Test-Retest, Within-Visit Variability of Goldmann Visual Fields in Retinitis Pigmentosa

Ava K. Bittner, Mian Haris Iftikhar, and Gislin Dagnelie

PURPOSE. Reliable outcome measures are needed to estimate changes in peripheral vision during future treatment clinical trials for retinal degeneration patients. The authors examined the short-term variability of Goldmann visual field (GVF) results converted to retinal areas in retinitis pigmentosa (RP) subjects.

METHODS. Two within-visit GVF s were obtained from one eye each of 37 RP subjects with visual acuity better than 20/400 by a single experienced operator using the V4e (n = 28) or III4e (n = 12) target, or both. Planimetric GVF measures were digitized and converted to retinal areas in square millimeters by a single independent user. The 95% coefficient of repeatability (CR.95) for percentage change in central retinal area was determined from the test-retest difference.

RESULTS. There were no significant systematic trends toward either increase or decrease between the first and second GVF. For the III4e target, the CR.95 was 23.7% on average across all 12 subjects. For the V4e target, the CR.95 was 32.8% on average across all 28 subjects. However, 3 of 8 subjects with a geometric mean retinal area <10 mm² (~7° radius) for the V4e target exhibited unusually large variability (50%–100%), and the CR.95 was 19.2% when these three subjects were excluded. Variability was not statistically significantly related to visual acuity, age, presence of cystoid macular edema, or subjects’ stress or anxiety levels.

CONCLUSIONS. Inherent test-retest variability (CR.95) of functional retinal areas derived from GVF results in a clinical RP population can be limited to <20% by using a single experienced operator, making the GVF the measure of choice for changes in peripheral vision. (Invest Ophthalmol Vis Sci. 2011; 52:8042–8046) DOI:10.1167/iovs.11-8321

Reliable outcome measures documenting visual function changes in mid- and far-peripheral areas are vital for clinical trials examining the safety and efficacy of treatments for retinal degenerative disease patients with peripheral visual field (VF) loss. Many treatments may preferentially affect VF area rather than, or in addition to, local VF sensitivity, and these changes may occur in peripheral locations not probed by most automated peripherals. Examples of such treatments already in early-phase clinical trials are transcorneal electrical stimulation for retinitis pigmentosa (RP), gene therapy for Leber’s congenital amaurosis (LCA), and oral synthetic retinoid therapy for some forms of LCA or RP (Koenekoop RK, et al. IOVS. 2011;52:ARVO E-Abstract 3323). Therefore, it is valuable to establish the typical short-term variability of VF measures obtained with an instrument likely to be used in such trials for its capacity to assess visual areas located beyond 60° eccentricity, such as the Goldmann VF (GVF).

There are only a few previous publications of VF variability in RP. A search of PubMed and similar sources revealed only two publications specifically examining GVF variability and two other publications on the variability of Humphrey VF sensitivity measures in RP. Ross et al. found that the variability of GVF area between sessions determined by planimetry using the V4e or III4e target was 12% on average in normally sighted observers and up to 50% in a group of RP subjects with visual acuity (VA) of 20/100 or better. The authors did not comment on the magnitude of the variability of GVF area in relation to subjects’ mean GVF extent. Examination of their published data indicates that as the GVF area decreased, some subjects’ eyes demonstrated increased variability between sessions, whereas others maintained low variability. Berson et al. determined intervisit GVF variability within 2 months in RP subjects with VA of 20/200 or better and found that a 22% decrease or a 29% increase in GVF diameter with the V4e test target was significant at the P < 0.01 level. We are not aware of any other publications in more than 25 years that have attempted to confirm these reports of GVF repeatability in other populations with RP or involving other trained GVF operators. Intersession Humphrey 30-2 VF variability in RP has been reported to range up to approximately 5 to 10 dB for each test location with the size V and III targets, which would be equivalent to test-retest differences ranging up to 16% in patients with good mean sensitivity (~32 dB) or up to 100% in patients with poor mean sensitivity (~10 dB). Thus, the limitation of the Humphrey VF is that it also appears to demonstrate significant variability across RP subjects, with a potential for particularly high variability in areas in which sensitivity is reduced; moreover, unlike the GVF, it is unable to assess areas of peripheral vision.

As one maps peripheral visual field locations, increasing planimetric distortion occurs relative to the retinal areas involved. This is due to two sources: the cartographic distortion between the chart and the perimeter bowl (i.e., from the flat paper to the hemispheric bowl of the GVF instrument) and the geometric optical projection of the perimeter bowl onto the retina (i.e., from the hemispheric bowl of the GVF instrument to the concave retinal surface). This has led to the recommendation to convert the GVF chart results to retinal rather than planimetric areas when monitoring changes in the peripheral VF in patients with retinal degeneration because natural progression and possible treatment effects are acting on the retina rather than the chart. Otherwise, the importance of changes occurring in peripheral VF areas will be overestimated relative to central areas.

As part of an investigation of possible causes of vision fluctuations reported by RP patients, we conducted a study of...
the effects of laboratory-induced stress on vision using a highly standardized protocol, the Trier Social Stress Test, by obtaining GVF results in RP subjects before and after the stress-inducing procedure involving mental arithmetic and public speaking. However, we found no statistically significant relationships between GVF changes and the magnitude of either the physiological stress level measured or the anxiety level reported by the subjects. Because the effects of stress played no demonstrable role in GVF variability in this subject group, the GVF changes between test and retest can be used as an estimate of GVF variability in a representative population of RP patients.

Therefore, our objectives in study were to explore the magnitude of within-visit GVF fluctuations measured by a single experienced operator in a group of RP patients with a relatively wide range of VA and VF loss and to determine whether GVF variability was correlated with subjects’ age, sex, level of vision, or cystoid macular edema (CME) status.

SUBJECTS AND METHODS

The protocol for the study was approved by the Institutional Review Board of the Johns Hopkins University School of Medicine and followed the tenets of the Declaration of Helsinki. Informed consent was obtained from the subjects after explanation of the nature and possible consequences of the study.

Subjects

Study subjects included 37 patients diagnosed with RP. Most of the subjects (n = 27; 73%) were recruited through the clinical practices at the Johns Hopkins Wilmer Eye Institute, from low vision optometrists or retinal specialists at our center, and these subjects were all experienced in the GVF test. The remaining 10 subjects received diagnoses of RP from retinal specialists outside our center and self-referred after learning of the study through online listings. Persons older than age 18 were eligible for the study if they had a confirmed diagnosis of RP from retinal specialists outside our center and self-referred after experiencing the GVF test. The remaining 10 subjects received diagnoses of RP from retinal specialists outside our center and self-referred after learning of the study through online listings. Persons older than age 18 were eligible for the study if they had a confirmed diagnosis of RP, with any level of vision, provided they could read reverse-contrast, large-sized font on a personal computer (i.e., VA better than 20/400). The mean age of our 37 subjects was 46 years (range, 19–76 years), and 21 were women (57%).

Data Collection

Data collection occurred from December 2008 through April 2010. The subjects enrolled at the Johns Hopkins Wilmer Eye Institute’s Lions Vision Center, and all vision tests were administered by a single examiner (AKB) during a single session. Best-corrected VA was measured in each eye with the Early Treatment of Diabetic Retinopathy Study (ETDRS, Lighthouse International, New York, NY) charts at 3 m or closer if fewer than 10 letters were identified. Best-corrected Pelli-Robson contrast sensitivity (CS) (Metropia, Ltd., Essex, UK) was also assessed for each eye at 1 m.

GVF was measured twice on the same day, within 1 to 2 hours, in one eye using the V4e and/or III4e test targets. We obtained results for both targets in three subjects, for the V4e target only in 25 subjects, and for the III4e target only in nine subjects. Typically, the V4e target was used in the repeated measurement for subjects with GVF diameters <60° for the V4e target and the III4e target was usually repeated for subjects with GVF diameters ≥60° for the V4e target. The eye that was used for the repeated GVF measurement was randomly selected such that the findings from 19 right eyes and 18 left eyes are reported here.

The GVF areas were plotted by tracing along each of 24 meridians one at a time, moving at a rate of approximately 5°/s, from nonseeing (peripheral) to seeing (central) areas. All meridians were evaluated, and a few of the meridians (~20%–25%) were rechecked for consistency in responses. Irregular seeing areas and scotomas were further explored in tangential directions as needed to obtain a reliable contour. Areas that were noted as not consistently seen were not included in our calculations of the GVF area. The tested points were connected with straight lines to form the areas or isopters.

Spectral domain–optical coherence tomography (SD-OCT) measurements were obtained at the time of study enrollment from all subjects using the HRA+OCT (Spectralis; Heidelberg Engineering, Vista, CA) to detect the presence of CME. The acquisition protocol consisted of a series of horizontal raster scans covering 20° centered on the fovea. The raster scans were spaced 240 μm apart, with each raster scan consisting of a series of A-scans with a transverse resolution of 14 μm and in-depth resolution of 3.9 μm. The resolution of each line was enhanced by repeating and averaging the measurements on each line at least 15 times using image alignment software (TruTrack; Heidelberg Engineering).

Statistical Analysis

A single operator at our center (MHI) used dedicated computer software to digitize the GVF maps and to calculate the planimetric and retinal areas for each subject. The central GVF area was defined as a region encompassing central fixation that was either completely intact or included one or more missing scotomas near the center or in the midperiphery, which were subtracted to determine the net seeing central area. The total GVF area included the central area as well as any isolated peripheral islands, which were measurable for eight subjects with the V4e target. We present VF areas on a logarithmic scale because it has been documented in several longitudinal, natural history studies that the remaining viable retinal area declines over time according to a negative exponential function.

In this study, we established the percentage change between test and retest for the central and total GVF areas in each subject as well as means for the III4e and V4e groups. We defined variability in GVF as the 95% coefficient of repeatability (CR-95) calculated as 1.96 times the across-subjects mean of the absolute percentage test-retest difference. This CR-95 establishes a test-retest interval beyond which a GVF change found in RP patients similar to that in our population is indicative of a true effect.

We used simple linear regressions to compare the percentage change in retinal areas with subjects’ VA, CS, mean log retinal area, and age, as well as Welch’s two-sample t-tests with unequal variances to test for significant differences in the percentage change in retinal area for each test target according to dichotomous characteristics (i.e., CME, sex; Stata/IC version 10.0; Stata Corp., College Station, TX).

RESULTS

Figure 1 displays the GVF test-retest difference in percentage for each subject as a function of their mean central retinal or planimetric (chart) area determined with the GVF V4e and III4e targets in a Bland-Altman plot. The top panel shows test-retest values, whereas the bottom panel shows absolute percentage differences. Solid symbols represent retinal areas, based on a model eye transform of 0.272 mm², whereas open symbols represent planimetric areas, using the same conversion factor, so that planimetric and retinal areas were the same for small central fields. Note that the discrepancy between retinal and planimetric areas was minimal for areas up to 100 mm² and became increasingly important for larger fields. In the top panel, the mean GVF test-retest differences across subjects were ~1.0% and ~0.2% for the V4e and III4e targets, respectively, for retinal areas and ~1.7% and 0.2%, respectively, for planimetric areas; there was no significant trend toward either an increase or decrease in areas when comparing the GVF test and retest results. These results confirm that there were no apparent significant test-retest effects related to learning/practice, fatigue, or stress/anxiety across subjects.

The data in the top panel of Figure 1 suggest that, especially for small visual fields with areas below 10 mm² (i.e., visual fields smaller than 14° in diameter), some test-retest differ-
ences were much larger than the typical range of \(-25\%\) to \(25\%\). Therefore, in the bottom panel of Figure 1, three outliers with mean areas below 10 mm\(^2\) for the V4e target and differences exceeding 50\% have been eliminated. In addition, all negative differences have been replotted as absolute positive values to permit examination of a possible general relationship between test-retest differences and mean retinal area. Simple linear regression by target and for each type (V4e-retina: \(r^2 = 0.05, P = 0.27\); V4e-chart: \(r^2 = 0.04, P = 0.34\); III4e-retina: \(r^2 = 0.12, P = 0.28\); III4e-chart: \(r^2 = 0.06, P = 0.43\)) shows that the remaining functional area hardly contributes to the test-retest difference.

Table 1 shows the mean and range of the absolute test-retest differences for both target sizes, before and after elimination of subjects with changes in excess of 45\%, and for subgroups according to mean area greater than or less than 10 mm\(^2\) with the V4e target. For the V4e target, two additional subjects showed large test-retest changes \(>45\%\) when peripheral islands were also included with central areas for the total area; hence five rather than three subjects were eliminated for the reduced set. For the III4e target, only a few subjects \((n = 3)\) had measurable peripheral areas, and including these areas did not affect the results.

The third and fifth columns in the table show the CR\(_{95}\), computed as 1.96 times the absolute mean test-retest difference (in percentages), and 95\% of repeated measures are expected to have a test-retest difference that does not exceed these values. Therefore, these values represent the percentage change in GVF area to be exceeded in a clinical study for a change to be considered significant. Note that the CR\(_{95}\) can be conservatively set at 20\% for both target sizes, central and total areas, and retinal and planimetric areas, after elimination of a small number of subjects with test-retest differences near or above 50\% or when only considering subjects with mean areas \(>10\) mm\(^2\).

Although retinal and planimetric areas have been used by most authors examining changes in GVFs over time, either as the natural progression or in response to treatment, some authors have used visual field diameter as their measure of choice, under the assumption that this measure should be highly correlated with the square root of the planimetric area. Figure 2 plots the GVF horizontal diameter against the square root of the mean retinal and planimetric (chart) areas for both the III4e and V4e targets. The diagonal line represents the anticipated relationship if the visual field contour is perfectly circularly symmetric visual field.

Table 1. Percentage Changes in Retinal and Planimetric Areas According to GVF Isopters and Subject Groups

<table>
<thead>
<tr>
<th>GVF Isopters/Subjects</th>
<th>Retinal Area Mean (Range)</th>
<th>CR(_{95})</th>
<th>Planimetric Area Mean (Range)</th>
<th>CR(_{95})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central III4e/all subjects ((n = 12))</td>
<td>12.1 (0.2–45.2)</td>
<td>23.7</td>
<td>12.7 (3.0–45.2)</td>
<td>24.9</td>
</tr>
<tr>
<td>Central V4e/all subjects ((n = 28))</td>
<td>16.7 (0.3–96.7)</td>
<td>32.8</td>
<td>16.5 (0.03–98.9)</td>
<td>32.4</td>
</tr>
<tr>
<td>Total V4e/all subjects ((n = 28))</td>
<td>18.5 (0.2–96.7)</td>
<td>36.3</td>
<td>22.0 (0.03–138.8)</td>
<td>43.2</td>
</tr>
<tr>
<td>Central III4e/reduced set ((n = 11))</td>
<td>9.1 (0.2–19.7)</td>
<td>17.8</td>
<td>9.7 (3.0–17.6)</td>
<td>19.1</td>
</tr>
<tr>
<td>Central V4e/reduced set ((n = 25))</td>
<td>9.8 (0.3–26.7)</td>
<td>19.2</td>
<td>9.7 (0.03–26.3)</td>
<td>19.0</td>
</tr>
<tr>
<td>Total V4e/reduced set ((n = 25))</td>
<td>9.0 (0.2–22.8)</td>
<td>17.7</td>
<td>9.4 (0.03–26.3)</td>
<td>18.3</td>
</tr>
<tr>
<td>Central V4e/area &gt;10 mm(^2) ((n = 20))</td>
<td>9.5 (0.3–26.7)</td>
<td>18.7</td>
<td>9.6 (0.03–26.3)</td>
<td>18.8</td>
</tr>
<tr>
<td>Central V4e/area &lt;10 mm(^2) ((n = 8))</td>
<td>34.7 (0.5–96.7)</td>
<td>67.9</td>
<td>33.9 (0.4–98.9)</td>
<td>66.4</td>
</tr>
</tbody>
</table>
centered and circular, without any included scotoma. The figure demonstrates that for real visual fields, even with small diameters and areas, the data points fall below the line, in part because of the irregularity of the visual field contour and the existence of included scotomata but, more important, because of the tendency of visual fields to have a wider horizontal than vertical diameter. The figure also confirms that a disparity between retinal and planimetric areas becomes increasingly important for subjects with GVF diameters >60° (i.e., radius >30°). As expected, the planimetric area calculations exceed the retinal areas contributing to visual function for a third of our RP subjects (n = 12) who had a GVF diameter >60°.

It has often been reported that there is a correlation between loss of visual field area and central measures such as VA and CS. As part of our analysis of test-retest differences, we examined whether VA or CS loss in advanced RP might be associated with increased test-retest differences. Figure 3 shows the relationship between the percentage change in central retinal area and VA (top panel) or CS (bottom panel) across subjects. In the VA plot, the subjects with large variability for the V4e target have near-average VA levels, and regression either with ($r^2 = 0.03, P = 0.26$) or without ($r^2 = 0.007, P = 0.61$) these subjects shows no statistically significant dependence of test-retest difference on VA. For the CS plot, a tendency is seen for the largest test-retest differences to occur in subjects with relatively poor CS. Simple linear regression analysis including all subjects confirmed that greater percentage changes in central retinal area were statistically significantly related to reduced CS ($r^2 = 0.19, P = 0.005$). Removing the subjects with test-retest differences near or greater than 50% dramatically reduced the relationship ($r^2 = 0.07, P = 0.11$). There were no statistically significant relationships between the percentage change in log central retinal area and age for subjects tested with the V4e ($P = 0.67$) and III4e ($P = 0.29$) targets.

Nearly one-fourth of the RP subjects (n = 9; 24.3%) had CME detectable with OCT in the study eye. Using t-tests, we found that, on average, subjects with CME did not exhibit statistically significantly larger amounts of variability expressed as the percentage change in log retinal area when compared to those without CME for both the V4e ($P = 0.07$) and III4e ($P = 0.77$) targets. There was actually a trend for subjects tested with the V4e target who had CME to exhibit less variability in their GVF measures than those without CME. We also did not find any statistically significant differences in percentage change in log retinal area according to sex for subjects tested with either the V4e ($P = 0.27$) or the III4e ($P = 0.77$) targets.

**DISCUSSION**

In the present study we have shown that the typical variability of GVF results ranged up to ±20% for the test-retest differences across most RP patients tested with V4e and III4e targets. It is important to note that large amounts of test-retest variability >50% are possible but occurred in a small sample (3 of 8 subjects with GVF diameter <14° with the V4e target). We excluded these outliers from our estimate of the typical GVF variability because clinical trials may use a screening process to identify and exclude a minority of potential participants exhibiting high test-retest variability. The present study demonstrated that increased RP severity (assessed by mean GVF, VA, CS, age, CME) did not tend to significantly predict increased GVF variability across subjects, although the greatest test-retest differences were present in some but not all the subjects with relatively poor CS or reduced mean GVF area/diameter. We used two different methods to calculate GVF within-session variability based on planimetric and retinal areas. Our findings suggest that retinal areas should be used for patients with central GVF diameters >60° because planimetric areas and the resultant variability are greater (and, in fact, distorted) for subjects with preserved far peripheral visual fields.

The primary difference between the present study and the previous study of GVF variability by Ross et al.3 was the longer test-retest follow-up time in that study (mean, 20 days) because it examined variability between sessions and between operators rather than within session for a single operator. Within-visit variability is more likely to reflect the inherent fluctuations from patient- and examiner-related factors, whereas between-visit variability may also include true vision fluctuations or changes in addition to the inherent variability we measured within a 1- to 2-hour period during a single visit. To explore this, a future study could compare within- and between-visit GVF variability within subjects in the placebo group tested with a single operator during an upcoming clinical trial. The previous study by Ross et al.3 documented that the planimetric area test-retest difference was approximately 12% on average for both sessions and within operator for RP patients tested with the III4e target,3 which is the same as the within-session test-retest difference across all subjects tested with the III4e target in the present study. The Ross et al. study3 did not test a large proportion of patients with the V4e target (n = 5); therefore, its ability to compare variability between test target sizes was limited. In addition, the participants in that study tended to have larger GVF diameters on average than our participants,3 and the study included only one participant with
a GVF diameter <14°, limiting the ability to assess variability among persons with severely constricted GVFs.

The study by Berson et al.4 on intervisit GVF repeatability used 99% rather than 95% probability and found asymmetric and slightly larger (−21%, +29%) confidence limits for variability in GVF diameter compared with our symmetric (−19%, +19%) limits derived from log retinal areas. However, the Berson et al.4 use of diameter rather than area would, on average, be expected to yield a confidence interval half the size of ours, whereas (under assumption of a normal distribution) the 99% versus 95% confidence would increase the interval by a factor of 1.5; thus, the findings of the that study4 are, in fact, similar to ours. The asymmetry in their results matches the symmetry in ours if one bears in mind that the differences in our study, derived from log areas, are proportions, and that the proportion 100/79 is very similar to the proportion 129/100.

Several possibilities explain the high test-retest variability we encountered in a limited a number of subjects in whom the percentage change exceeded 45%. Our results indicate that this tends to occur in some patients with advanced RP with severely reduced GVF and CS or when far peripheral islands are included in the total area. The increased variability may be due in part to a limited number of diseased or sporadically located photoreceptors that are less able to consistently respond to the GVF test stimulus. However, patient-related factors, such as reduced attention during the test procedure because of sleepiness or distracting thoughts were not assessed during our study and therefore cannot be ruled out as another possible source of variability.

Although we did not find that laboratory-induced stress was related to GVF decrements across subjects, previous research in normally sighted subjects without RP suggested that real-life stressful situations produce greater perceptual peripheral VF constrictions than laboratory-induced stress procedures, possibly because of increased anxiety.10 In addition, normally sighted subjects with higher life event stress experienced much greater decrements in peripheral vision during a laboratory-induced stress condition (mean change, 14.3°) than those with lower life event stress scores (mean change, 2.2°).11 We administered the Life Experiences Survey to assess positive and negative major life changes within the past year (e.g., marriage, new job, death of a close friend/relative), and found that the three subjects with test-retest differences >50% for the V4e target had higher negative life event scores than the less variable subjects. However, we are limited in our ability to generate conclusions regarding this association because of the small proportion of subjects with high variability and high negative life event scores; therefore, we are recommending this as an area for future research in a larger sample.

The results of this study can be used to check for the proficiency of newly trained operators by comparing the variability of their results with those of our experienced operator. Caution must be used in generalizing our results to all GVF operators because the variability for untrained or newly trained operators is unknown and may be an appropriate topic to explore during future research. It is also preferable to have a single, well-trained, and experienced operator perform all GVF tests within a center, or at least within a clinical trial. At the very minimum, given that previous research has demonstrated that the within-session GVF test-retest difference is approximately 13.5% on average when comparing results between two operators, we recommend assigning a single well-trained operator to a particular subject for longitudinal monitoring. However, when this is not feasible or changes in staff are anticipated over long periods of time, a rigorous training program to certify operators in proper, consistent GVF technique is highly recommended. Most of our subjects had previous experience with the GVF, and we recommend a “practice” GVF session to familiarize patients who are new to the GVF procedure.

Our findings may be used to guide the design and conduct of future clinical trials in which measurement of peripheral VF is required to study the natural progression of a retinal disease or the possible effect of treatment. There is a paucity of GVF instruments, and most are located in academic centers. This is unfortunate given the potential value and role of the GVF in future clinical trials for patients with retinal degeneration and should be an important consideration for manufacturers of perimetric equipment. This study highlights the need to screen potential clinical trial subjects for large within-session GVF variability, especially when enrolling those who are legally blind because of constricted visual fields. An important finding of this study is that longitudinal GVF changes >20% may be considered outside the range of typical variability and therefore indicative of significant change as long as researchers have taken proper precautions to reduce subject-based variability by excluding prospective subjects who show large within-session GVF changes from participation in clinical trials and to ensure operator-based variability by using a single well-trained, experienced operator for within-subject repeated testing.

Acknowledgments

The authors thank Paul Dagnelie and Liancheng Yang for creating the software used to digitize the GVF charts and calculate the planimetric and retinal VF areas used in this study (the C language software may be obtained from the authors), and Mohamed Ibrahim for obtaining the OCT measures.

References