The Repeatability of Mean Defect with Size III and Size V Standard Automated Perimetry

Michael Wall,1,2 Carrie K. Doyle,1,2 K. D. Zamba,3 Paul Artes,4 and Chris A. Johnson1

PURPOSE. The mean defect (MD) of the visual field is a global statistical index used to monitor overall visual field change over time. Our goal was to investigate the relationship of MD and its variability for two clinically used strategies (Swedish Interactive Threshold Algorithm [SITA] standard size III and full threshold size V) in glaucoma patients and controls.

METHODS. We tested one eye, at random, for 46 glaucoma patients and 28 ocularly healthy subjects with Humphrey program 24-2 SITA standard for size III and full threshold for size V each five times over a 5-week period. The standard deviation of MD was regressed against the MD for the five repeated tests, and quantile regression was used to show the relationship of variability and MD. A Wilcoxon test was used to compare the standard deviations of the two testing methods following quantile regression.

RESULTS. Both types of regression analysis showed increasing variability with increasing visual field damage. Quantile regression showed modestly smaller MD confidence limits. There was a 15% decrease in SD with size V in glaucoma patients (P = 0.10) and a 12% decrease in ocularly healthy subjects (P = 0.08).

CONCLUSIONS. The repeatability of size V MD appears to be slightly better than size III SITA testing. When using MD to determine visual field progression, a change of 1.5 to 4 decibels (dB) is needed to be outside the normal 95% confidence limits, depending on the size of the stimulus and the amount of visual field damage. (Invest Ophthalmol Vis Sci. 2013;54:1345–1351) DOI:10.1167/iovs.12-10299

Various formulas have been used to estimate the global visual field sensitivity. Flammer1 introduced the term mean defect (MD), which is calculated by computing the average age-corrected loss per test location. Shortly thereafter, Heijl et al.2,3 suggested the term mean deviation for the average age-corrected loss per test location weighted for intra- and intersubject variability related to eccentricity. It is calculated by determining the mean of the total deviation values weighted for the variability of the normal values at that test location; test loci with lower variability (the more central locations) have greater weight in the calculation. These robust perimetric indices are useful measures of central tendency because they are less susceptible to the fluctuations known to be common for individual test locations. It appears that differences between the mean deviation and mean defect are minimal.4

While the MD is not sensitive to small and shallow defects, optic neuropathies including glaucoma, often have diffuse involvement of nerve fiber bundles resulting in generalized loss throughout the visual field.5 This combination of capturing diffuse loss and relatively lower variability gives the MD useful characteristics as a summary statistic for visual field change.

Tattersall and coworkers6 estimated the variability of the MD using a large cohort of stable glaucoma patients recorded over 3 years in which the first and last Advanced Glaucoma Intervention Study (AGIS) scores were the same. A variety of Humphrey Field Analyzer testing methods were pooled including Full Threshold, Fastpac, and Swedish Interactive Threshold Algorithm (SITA) Standard. Tattersall and coworkers collapsed their cohort by five stages of defect severity. They reported an increased width of the 99% confidence interval with increasing visual field damage with a range of 0.3 to 1.3 decibels (dB).

To better understand the variability of the MD with both Goldmann size III and size V testing and its change with increasing visual field damage, we obtained repeated measures from a cohort of normal and glaucoma patients tested once a week for 5 weeks.

SUBJECTS AND METHODS

Subjects

The visual testing protocol was approved by the University of Iowa Institutional Review Board. One eye from each of 46 glaucoma patients and 28 ocularly healthy participants was tested once a week for 5 weeks. They all gave written informed consent to participate in the study. The tenets of the Declaration of Helsinki were followed. The normal subjects were volunteers, paid in accordance with the Institutional Review Board guidelines, who answered advertisements inviting them to participate in research.

Normal participants were included if they had (1) no history of eye disease, (2) refractive error within ±5 diopter (D) sphere and ±2 D astigmatism, (3) no history of diabetes mellitus or systemic arterial hypertension, and (4) a normal ophthalmologic examination including 20/25 or better Snellen visual acuity. The subjects either had undergone a complete eye exam within 24 months prior to this study or were examined by an ophthalmologist on the day of testing to ensure normal ocular health. One eye of each participant was randomly chosen as the study eye.

The glaucoma patients were invited from the glaucoma clinic at the University of Iowa Department of Ophthalmology and Visual Sciences if they met entry criteria. They were enrolled if they had glaucomatous
optic disc changes with abnormal standard automated perimetry (SAP; glaucomatous visual field defects, i.e., three or more adjacent test locations that fell outside normal limits in a clinically suspicious area at the $P < 0.05$ level or two adjacent locations falling outside normal limits with at least one at the $P < 0.01$ level). In addition, mean deviation was in the range of 0 to $-30$ dB on SAP. We included patients with primary, secondary, or normal tension glaucoma. The patients did not have another disease affecting vision and were capable of performing SAP and returning for follow-up visits. Patients were excluded if they had cataract causing visual acuity of worse than 20/30, pupil size less than 2.5 mm, or age less than 19 years. If both eyes qualified for the study, one eye was chosen at random as the study eye.

The average age (and SD) of the normal subjects was 45.4 ± 16.1 years with a range of 24 to 77 years; their mean deviation was $-0.02$ ± 0.73 dB with Pattern Standard Deviation (PSD) of 1.34 ± 0.25 dB. Nineteen of the healthy participants were women and nine were men. Most of the glaucoma patients were experienced with automated perimetry; most of the normal subjects were naive perimetry subjects. Twenty-eight of the glaucoma patients were men and 18 were women. The average age of the glaucoma patients was 65.7 ± 14.0 years; range, 22 to 82 years; their mean deviation was $-9.80$ ± 7.15 dB with a PSD of 7.89 ± 4.40 dB.

**Visual Testing**

**Humphrey Field Analyzer.** All subjects underwent automated perimetry using program 24-2 of the Humphrey Field Analyzer (HFA; Carl Zeiss Meditec, Dublin, CA). For SAP, the stimuli of Goldmann size III (0.5 diameter, 4 mm$^2$) were used with the SITA standard 24-2 algorithm. Goldmann size V stimuli (1.72 diameter, 64 mm$^2$) were used along with the Full Threshold testing strategy; there is no SITA program currently available for size V stimuli. SITA was used for size III testing for practical considerations. These data were collected to determine confidence limits for a glaucoma change probability analysis for a longitudinal glaucoma study comparing SITA size III with full threshold size V. The subjects had Matrix and motion perimetry testing randomly intermixed, but the mean defect of these results was not analyzed. While using size III full threshold for comparison would be

---

**Figure 1.** The average mean deviation for the five tests is plotted against their standard deviation for the glaucoma patients. More fluctuation is present in (A) size III SITA standard testing than in (B) size V full threshold testing.

**Table 1.** A Comparison of the Standard Deviation of the Five Retests for Glaucoma Patients and Control Subjects

<table>
<thead>
<tr>
<th></th>
<th>Glaucoma, n = 46</th>
<th>Control, n = 28</th>
<th>Total, n = 74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean SD, III</td>
<td>1.01</td>
<td>0.54</td>
<td>0.69</td>
</tr>
<tr>
<td>Mean SD, V</td>
<td>0.83</td>
<td>0.45</td>
<td>0.60</td>
</tr>
<tr>
<td>Ratio, SD V/SD III</td>
<td>0.85</td>
<td>0.88</td>
<td>0.86</td>
</tr>
<tr>
<td>$P$-value*</td>
<td>0.10</td>
<td>0.08</td>
<td>0.06</td>
</tr>
</tbody>
</table>

SDs were log-transformed for the calculation of means, ratios, and $t$-tests.

* $P$ value refers to a paired $t$-test.

---

**Table 2.** Intraclass Correlations

<table>
<thead>
<tr>
<th>Group</th>
<th>Size III</th>
<th>Size V</th>
<th>No. of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0.55</td>
<td>0.56</td>
<td>28</td>
</tr>
<tr>
<td>Glaucoma</td>
<td>0.97</td>
<td>0.98</td>
<td>46</td>
</tr>
</tbody>
</table>
FIGURE 2. Results of quantile regression of size III (A) and size V (B) in glaucoma patients; the insert shows the results of 28 normal control subjects. Note the increasing width of the 95% repeatability interval with increasing damage.
Figure 3. (a-d) Mean defect over time for the two stimulus sizes and the two groups tested. The graphs show the averaged mean defect at each of the five test sittings separated by approximately 1 week. They all show a small learning effect of approximately 0.3 dB overall in the normal subjects and glaucoma patients with size III and approximately 0.5 dB overall in the size V glaucoma patients.
ideal, our pilot data and the work of Artes and coworkers show that the variability differences between SITA and full threshold are minor.

We followed the manufacturer's recommendations for using wire-rimmed corrective lenses (House of Vision Instrument Company, Lisle, IL). Care was taken to prevent lens rim artifacts. The subjects had testing in one eye, chosen at random, but the same eye was used for all tests. All visual field examinations met the following reliability criteria: fixation losses < 20% or normal gaze tracking, false-positive rate <

![Bland-Altman plots showing the difference of the mean defect (MD_{Iowa}) and the MD_{Statpac} in glaucoma subjects (left) and normal subjects (right) of the mean (A) and standard deviation (B) of the five trials.]}
10%, and false-negative rate < 33%. The four tests were administered in a random order with at least a 5-minute rest break between the two test procedures.

Statistical Analysis. To assess the relationship between visual field damage and variability of the MD, standard deviation of the MD was regressed against the mean MD for the five repeated tests. To compare the variability of MD between size III and size V, the ratio of the SDs (size V, size III) was calculated for each patient. The average reduction in variability was then derived as the geometric mean of these ratios. To further assess variability, intraclass correlations were calculated using R statistical software (provided in the public domain by The R Project for Statistical Computing, http://www.r-project.org).

We also compared our calculated mean defect with the Humphrey Field Analyzer Statpac mean deviation. The data we used to calculate mean defect for each subject in the present study is from a separate set of 60 normal subjects tested twice. The data were standardized to age 45 and the 50th percentile value was used to calculate the mean defect for the current study (after standardizing the raw data of this study’s subjects to age 45 and taking the mean value).

Differences between groups was considered significant if $P < 0.05$. Quantile regression of individual MD pairs was performed using R statistical software to establish the 90% limits of differences between test and retest.

RESULTS

The mean test time for normal subjects with SITA standard was 4.7 ± 0.51 minutes and for full threshold size V 9.4 ± 0.62 minutes. For glaucoma subjects, the mean test time was 6.7 ± 1.32 minutes with SITA standard III and 11.7 ± 1.88 minutes with size V full threshold. The glaucoma patients had a mean deviation (Statpac) of −9.80 ± 7.15 dB with a PSD of 7.89 ± 4.40 dB. The normal subjects had a mean deviation of −0.02 ± 0.73 dB with a PSD of 1.34 ± 0.25 dB. Our calculated mean defect with size III was −9.14 ± 6.62 dB with SITA size III testing in glaucoma patients and −7.25 ± 4.75 dB with size V. The average (geometric mean) of the standard deviations for the five repeat visual field results in glaucoma subjects was 1.01 dB with SITA size III testing and 0.85 dB with size V.

Figure 1 shows the well-known relationship of increasing variability with increasing visual field damage and compares size III and size V results. The $R^2$ of this relationship with size III was 0.20, and with full threshold size V $R^2$ was 0.29. There was a 15% decrease in SD with size V in glaucoma patients and a 12% decrease in ocularly healthy subjects (Table 1).

Quantile regression (Figs. 2, 3) show the 95% repeatability interval becomes larger with increasing sensitivity and size III goes from needing a change of 1.4 dB at 0 dB MD to 5.15 dB at 30 dB. The values for size V are 1.1 dB at 0 dB to 4.7 dB at 30 dB.

When comparing the results from the five test sittings over time a small learning effect is present (Fig. 3). The effect is approximately 0.3 dB overall in the normal subjects and glaucoma patients with size III and approximately 0.5 dB overall in the size V glaucoma patients. This analysis also suggest lower variability of size V full threshold testing when comparing the $R^2$ values: normal subjects with sizes III (0.22) and V (0.70) and glaucoma patients with sizes III (0.50) and V (0.71).

The Bland-Altman plots of our glaucoma subjects and normal subjects comparing the mean deviation (HFA calculation) with our calculated mean defect show good agreement (Fig. 4A); the standard deviations are compared in Figure 4B and again show good agreement. Table 2 shows the results of intraclass correlations. There were no significant differences between size III and size V in either glaucoma patients or healthy subjects.

DISCUSSION

Optic neuropathies are characterized by visual loss within the central 30° radius visual field. 5,8–9 Although definite localized defects are present in most optic neuropathies, progression or improvement is also commonly generalized or widespread. The MD is less sensitive to localized visual field change than pointwise or nerve fiber bundle zone indices, but it is a good candidate to estimate global, overall change. Because the effective dynamic range of size V is approximately one log unit greater than size III and its variability for individual test locations is substantially lower, 10 coupled with a similar ability to detect glaucomatous defects, 10,11 use of the MD for this stimulus size would appear to be a desirable measure. However, we found the variability of the full threshold size V MD to be only slightly lower than with size III testing. The discrepancy in variability between the substantially lower variability at individual test locations for size V and the modestly lower variability found with MD likely have contributions from three sources. First, fewer 0 dB test locations were present when testing with size V; in other words, 0 dB stimuli found with size III become measurable with the size V stimulus and therefore add to the variability. Second, since mean defect is an average, there was a dampening effect of the high variability of individual test locations. Since most of our subjects had most test locations above 20 dB, the low variability test locations may have dampened the effect of the higher variability locations. Third, a higher variability of full threshold testing existed for individual test locations between 5 and 20 dB.
When using MD to determine a significant change between two tests, a change of 1.4 to 5 dB is needed to be outside the 90% repeatability limits, depending on the size of the stimulus and the amount of visual field damage. This estimate is substantially greater than the results of Tattersall et al. For example, with an MD of –10 dB, the width of the 90% repeatability interval is 5.3 dB for size III (requiring a change of 2.6 dB for size III to fall outside the 95% limit). For size V, the width is 4.6 dB with a change of 2.3 dB needed. The corresponding figure from Tattersall et al. for the 99% confidence interval for size III is 1.6 dB with a change of 0.8 dB needed to identify positive or negative change. Tattersall and coworkers compacted their data into five bins and calculated the mean and confidence interval for each bin. It is possible that with the analyses of Tattersall, binning the data may have contributed to the differences.

We used quantile regression to describe the limits of test–retest differences in our data. Standard regression estimates how the mean of the dependent variable changes with the independent variable. In contrast, quantile regression estimates changes in arbitrary quantiles, for example the 90th percentile. Because we are interested in the limits of the distribution rather than the mean, quantile regression appears to be the appropriate technique for this data.

Artes et al. compared MD and the visual field index in 109 glaucoma patients with SAP size III data obtained from a longitudinal study. They plotted the difference between the level of MD at single tests and their “best available estimates.” The variability shown in their graph (Fig. 5) is similar in magnitude to ours. Funkhouser and Funkhouser compared the mean defect and the mean deviation in 169 visual field examinations in glaucomatous eyes with varying degrees of damage, and the differences were so small they concluded the two indices could be used interchangeably in glaucoma. They attribute this to the presence of generalized loss in glaucoma. We used Bland–Altman plots to compare the mean deviation with mean defect. Figure 3 shows good agreement of both the mean defect with the mean deviation and the standard deviations of the two methods in glaucoma patients and normal subjects.

We found a modest learning effect over the five testing periods (Fig. 3). The effect was approximately 0.3 dB overall in the normal subjects and glaucoma patients with size III and approximately 0.5 dB overall in the size V glaucoma patients. This effect was smaller than the learning effects usually found. We believe this was due to cross learning from having four visual field tests performed at each visit.

Our study would be strengthened by including more subjects, particularly for the more damaged MD values. Also, an equal distribution of subjects across MD values would add strength to the findings. While differences in variability between STIA standard and full threshold testing with size III are small, the study would have been stronger had full threshold testing of size III been done. There would be a modest increase variability of size III had full threshold testing been used. The differences between full threshold and STIA testing exist between 5 and 20 dB and peak at 10 dB, where there is approximately a 20% increase in variability with full threshold testing. This effect would increase the variability of size V full threshold relative to size III STIA standard testing.

The repeatability of size V MD appears slightly better than size III STIA testing with a reduction of 15% in glaucoma subjects and 12% in healthy control subjects. When using MD to determine visual field progression, a change of 1.5 to 4 dB is needed to be outside the normal 95% confidence limits, depending on the size of the stimulus and the amount of visual field damage. We caution about using one test alone outside the repeatability limits to determine visual field change. While variability clearly increases with increasing damage and limits for change vary accordingly, further work needs to be done to determine criteria for identifying visual field change.

References