Interseession Repeatability of Humphrey Perimetry Measurements in Patients with Retinitis Pigmentosa

Linda S. Kim, J. Jason McAnany, Kenneth R. Alexander, and Gerald A. Fishman

PURPOSE. To determine the relationship between interseession test repeatability in static perimetry and the degree of local sensitivity reduction in patients with retinitis pigmentosa (RP).

METHODS. Visual field data were obtained from 27 patients with RP using FASTPAC 30–2 of the Humphrey Field Analyzer and stimulus sizes III and V. Each test was repeated at two subsequent visits after an initial practice session. An interseession sensitivity difference was defined for each patient at each measured point in the visual field by subtracting the sensitivity value obtained during session 2 from the value obtained during session 3 at the corresponding test location. In addition, the mean sensitivity value for sessions 2 and 3 was determined at each location for each patient.

RESULTS. Bland-Altman plots of intersession sensitivity differences as a function of mean sensitivities showed that test-retest variability was greatest at test regions with sensitivities of approximately 10 dB, with lower variability at locations with higher or lower sensitivity. The pattern of findings was similar for sizes III and V stimuli. The mean interseession sensitivity difference was +0.26 dB for stimulus size III and +0.45 dB for stimulus size V, representing a small improvement in sensitivity from session 2 to session 3.

CONCLUSIONS. Intersession repeatability of visual field measurements in patients with RP is nonmonotonically related to the magnitude of the sensitivity loss. Therefore, apparent changes in visual field sensitivity as a result of time or treatment should be evaluated in light of the degree of the expected variability at particular sensitivity levels. (Invest Ophthalmol Vis Sci. 2007; 48:4720–4724) DOI:10.1167/iovs.07-0690

Retinitis pigmentosa (RP) is a group of progressive retinal degenerations characterized by night blindness, retinal pigmentary changes, subnormal or nondetectable electroretinographic amplitudes, and progressive peripheral visual field loss. The status of the visual field in patients with RP is typically evaluated by kinetic or static perimetry, sometimes both. Kinetic perimetry has characteristically been used to illustrate patterns and rates of visual field change in patients with RP.1–6 Static perimetry emphasizes sensitivity changes at specific locations within the central visual field.

A major consideration in evaluating the results of automated static perimetry is the degree of repeatability of the visual field sensitivity measurements. This is a particularly important factor in the design and interpretation of future clinical trials of potential therapeutic agents. In particular, it is not apparent whether test-retest repeatability in patients with RP is related to the degree of sensitivity loss, as it is in patients with glaucoma.7,8 For example, Artes et al.7 reported that intersession variability was greatest at a sensitivity of approximately 7 dB in patients with glaucoma but was lower for visual field locations with higher and lower sensitivities. Similarly, Wyatt et al.8 showed that intersession sensitivity differences in patients with glaucoma were small for visual field locations with high and low sensitivity, whereas sensitivity differences were large for field locations with intermediate sensitivity. In comparison, Seiple et al.9 examined the variability of visual field sensitivity in eight patients with RP, with variability defined as the overall standard deviation (SD) of the measurements on each of four visits. They observed that four of the eight patients with RP had SDs that were greater than those of the control subjects on at least two of the four testing sessions. However, they reported that the patients’ repeatability was not related to their mean sensitivity.

The purpose of the present study was to evaluate, using the Humphrey Field Analyzer (HFA), the short-term repeatability of visual field sensitivity as a function of sensitivity loss at individual points within the visual field in a group of patients with RP. Because Swanson et al.10 reported that the change in perimetric sensitivity with stimulus size in patients with RP was different for sizes III and V stimuli, we evaluated the degree of test-retest repeatability for these two stimulus sizes.

METHODS

Visual field data were obtained from 28 patients (13 men, 15 women) with typical RP. However, the data from one female patient were excluded from analysis because of excessive fixation loss and a high false-positive rate. The mean age of the 27 patients included in the analysis was 40 years (range, 18–64 years), and the genetic types were autosomal dominant (n = 2), autosomal recessive (n = 8), X-linked recessive (n = 1), and isolated with no other known family members affected (n = 16). Mean visual acuity of the patients was 0.11 log MAR (minimum angle of resolution; 20/25 Snellen equivalent), and the range was −0.10 to +0.56 log MAR (Snellen equivalent, 20/16 to approximately 20/65).

At the initial testing session, the patients underwent routine eye examination to determine eligibility for entry into the study; this included visual acuity testing using the Lighthouse chart (Lighthouse Low Vision Products, Long Island City, NY), slit lamp biomicroscopy, measurement of intraocular pressure, and fundus examination. Patients’ refractive errors ranged from −8.00 to +2.50 diopters, and their visual fields were greater than 20° horizontally to the Goldmann III stimulus using kinetic perimetry. Lens status ranged from no opacity to grade 1+ posterior subcapsular opacity, though four patients (two men, two women) had posterior chamber intraocular lens implants in each eye. All patients had normal intraocular pressure, no family history of glaucoma or diabetes mellitus, and open anterior chamber angles. None was taking medication known to affect the visual field.
and all were willing and available to take part in the study for the entire three sessions. Patients with syndromic forms of RP, such as Usher or Bardet-Biedl syndrome, were excluded from the study, as were RP patients with diabetic retinopathy, glaucoma, and nystagmus. All patients underwent fundus photography and optical coherence tomography (Stratus OCT 3; Carl Zeiss Meditec Inc., Dublin, CA) to exclude those with cystoid macular edema.

At the first session, the eye with the better Goldmann visual field was examined with the HFA Il-2 series and the FASTPAC algorithm (Carl Zeiss Meditec Inc.) using Program 30-2 with stimulus sizes III and V. A rest period of approximately 5 minutes was given between the two stimulus size programs. Thirteen of the 27 patients had been previously tested with the HFA, and all patients were experienced in Goldmann visual field testing. The distance refractive correction in the form of a full aperture trial lens and any near correction appropriate for the viewing distance of the perimeter bowl were used. The nontested eye was occluded with an opaque patch. In addition to the standard Heijl-Krakau technique (checking for fixation losses when a stimulus is presented within the blind spot), fixation was monitored continuously with the gaze tracker and video monitor. The first session was considered a familiarization period for the study, and the results were discarded before data analysis to reduce the influence of a learning effect. The protocol at the second and third sessions was identical to that at the first session. Sessions were each separated by no more than a 14-day interval and were carried out within a maximum period of 6 weeks.

The study adhered to the tenets of the Declaration of Helsinki and was approved by a University of Illinois at Chicago institutional review board. Informed consent was obtained from each patient after the nature of the procedures had been explained. Examinations were conducted in accordance with Health Insurance Portability and Accountability Act regulations.

Analysis

Results for left eyes were converted to a right eye format. Measurements made at stimulus locations immediately above and below the physiological blind spot were excluded from the analyses. Additionally, to preclude a floor effect,9 measurements at locations that had a sensitivity of 0 dB at one or both sessions were excluded. For test locations at which two sensitivity measurements were available, the mean of the two measurements was used as the sensitivity value.

The agreement between the sensitivity values obtained in sessions 2 and 3 was analyzed using the approach of Bland and Altman.11,12 First, a sensitivity difference in decibels was defined for each patient at each measured point in the visual field by subtracting the sensitivity value obtained during session 2 from the value obtained during session 3 at the corresponding test location. Second, the mean sensitivity difference for sessions 2 and 3 was determined at each measured test location for each patient. Then, for each patient and test location, the sensitivity difference was plotted against the mean sensitivity. The Bland-Altman analysis of intersession sensitivity agreement was performed separately for stimulus sizes III and V.

Bland-Altman analysis was also used to examine the effect of stimulus size on sensitivity, in accordance with the approach of Swanson et al.9 For this analysis, the sensitivity at each test location for a given stimulus size was defined as the mean sensitivity of sessions 2 and 3. The size effect at each test location was determined by subtracting the sensitivity value obtained with the size III stimulus from the value obtained with the size V stimulus. The mean sensitivity value for size III and V stimuli was also obtained at each test location. The size effect was plotted with respect to the mean sensitivity for the two stimulus sizes for all patients and test locations.

Results

Figure 1 plots sensitivity differences between sessions 2 and 3 as a function of mean sensitivity for the size III stimulus, for all subjects and test locations. Results obtained with the size V stimulus are presented in Figure 2. The overall mean sensitivity difference pooled across patients and test locations was $+0.26$ dB for the size III stimulus and $+0.45$ dB for the size V stimulus. These values represent a slight, but statistically significant, improvement in sensitivity from session 2 to session 3 for both stimulus sizes ($t = 2.66, P < 0.01$, and $t = 5.49, P < 0.001$, for the size III and V stimuli, respectively). Data in Figures 2 and 3 were then fit by linear regression to determine whether the mean sensitivity difference between sessions 2 and 3 varied as a function of the mean sensitivity. Slopes of the best-fit regression lines were not significantly different from zero ($t = -0.17, P = 0.87$, and $t = -0.54, P = 0.59$, for the size III and size V stimuli, respectively). The nonsignificant slopes indicated that the mean sensitivity difference did not vary as a function of the mean sensitivity. Thus, straight lines with a slope of zero and y-intercepts of 0.26 and 0.45 were fit to the size III and V data sets, respectively, represented by the dashed horizontal lines in Figures 1 and 2.

The variability of the sensitivity differences was then examined using the approach of Bland and Altman.11,12 In this analysis, the value of each data point in Figures 1 and 2 was subtracted from the overall mean sensitivity value for the appropriate stimulus size. The result of this computation is equivalent to calculating the residuals from the dashed lines in Figures 1 and 2. The absolute value of these residuals was then fit by linear regression. In accordance with Bland and Altman,12 the 95% limits of repeatability were given by

$$D \pm 1.96 \sqrt{\frac{\pi/2}{n} \times R},$$

where $D = 0.26$ and $0.45$ for the size III and V stimuli, respectively, and $R = b_0 + b_1A_{ij}$. The constants $b_0$ and $b_1$ are


The slopes of the repeatability limits were significantly different from zero (\(P<0.00001\)), which indicates that the size effect increased as mean sensitivity decreased. The absolute values of the residuals from this regression line were then fit by linear regression. The 95% limits of agreement obtained by this analysis are indicated by the solid lines in Figure 5. These linear fits captured the pattern of data, and more complex fits were not needed.

Also indicated in Figure 5 are two forms of spatial summation, as described by Swanson et al.\(^{10}\) For linear spatial summation, the predicted change in sensitivity with an increase in stimulus size is equal to the change in stimulus area (1.2 log units, or 12 dB). For probability summation, the predicted sensitivity change is equal to one fourth the change in stimulus area (0.3 log units, or 3 dB). Most patients' results fell within these two limits. Possible explanations for the trend toward linear summation with decreasing mean sensitivity are considered in the Discussion.

**DISCUSSION**

Patients with RP in this study showed a slight improvement in overall visual field sensitivity from session 2 to session 3, likely because of a practice effect. More importantly, the test-retest repeatability was nonmonotonically related to the level of sensitivity loss. That is, the intersession variability was greatest at sensitivity levels of approximately 10 dB, similar to the pattern of results reported previously in patients with glaucoma.\(^7,8\) The overall pattern of intersession variability would continue to increase at low sensitivity levels. By this analysis, the intersession variability was greatest at sensitivity levels of approximately 10 dB, similar to the pattern of results reported previously in patients with glaucoma.\(^7,8\)

The general pattern of results obtained with the size V stimulus and pooled across test locations and patients, with conventions as in Figure 1.

As can be seen in a comparison of Figures 1 and 2, the 95% limits of repeatability reasonably capture the pattern of data shown in Figures 1 and 2. However, these linear limits of repeatability predicted that the intersession sensitivity differences are expected to lie. The slopes of the repeatability limits were +0.19 and +0.18 for the size III and V stimuli, respectively, which were significantly different from zero (\(t = 4.36, P < 0.00001\), and \(t = 4.41, P < 0.00001\), for the size III and size V stimuli, respectively). Statistically significant linear slopes indicated that the variability of the sensitivity differences increased as mean sensitivity decreased.

For mean sensitivity values that were greater than approximately 5 dB, the 95% limits of repeatability reasonably captured the pattern of data shown in Figures 1 and 2. However, these linear limits of repeatability predicted that the intersession variability would continue to increase at low sensitivity values, a trend not present in the data. To better model the variability of the sensitivity differences, the absolute values of the residuals were fit with a fourth-order polynomial. The 95% limits of repeatability obtained from these fits are indicated by the black curved lines in Figures 1 and 2. For both stimulus sizes, this model better described the decrease in variability at low mean sensitivity levels. By this analysis, the intersession variability was greatest at a mean sensitivity value of approximately 10 dB, and the variability decreased markedly below sensitivity values of approximately 5 dB.

As can be seen in a comparison of Figures 1 and 2, the overall pattern of results obtained with the two stimulus sizes did not differ substantially. For example, the slopes of the linear 95% limits of agreement did not differ significantly for the two stimulus sizes (\(t = 0.11, P = 0.91\)). However, there was a significant difference in the overall mean sensitivities for the two stimulus sizes, pooled across all patients and test locations (19.84 dB and 23.47 dB for the size III and V stimulus, respectively; \(t = 2.20, P < 0.05\)). Thus, overall sensitivity for the larger stimulus was greater, as expected.

To determine whether the effect of stimulus size varied with mean sensitivity, Bland-Altman analysis was performed. First, the sensitivity difference for the two stimulus sizes was plotted against the mean sensitivity at each test location for each patient; results are shown in Figure 3. Next, these data were fit by linear regression, as indicated by the dashed line in Figure 3. The slope of this regression line (\(-0.23\)) was significantly different from zero (\(t = 4.92, P < 0.00001\)), which indicates that the size effect increased as mean sensitivity decreased. The absolute values of the residuals from this regression line were then fit by linear regression. The 95% limits of agreement obtained by this analysis are indicated by the solid lines in Figure 5. These linear fits captured the pattern of data, and more complex fits were not needed.

**FIGURE 2.** Sensitivity difference between sessions 2 and 3 compared with mean sensitivity for the two sessions, obtained with the size V stimulus and pooled across test locations and patients, with conventions as in Figure 1.

**FIGURE 3.** Sensitivity difference between the size III and V stimuli versus mean sensitivity for the two stimulus sizes pooled across test locations and patients. **Thick dashed line**: best-fit linear regression line fit to the entire data set. **Thick solid lines**: 95% limits of agreement. **Thin horizontal dashed line**: zero sensitivity difference between the two stimuli. **Horizontal dot-dashed lines**: predicted size effect for the two forms of spatial summation.
variability in patients with RP was similar for stimulus sizes III and V.

A possible explanation for the nonmonotonic pattern of test–retest variability has been presented by Wyatt et al., who attributed their results in patients with glaucoma to an interaction between fixational eye movements and the visual field gradient (the change in sensitivity as a function of test location). For visual field regions with intermediate sensitivity, the visual field gradient tends to be steep so that fixational eye movements can result in relatively large changes in sensitivity between sessions. For visual field regions with the highest and lowest sensitivity, however, the visual field gradient tends to be shallow, and, as a result, small changes in fixation have little effect on sensitivity. However, Henson et al. concluded that fixation errors are not the major explanation for the increased variability at test locations with reduced sensitivity in static perimetry. The interstimulus spacing of the 30–2 program used in our study was insufficient to allow us to evaluate the eye-movement/gradient hypothesis in our patients with RP. It may also be that neural units in regions of intermediate sensitivity are “noisier” because of the juxtaposition of relatively healthy and badly damaged regions of retina, leading to increased variability. In our study, we sought to minimize a “floor effect” by excluding data from those visual field locations at which one or both measurements were zero. Nevertheless, the reduced variability we observed at the lowest sensitivity levels might have been related to a range restriction. That is, at low mean sensitivities, there is only a small range of possible sensitivity differences before one of the measurements exceeds the limits of the HFA and therefore is discarded, thus limiting the allowable intersession variability.

Our results for light-adapted static perimetry are in contrast to those of Roman et al. who performed static perimetry under dark-adapted conditions using a modified HFA in various patients with retinal degeneration, including RP. With a full-field achromatic stimulus, Roman et al. found a constant degree of intersession variability across sensitivity levels in a Bland-Altman plot. However, with a full-field stimulus, any eye movements that might have occurred between measurements would have had little effect on sensitivity, minimizing the intervisit variability. Roman et al. also examined test–retest variability using a focal chromatic stimulus under dark-adapted conditions and found no change in test–retest variability as a function of sensitivity level. However, they presented data from only two patients, covering a relatively narrow range of visual field sensitivities. Our results, shown in Figures 1 and 2, demonstrate that nonuniformities in the pattern of test–retest repeatability only emerge when data from a large number of patients covering a broad range of visual field sensitivities are examined.

Birch et al. measured dark-adapted thresholds longitudinally in a group of patients with RP and reported that the intervisit variability was considerably higher for perimetric thresholds than for visual acuity or the electroretinogram. Their criterion for change in visual sensitivity of 5 to 6 dB is similar to that for light-adapted perimetry observed in our study with RP. The interstimulus spacing of the 30–2 program used in our study was insufficient to allow us to evaluate the eye-movement/gradient hypothesis in our patients with RP. It may also be that neural units in regions of intermediate sensitivity are “noisier” because of the juxtaposition of relatively healthy and badly damaged regions of retina, leading to increased variability.

In our study, we sought to minimize a “floor effect” by excluding data from those visual field locations at which one or both measurements were zero. Nevertheless, the reduced variability we observed at the lowest sensitivity levels might have been related to a range restriction. That is, at low mean sensitivities, there is only a small range of possible sensitivity differences before one of the measurements exceeds the limits of the HFA and therefore is discarded, thus limiting the allowable intersession variability.

Our results for light-adapted static perimetry are in contrast to those of Roman et al. who performed static perimetry under dark-adapted conditions using a modified HFA in various patients with retinal degeneration, including RP. With a full-field achromatic stimulus, Roman et al. found a constant degree of intersession variability across sensitivity levels in a Bland-Altman plot. However, with a full-field stimulus, any eye movements that might have occurred between measurements would have had little effect on sensitivity, minimizing the intervisit variability. Roman et al. also examined test–retest variability using a focal chromatic stimulus under dark-adapted conditions and found no change in test–retest variability as a function of sensitivity level. However, they presented data from only two patients, covering a relatively narrow range of visual field sensitivities. Our results, shown in Figures 1 and 2, demonstrate that nonuniformities in the pattern of test–retest repeatability only emerge when data from a large number of patients covering a broad range of visual field sensitivities are examined.

Birch et al. measured dark-adapted thresholds longitudinally in a group of patients with RP and reported that the intervisit variability was considerably higher for perimetric thresholds than for visual acuity or the electroretinogram. Their criterion for change in visual sensitivity of 5 to 6 dB is similar to that for light-adapted perimetry observed in our study with RP. The interstimulus spacing of the 30–2 program used in our study was insufficient to allow us to evaluate the eye-movement/gradient hypothesis in our patients with RP. It may also be that neural units in regions of intermediate sensitivity are “noisier” because of the juxtaposition of relatively healthy and badly damaged regions of retina, leading to increased variability.

In our study, we sought to minimize a “floor effect” by excluding data from those visual field locations at which one or both measurements were zero. Nevertheless, the reduced variability we observed at the lowest sensitivity levels might have been related to a range restriction. That is, at low mean sensitivities, there is only a small range of possible sensitivity differences before one of the measurements exceeds the limits of the HFA and therefore is discarded, thus limiting the allowable intersession variability.

Our results for light-adapted static perimetry are in contrast to those of Roman et al. who performed static perimetry under dark-adapted conditions using a modified HFA in various patients with retinal degeneration, including RP. With a full-field achromatic stimulus, Roman et al. found a constant degree of intersession variability across sensitivity levels in a Bland-Altman plot. However, with a full-field stimulus, any eye movements that might have occurred between measurements would have had little effect on sensitivity, minimizing the intervisit variability. Roman et al. also examined test–retest variability using a focal chromatic stimulus under dark-adapted conditions and found no change in test–retest variability as a function of sensitivity level. However, they presented data from only two patients, covering a relatively narrow range of visual field sensitivities. Our results, shown in Figures 1 and 2, demonstrate that nonuniformities in the pattern of test–retest repeatability only emerge when data from a large number of patients covering a broad range of visual field sensitivities are examined.

Birch et al. measured dark-adapted thresholds longitudinally in a group of patients with RP and reported that the intervisit variability was considerably higher for perimetric thresholds than for visual acuity or the electroretinogram. Their criterion for change in visual sensitivity of 5 to 6 dB is similar to that for light-adapted perimetry observed in our study with RP. The interstimulus spacing of the 30–2 program used in our study was insufficient to allow us to evaluate the eye-movement/gradient hypothesis in our patients with RP. It may also be that neural units in regions of intermediate sensitivity are “noisier” because of the juxtaposition of relatively healthy and badly damaged regions of retina, leading to increased variability.

In our study, we sought to minimize a “floor effect” by excluding data from those visual field locations at which one or both measurements were zero. Nevertheless, the reduced variability we observed at the lowest sensitivity levels might have been related to a range restriction. That is, at low mean sensitivities, there is only a small range of possible sensitivity differences before one of the measurements exceeds the limits of the HFA and therefore is discarded, thus limiting the allowable intersession variability.

Our results for light-adapted static perimetry are in contrast to those of Roman et al. who performed static perimetry under dark-adapted conditions using a modified HFA in various patients with retinal degeneration, including RP. With a full-field achromatic stimulus, Roman et al. found a constant degree of intersession variability across sensitivity levels in a Bland-Altman plot. However, with a full-field stimulus, any eye movements that might have occurred between measurements would have had little effect on sensitivity, minimizing the intervisit variability. Roman et al. also examined test–retest variability using a focal chromatic stimulus under dark-adapted conditions and found no change in test–retest variability as a function of sensitivity level. However, they presented data from only two patients, covering a relatively narrow range of visual field sensitivities. Our results, shown in Figures 1 and 2, demonstrate that nonuniformities in the pattern of test–retest repeatability only emerge when data from a large number of patients covering a broad range of visual field sensitivities are examined.

Birch et al. measured dark-adapted thresholds longitudinally in a group of patients with RP and reported that the intervisit variability was considerably higher for perimetric thresholds than for visual acuity or the electroretinogram. Their criterion for change in visual sensitivity of 5 to 6 dB is similar to that for light-adapted perimetry observed in our study with RP. The interstimulus spacing of the 30–2 program used in our study was insufficient to allow us to evaluate the eye-movement/gradient hypothesis in our patients with RP. It may also be that neural units in regions of intermediate sensitivity are “noisier” because of the juxtaposition of relatively healthy and badly damaged regions of retina, leading to increased variability.

In our study, we sought to minimize a “floor effect” by excluding data from those visual field locations at which one or both measurements were zero. Nevertheless, the reduced variability we observed at the lowest sensitivity levels might have been related to a range restriction. That is, at low mean sensitivities, there is only a small range of possible sensitivity differences before one of the measurements exceeds the limits of the HFA and therefore is discarded, thus limiting the allowable intersession variability.

Our results for light-adapted static perimetry are in contrast to those of Roman et al. who performed static perimetry under dark-adapted conditions using a modified HFA in various patients with retinal degeneration, including RP. With a full-field achromatic stimulus, Roman et al. found a constant degree of intersession variability across sensitivity levels in a Bland-Altman plot. However, with a full-field stimulus, any eye movements that might have occurred between measurements would have had little effect on sensitivity, minimizing the intervisit variability. Roman et al. also examined test–retest variability using a focal chromatic stimulus under dark-adapted conditions and found no change in test–retest variability as a function of sensitivity level. However, they presented data from only two patients, covering a relatively narrow range of visual field sensitivities. Our results, shown in Figures 1 and 2, demonstrate that nonuniformities in the pattern of test–retest repeatability only emerge when data from a large number of patients covering a broad range of visual field sensitivities are examined.

Birch et al. measured dark-adapted thresholds longitudinally in a group of patients with RP and reported that the intervisit variability was considerably higher for perimetric thresholds than for visual acuity or the electroretinogram. Their criterion for change in visual sensitivity of 5 to 6 dB is similar to that for light-adapted perimetry observed in our study with RP. The interstimulus spacing of the 30–2 program used in our study was insufficient to allow us to evaluate the eye-movement/gradient hypothesis in our patients with RP. It may also be that neural units in regions of intermediate sensitivity are “noisier” because of the juxtaposition of relatively healthy and badly damaged regions of retina, leading to increased variability.


