blue lines in Figure 2 are the average ERG from two subjects with genetically confirmed congenital stationary night blindness type 1 (CSNB1) who have impaired on-pathway bipolar signaling,53,54 and thus represent a b-wave “floor” for comparison. By corollary, there is minimal or no signaling from photoreceptors to ON-bipolar cells in four of the eyes in Figure 2B.

The variability of the ERG b/a ratio was substantially less than for absolute ERG amplitudes (Table 4). The repeatability coefficient and associated 95% CIs of the b/a ratio were considerably larger for the Better Eye group compared with the Worse Eye group (Fig. 3; Table 4). This is likely due to the low variability of the four eyes in Figure 2B (Worse Eye group) with “b-waves” that are similar to the floor response, which in turn minimizes their contribution to the variability of the b/a ratio. Based on the data from these four subjects, the floor for the b/a ratio in XLRS is approximately 0.6.

The variability of light adapted 30-Hz flicker response was small and similar to that of the ERG a-wave (Table 4). As such, the 30-Hz response may prove useful as another measure of safety.

Variability of Central Retinal Sensitivity Measured with Retinal Guided Perimetry

Figure 4 shows retinal sensitivity maps recorded 6 months apart from the Worse VA eye (20/80) of

http://tvstjournal.org/doi/full/10.1167/tvst.3.5.5

TVST | 2014 | Vol. 3 | No. 5 | Article 5

Downloaded from iovs.arvojournals.org on 10/06/2019
subject three. Although his fixation shifted by 3°, and fixation stability had deteriorated, the maps show that retinal sensitivity was measured from almost identical areas of the retina. These maps highlight the importance of the retinal tracking and follow-up features of the MP1 microperimeter in measuring retinal sensitivity in XLRS subjects, many of whom have eccentric fixation and/or poor fixation stability. Three other eyes had noticeable changes in fixation during the course of the study without major changes in visual acuity. Such changes in fixation likely reflect a shift in preferred retinal locus to a nearby retinal region with similar acuity.

The tracking features of the MP1 also contribute to minimizing variability in mean macular sensitivities. The repeatability coefficients for mean macular and paramacular sensitivity were in the range of 2 dB for both Better and Worse eye groups (Table 4). Variability in mean sensitivity across the central 10° (CMS) was similar for the Worse Eye group but higher (3.3 dB) in the Better Eye group (Table 4). No correlation was found between fixation stability and reliability (false positive responses) in our cohort of seven subjects (data not shown).

The repeatability coefficients for point-wise sensitivity were considerably higher than obtained for mean sensitivities (Table 4). Despite the tracking features of the MP1, slight changes in the location of test points can occur between visits (white arrows in Figure 4). We have also observed instances where the MP1 tracking briefly locked onto different features during real-time tracking. Such erroneous tracking is observed as fixation points in another part of the retinal field separated from the main cluster of fixation points (not shown). Misalignment between visits and/or erroneous real-time tracking likely account for the much higher variability in point wise sensitivity (PWS). Taking into account the 95% CIs of the repeatability coefficients, point-wise sensitivity could vary by 10.4 dB between visits for the Better Eyes and by 7.0 dB for Worse Eyes.

Figure 4. MP1 microperimetry maps. Microperimetry measured at baseline (A) and 6 months later (B) from the Worse Eye of subject 3. The arrows highlight slight changes in the location of test points relative to nearby blood vessels. The bivariate contour ellipses in each graph describe the variation in the fixation (blue dots) during testing.
Figure 5 shows a representative example of the variation observed in central retinal thickness in XLRS subjects. At baseline central retinal thickness measured 538 μm (0.405 logOCT). Three months later central retinal thickness had increased 16% in this subject to 622 μm (0.462 logOCT) but by the 6-month visit, had returned to essentially the same level measured at baseline (534 μm; 0.393 logOCT). The variability of central retinal thickness was higher for the Better Eye group as indicated by the larger values for both RC and associated 95% CIs (Table 4). The lower variability in the Worse Eye group may represent a floor effect due to the collapse of the schisis pockets and/or retinal atrophy in these eyes.

Discussion

Here, we have outlined a robust method that provides a quantitative description of the variability of an outcome measure, and in addition quantifies the 95% CI for this variability measurement for a small cohort of patients. The coefficients of repeatability and associated CIs provide the basis of defining the minimum change required to assess safety and efficacy of a treatment. We emphasize that these quantitative descriptions of variability alone do not necessarily define clinically meaningful change needed to address safety and efficacy issues. Instead, these RCs define the minimum level of change required in a parameter to be considered statistically different from baseline. For example, the analysis of Worse Eye acuity indicated that a change of between four and eight letters at follow-up could be used as evidence of significant change from baseline. Grover et al. reported a similar criterion of seven letters for significant change in visual acuity for subjects with retinitis pigmentosa. However, a clinically meaningful change in visual acuity is frequently taken to be 15 letters (i.e., doubling of the visual angle).55

To our knowledge, no previous study has examined variability of ERG a-wave to the ISCEV standard maximal response flash. For the XLRS subjects in this study, the RCs for the ERG a-wave indicated that the criterion for a significant reduction in a-wave amplitude was approximately 40%. When the 95% CIs were taken into account, the criterion increased to a reduction of 50% to 56% for ERG a-wave amplitude. In previous studies, examining variability of the dim flash ERG b-wave in retinitis pigmentosa patients and control subjects, the criteria for significant reduction in ERG b-wave amplitude ranged from 31% to 51%.40,41,43,44 Lower short-term variability (< 2 weeks) in healthy subjects was obtained by carefully controlling recording conditions,44 and considerably lower thresholds for significant reductions of rod (23%) and cone (37%) a-waves in response to high intensity flashes have been reported.42 Variability for 30-Hz flicker amplitude is typically higher than for the scotopic b-wave.42,43 These combined results suggest that a 50% reduction in amplitude may be a suitable criterion for a clinically meaningful change in the ERG a-wave recorded to a standard ISCEV maximal flash.

As for the ERG b-wave, to our knowledge, no previous study has examined variability of ERG b/a ratio to the ISCEV standard maximal response flash. The ERG b/a wave ratio serves as an excellent parameter by which to judge efficacy in a clinical trial for XLRS. Successful treatment of XLRS could be expected to improve signaling from photoreceptors to the bipolar cells, which in turn would be reflected as an increase in the ERG b/a ratio. The upper limits in the criteria for significant change in the ERG b/a ratios
were 0.44 for the Better Eye and 0.23 for the Worse Eye groups. Examination of the criteria for significant change in ERG b/a ratio may help define eligible patients for a clinical trial. For our system, mean b/a ratio for a healthy 25-year-old subject is 1.88. Allowing for the maximum criterion for significant improvement of 0.44 for the ERG b/a ratio, candidates for a treatment trial to evaluate efficacy would ideally have a maximal b/a wave ratio of less than or equal to 1.44.

In the present study, four of the subjects had no b-wave response, consistent with an absence of photoreceptor to bipolar cell signaling, and the floor for the b/a ratio was estimated at 0.6. Other studies have shown smaller b/a ratios of 0.3, as in four affected male relatives, age 32 to 45 years, with a mutation causing null-RS expression, and this parameter may change slowly with time. Therefore, the ERG b-wave is likely not a good candidate for assessing safety in a clinical trial of XLRS.

Defining what constitutes a clinically meaningful change in retinal sensitivity is perhaps less intuitive than for visual acuity and ERG measures. For MP1 mean macular sensitivity, the repeatability coefficients of 1.7 to 2.2 dB reported herein are similar to the values reported for age-related macular degeneration (AMD), ABCA4 retinopathy, and for intrasession variability in patients with unspecified macular disease. However, mean sensitivity may not be the most appropriate measure unless a treatment is expected to affect a large section of the retina, such as the central 20° field recorded with the 10-2 pattern. To that end, we examined mean sensitivity across the central 10°, which exhibited similar, albeit slightly higher, variability (Table 4). The disadvantage of mean sensitivity is that spatial information is lost. It is possible that a treatment for XLRS may restore function only at the margins of the central scotoma. Under this scenario, the sensitivity of some individual points may change substantially without affecting overall mean sensitivity. To account for this possibility, an understanding of the variability of point-wise sensitivity is important. The coefficients of repeatability for PWS in our XLRS patients were 6.8 and 5.4 dB for the Better and Worse Eye groups, respectively. These values are consistent with the 5.6 dB reported by Chen and colleagues in an intrasession variability study of 50 maculopathy patients. Cideciyan et al. reported a slightly lower repeatability coefficient of 4.2 dB for PWS in ABCA4 retinopathy with the MP1, but this lower value likely reflects the smoothing of individual thresholds by applying a 3-point spatial moving average. Wu et al. similarly found a lower coefficient of repeatability for PWS of 4.2 dB in AMD patients using a macular integrity assessment (MAIA) fundus-guided perimeter that uses a scanning laser ophthalmoscope for fundus tracking. Of concern are the large confidence intervals (CIs) in the RCs for our PWS data, which extends the criteria for significant change of an individual point to 8.0 dB for the Worse Eye and 10.4 dB for the Better Eye groups. The MP1 has a stimulus range of 20 dB and our results indicate that a change of up to one-half the range is required to be considered significant. A possible alternative would be the use of scotopic two-color microperimetry that extends the range of the MP1 to 50 dB by incorporating measurement of rod function.

XLRS typically involves the macula although schisis cavities may extend into the peripheral retina, an area that could be examined in XLRS patients using standard automated perimetry (SAP). Coefficients of repeatability for PWS with SAP compare with the values reported here for XLRS and range from 5 to 7 dB in diabetics, glaucoma, and retinitis pigmentosa. However, PWS variability of SAP dramatically increases with retinal eccentricity and with decreasing retinal sensitivity, likely reflecting unsteady fixation or changes in eccentric fixation locus that cannot be compensated for with standard perimetry. Kinetic perimetry provides a relatively precise measurement of the edge of scotomas in retinal disease. Bittner et al. reported a favorable 43% coefficient of repeatability for intrasession change in planimetric area in retinitis pigmentosa. The disadvantage of kinetic perimetry is that spatial information about the variability of sensitivity within the seeing area is lost and variability increases substantially in subjects with poor acuity and narrow visual fields. In the present study, we used fundus-guided perimetry to examine retinal sensitivity across the central 20° where schisis cavities are most prevalent in XLRS. The major advantages of retinal-guided perimetry over SAP and kinetic perimetry for measuring retinal sensitivity in maculopathies such as XLRS, include (1) the ability to compensate for eccentric and/or poor fixation stability, (2) the follow-up feature that enables the exact same areas of the retina to be tested longitudinally in a clinical trial, and (3) the ability to study structure–function correlations by overlaying retinal images with maps of retinal function. Changes in peripheral retina function are not likely to be localized following intravitreal delivery of treatment and may well be reflected by changes in the full-field ERG.

In our XLRS patients, the criteria for significant change in central subfield retinal thickness (0.107 logOCT) corresponded to a 22% decrease or 28%
increase in thickness. That is, for a XLRS subject with central retinal thickness of 500 µm at baseline, a decrease of greater than 110 µm or an increase greater than 140 µm at follow-up would be considered significant. A retrospective study of diabetics with refractory diabetic macular edema (DME) similarly reported central retinal thickness varied by up to 28% of median thickness over a 7-month period.⁶⁴ In contrast, for diabetic subjects without macular edema or with regressed macular edema, OCT measurements of central subfield retinal thickness varied by less than 11% of median thickness over periods extending from 17 to 22 months.⁶⁴,⁶⁵

In summary, an appreciation of range of the true repeatability coefficient for a parameter is fundamental to setting the minimum safety and efficacy limits for clinical trials. The 95% CI of the RC is crucial when variability is derived from a small number of subjects as in our study. Our derivation of variability, which was based on four visits across 6 months is a more general form of the classical test–retest paradigm and provides the additional advantage of assessing variability as a function of time. This is an important consideration in designing clinical trials where change in a parameter will be measured for an extended period relative to baseline.

**Acknowledgments**

Supported by the National Institutes of Health Intramural Research Programs of the National Institute on Deafness and Other Communication Disorders and the National Eye Institute.

Disclosure: **B.G. Jeffrey**, None; **C.A. Cukras**, None; **S. Vitale**, None; **A. Turriff**, None; **K. Bowles**, None; **P.A. Sieving**, None

* BGJ and CAC contributed equally to the work presented here and should therefore be regarded as equivalent authors.

**References**


