Size Threshold Perimetry Performs as Well as Conventional Automated Perimetry With Stimulus Sizes III, V, and VI for Glaucomatous Loss

Michael Wall,1,2 Carrie K. Doyle,1 Trina Eden,1 K. D. Zamba,3 and Chris A. Johnson1

1Department of Ophthalmology, University of Iowa, College of Medicine, Veterans Administration Hospital, Iowa City, Iowa
2Department of Neurology, University of Iowa, College of Medicine, Veterans Administration Hospital, Iowa City, Iowa
3Department of Biostatistics, University of Iowa, College of Medicine, Veterans Administration Hospital, Iowa City, Iowa

Correspondence: Michael Wall, University of Iowa, College of Medicine, Department of Neurology, 200 Hawkins Drive #2007 RCP, Iowa City, IA 52242-1053; michael-wall@uiowa.edu.

Submitted: November 13, 2012
Accepted: April 23, 2013

PURPOSE. It is thought that large perimetric stimuli are insensitive for demonstrating visual field defects. To test the hypothesis that there is no difference in the total number of abnormal test locations with total deviation empiric probability plots in glaucoma patients, we compared results of glaucoma patients tested with sizes III (0.43° diameter), V (1.72°), and VI (3.44°), and size threshold perimetry (STP), a method that finds threshold by changing stimulus size.

METHODS. We derived normative limits for total deviation probability plots using the second method from 60 age-matched normals. We analyzed the probability plots of 120 glaucoma patients (mean deviation was −9.3 ± 6.1 dB with a range of −0.2 to −31.6) at the 42 nonblind spot locations common to the tests. We compared the number of abnormal test locations at the 5% level among the tests using one-way repeated measures ANOVA on ranks. We stratified the results by mean deviation.

RESULTS. There was a statistically significant difference in the number of abnormal test locations among the tests: III, 28.5; V, 29.7; VI, 27.0; and STP, 28.8. P = 0.001; Tukey pairwise comparisons were statistically significant for the assessments between sizes V and VI and between STP and size VI. When stratifying by mean deviation, with mild visual loss, size V was most sensitive, followed by STP; size VI appeared slightly less sensitive.

CONCLUSIONS. Size V and STP provide favorable stimulus methodology for detection of mild to moderate glaucoma. Size VI appears slightly less sensitive for glaucoma with mild loss.

Keywords: perimetry, visual testing, visual field, vision testing, stimulus size

Conventional automated perimetry, since its introduction in the late 1970s, has almost exclusively used the Goldmann size III stimulus. However, it has been shown that detection of defects from glaucoma and other optic neuropathies can be done at least as well with larger stimuli.1,2 In addition, these large stimuli have been shown to give better retest variability and extend the dynamic range of the test.3

In a study comparing response variabilities of Goldmann sizes III and V, we used frequency of seeing curves in glaucoma patients that were generated by using a custom test program to evaluate these patients at 2 dB intervals over a 2 to 3 log unit range. The patients were tested with size III and size V stimuli in areas of normal sensitivity and areas of 10 to 20 dB loss. The same test locations were used in the same sitting for both sizes. As shown by a steepening of the slope of the frequency of seeing curves, response variability substantially decreased when the size V stimulus was used. This study led us to test for differences in number of abnormal test locations found comparing Swedish Interactive Threshold Algorithm (SITA) Standard size III with Full Threshold size V. We found no clinically significant differences in the abilities of these test strategies and sizes to demonstrate visual loss.4

It has been reported that size V stimuli should not be used routinely because this strategy may fail to detect early visual loss.5 However, the original 10° frequency doubling technology stimulus is over 40 times the size of the 1.7° size V stimulus in area and Frequency Doubling Technology (FDT) testing is similar in sensitivity to conventional automated perimetry using a size III stimulus (0.43°) for glaucoma and other optic neuropathies.1,2,6

While increasing stimulus size from III to V in glaucoma patients does not appear to hinder defect detection,4 for visual field defects that have borders with steep slopes like some hemianopic scotomas or small focal retinal lesions, larger stimulus sizes may fail to detect this type of visual field defect.2 An ideal strategy would employ small stimuli in areas of normal or near normal sensitivity and larger stimuli in areas with more damage. This can be accomplished by finding threshold by changing size in increments, rather than changing luminance (size threshold perimetry, also called size modulation perimetry).7,8

Since larger perimetric stimuli can have greater dynamic range and lower variability, it is important to investigate how efficiently stimuli of different sizes discriminate between healthy and damaged visual fields. To test the hypothesis that the number of abnormal test locations found is no different for sizes III, V, VI, and size threshold perimetry (STP) in which stimulus size is changed rather than stimulus light intensity in glaucoma patients, we developed databases from the same set of healthy participants and generated empiric probability plots.
for these stimulus sizes using the same methodology. Our aim was to compare results of glaucoma patients with the database from the common set of healthy observers and confirm our findings with Size V stimuli, while extending our testing of the hypothesis to the larger size VI stimulus and STP.

**Methods**

**Subjects**

The visual testing protocol was approved by the University of Iowa institutional review board, and followed the tenets of the Declaration of Helsinki. Sixty ocularily healthy subjects and 120 glaucoma patients were tested at baseline and again at a separate sitting within 1 to 5 weeks. They all gave written informed consent to participate in the study. The healthy observers were volunteers, paid following protocol of Investigational Review Board, who answered advertisements inviting them to participate in research. The glaucoma patients were invited from the glaucoma clinic at the University of Iowa Department of Ophthalmology and Visual Sciences. The average age of the healthy observers was 61.2 with an SD of 8.9 years; the range was 42 to 79. Forty-three of the volunteers were women and 17 were men. Their mean deviation was 0.1 (SD 1.0) dB with a range of −2.4 to 2.1, (median 0.3; interquartile range or IQR 1.3), and pattern SD was 1.5 (SD 0.3 dB). The average age of the glaucoma patients was 67.0 (SD 9.6) years with a range of 47 to 86; 73 were women and 47 were men. Their mean deviation was −9.3 (SD 6.1) dB with a range of −6.2 to −31.6, (median −8.6; IQR 7.5), and pattern SD was 9.3 (SD 4.2 dB).

Healthy participants were included if they had (1) no history of eye disease and modest refractive error (no more optical correction than 5 diopters [D] of sphere or 3 D of cylinder), (2) no history of diabetes mellitus or systemic arterial hypertension, and (3) a healthy ophthalmologic examination including 20/25 or better corrected Snellen acuity. The subjects either had undergone a complete eye exam within 2 years prior to this study or were examined by an ophthalmologist on the day of testing. A perimetric examination was specifically not required.

Patients from the University of Iowa Hospitals and Clinics Glaucoma Service were offered admission if they met study entry criteria. They were enrolled if they had glaucomatous optic disc changes with abnormal conventional automated perimetry (glaucomatous visual field defects, i.e., three or more adjacent abnormal test locations in a clinically suspicious area at the P < 0.05 level or two adjacent locations abnormal with at least one at the P < 0.01 level). In addition, mean deviation was in the range of 0 to −32 dB on conventional automated perimetry. We included patients with primary, secondary, or normal tension glaucoma. The patients did not have another disease affecting vision and were capable of reliably performing conventional automated perimetry and returning for follow-up visits. Patients were excluded if they had cataract causing visual acuity of worse than 20/30, pupil returning for follow-up visits. Patients were excluded if they reliably performing conventional automated perimetry and secondary, or normal tension glaucoma. The patients did not automate perimetry. We included patients with primary, specifically not required.

Subjects were examined with conventional automated perimetry (stimuli sizes III, V, and VI), and with STP on the same day. STP is a method that uses a staircase procedure to determine a size threshold. The size threshold is defined as the smallest size stimulus seen by the subject at each test location where the subject made a response meeting prespecified criteria. The stimulus was a light gray filled circular patch of 80 apostilb; the background was a uniform (nonstippled) darker gray with a luminance of 50 apostilb. The stimuli were of 18 sizes, also with a diameter step factor of 10⁻¹. The angle subtended by the targets ranged from 0.13° to 8.46°. A trial proceeds as follows. First, a stimulus is displayed. Next the subject touches a light pen to the screen where the target was perceived. Last, the subject is given feedback about their performance with a cross hair shows the location of their light pen touch to the monitor relative to the target (a nonfilled circle). A 2/1 staircase procedure was used to estimate threshold. Further details can be found in a previous publication.⁷

Sizes III, V, and VI were tested with the Humphrey Field Analyzer (Carl Zeiss Meditec, Dublin, CA). The order of the testing using the four different methods was randomized. Size III testing was performed with the SITA standard 24-2 algorithm; sizes V and VI testing used the Full Threshold method. The size III stimulus is 0.45° in diameter, the size V is 1.72°, and the custom made size VI was 5.44°. Since there is no SITA strategy available for size V or size VI stimuli, the subjects were tested with the Humphrey 24-2 Full Threshold algorithm. We followed the manufacturer’s recommendations and used a corrective lens when necessary. Care was taken to prevent lens rim artifact. The healthy participants had testing in one eye chosen at random and the same eye was used for all four tests. The glaucoma subjects had one eye chosen that qualified for the study. If both eyes qualified, one eye was chosen at random. All visual field examinations met the following reliability criteria: fixation losses less than 20% or normal gaze tracking, a false positive rate less than 10%, and a false negative rate less than 33%. We define normal gaze tracking as the presence of only an occasional upward deflection representing an eye movement.

**Pointwise Probability Plots**

We computed normative limits and then generated empiric probability plots from data collected from testing 60 healthy participants with sizes III, V, VI, and STP testing. We used the results of second visual field examinations of these healthy observers and converted the data using PeriData (PeriData Software GmbH, Hürtel, Germany) to a spreadsheet format.

We then found the effect of age by regressing mean score on age. We found a loss of 0.067 dB/y for size III, 0.066 dB/y for size V, and 0.052 dB/y for size VI. The value for STP was 0.024 dB/y. We then adjusted all threshold values to age 45. The 60 values were then ranked from highest to lowest in a spreadsheet for each test location. Using the percentile function within the Microsoft Excel spreadsheet program (Microsoft Corporation, Redmond, WA), a percentile was calculated for each of the 60 ranked values and rounded to the nearest percentile. The 90th, 95th, 98th, and 99th percentiles were then empirically determined.

We then computed total deviation probability plots for each of the 120 glaucoma subjects representing all stages of glaucoma from 0.24 to 31.65 dB mean deviation (MD). We analyzed the probability plots of the 120 glaucoma patients by counting the number of test locations with loss at the less than 1%, less than 2%, less than 5%, and 10% levels and compared the number of abnormal test locations of the 42 test locations common to the four tests and the 52 locations common to the three Humphrey Field Analyzer tests. Since failure to detect a difference could be due to the fact that patients with moderate or severe visual field loss have many damaged test locations, we stratified the patients into 12 equal groups of 10 by mean
deviation. We then recounted the abnormal test locations by MD group.

We repeated this analysis for the pattern deviation results. This was done by calculating, for each location the sensitivity expected for the age of healthy control subjects. We then computed the total deviation values and found the 85th percentile of the total deviations, also called the general height (GH). We then subtracted the GH of the subject from the GH of the sensitivities expected by age computed above (GH shift).

While we used the second examination, differences between test 1 and test 2 were small. A small learning effect was present from test 1 to test 2 (Table 1). This is especially true for the healthy participants. We noted this previously and have attributed the presence of this small difference in mean sensitivity to transfer of training (learning) among the four tests that were given at each sitting in a random order.

### Statistical Analysis

Since the data were not normally distributed, repeated measures ANOVA on ranks was used to compare the number of normal and abnormal test locations among the three probability levels, groups, and visits. Tukey key tests were performed on the medians to evaluate pairwise differences. Differences between groups of all test results were interpreted as significant if the probability of their occurrence was less than 0.05. Mean sensitivities of the first and second tests were compared using a Wilcoxon rank sum test.

### RESULTS

With the glaucoma patients, we found there was a statistically significant difference in the number of abnormal test locations out of 42, at a \( P = 0.05 \) significance level, among the tests: III, 28.5; V, 29.7; VI, 27.0; and STP, 28.8, \( P = 0.001 \); Tukey pairwise comparisons were statistically significant for the comparisons between sizes V and VI and between STP and size VI (Fig. 1A). A similar pattern is present when comparing the 52 test locations common to the 24-2 pattern of sizes III, V, and VI (Fig. 1B). Similar findings are also present for the pattern deviation values (Figs. 2A, 2B).

We compared the differences in total deviation among sizes III, V, and VI (Fig. 3). We found the defect depth is greatest with size III followed by size V. Size V has slightly greater depth than size VI. Since the number of abnormal test locations are similar and we know larger sizes have lower variability it appears that the decrease in defect depth is approximately balanced by the reduction in variability (Table 2, Fig. 3). To evaluate the effect of visual field damage on number of abnormal test locations, we stratified the total deviation results by 12 mean deviation bins of 10 subjects each. Figure 4A shows the differences in the number of abnormal test locations flagged by the four types of probability plots. For the less damaged MD bins, size V appeared most sensitive. With mild glaucomatous loss, size VI appeared slightly less sensitive (3–4 test locations in magnitude per visual field) compared with size III. Figure 4B compares the 52 common test locations of the size III, V, and VI results; note the similar patterns of results. We analyzed the pattern deviation results in a similar fashion (Figs. 4C, 4D). Again, size VI appeared less able to detect abnormalities with STP and size V showing more abnormal test locations.

Figure 5 shows typical examples of our custom printouts that used the percentile values from the common set of Iowa normals. Both the number of abnormalities and the shape of the visual field defects are similar. This similarity of defect shape by probability plots was nearly always present when comparing sizes III, V, VI, and STP.

### DISCUSSION

Our results for visual field number of abnormal test locations in glaucoma show small, but statistically significant differences among tests of different stimulus size compared with a test that keeps luminance constant and changes stimulus size. It is not likely these differences are clinically significant except possibly for the decrease in detection using the size VI stimulus with early glaucomatous loss. We found Full Threshold size V testing
to be slightly more sensitive to detect glaucomatous loss than SITA Standard size III or Full Threshold size VI. Since various reports suggest slightly more abnormal test locations identified with SITA compared with Full Threshold methods, it is unlikely that the thresholding method is the reason that the results of these stimulus sizes are so similar. We also found changing stimulus size can perform as well as changing stimulus luminance for disease detection; this was especially true with the pattern deviation analysis.

The automated perimetry methodology that uses increments in size to find a visual threshold was introduced by Frisén in 1987. He used vanishing optotypes as a way to estimate peripheral acuity. The stimuli varied in size as $10^{-0.1}$. The test, high-pass resolution perimetry (ring test), continued, using a staircase procedure to estimate the visual threshold. The test was practical, efficient, and correlated highly with standard types of perimetry. Surprisingly, unlike conventional perimetry where a fixed Goldmann stimulus size III was used, variability did not increase substantially as sensitivity decreased. We have found similar results with two other types of perimetry where size is changed in small increments, STP, and motion perimetry. This property makes these tests strong candidates for longitudinal recognition of visual field progression since variability does not increase as visual function deteriorates. It is encouraging that we found STP can perform as well as changing stimulus luminance for detection of abnormal test locations in glaucoma.

**Figure 1.** Number of abnormal total deviation test locations per visual field of (A) 42 nonblind spot locations common to the four tests and (B) 52 nonblind spot locations examination from 120 glaucoma patients. Note the similarity of size III, V, and VI, and STP results when the same Iowa normal eyes are used to generate pointwise probability plots.
Our results for glaucoma using Goldmann stimulus size VI suggest that at some point, increasing stimulus size results in a decrease in the number of abnormal test locations. Another undesirable consequence of using large stimulus sizes is a reduced detection of some hemianopic defects. Since hemianopic scotomas often have very steep borders, subjects are able to see parts of the large stimulus leading to loss of defect detection and resolution.

However, using larger stimulus sizes has some distinct advantages. Variability is lower with larger sizes, often without a loss of signal (the difference between a healthy visual sensory apparatus and a damaged one) and the effective dynamic range is greater. Reducing measurement variability does not yield a net benefit if the signal is reduced in proportion. A reasonable compromise may be to use a method like STP that finds threshold by changing stimulus size. This includes the benefits of using large sizes coupled with the fine resolution of small defects enabled by using small stimulus sizes.

Why do larger stimulus sizes have these properties? Spatial summation is the property of the visual system that relates stimulus size to luminance; threshold intensity multiplied by the stimulus area equals a constant (Ricco's law). Spatial summation gradually increases with increasing distance from the fovea in healthy subjects likely due to the increases in receptive field size and overlap (density) and neural convergence. It has been proposed that the area of partial spatial summation, Ricco's area, increases with eccentricity to maintain a constant number of underlying retinal ganglion cells. It appears that with glaucomatous damage, a similar increase in Ricco's area occurs to maintain a constant number of receptive fields. Thus, we hypothesize that larger stimuli may still fall within Ricco's area of partial spatial summation and remain sensitive stimuli for disease detection. However,
Figure 3. Scatterplot of (A) total deviation values for individual size III thresholds plotted against size V total deviations. (B) Shows size III total deviations plotted against size VI and (C) shows the size V total deviation values plotted against size VI. Note the greatest defect depth for size III total deviation values followed by size V and size VI.
when the stimulus size exceeds Ricco’s area, oversampling of the receptive or perceptive field array may occur and optic nerve damage may not be detected. This may at least in part be reason for the poorer detection of the size VI stimuli in our cases of early glaucomatous damage.

When comparing perimetry types, most studies use the software of the manufacturer’s normative database rather than defining the reference standards for controls using a common set of healthy participants. Use of the manufacturer’s normative database is problematic due to different numbers of subjects included, use of one or both eyes, and wide differences in inclusion and exclusion criteria for the database. For example, the Humphrey Matrix required a normal Humphrey 24-2 visual field for inclusion. However, the Humphrey database for SITA Standard required only that the subject have a normal eye exam and no eye disease other than refractive error; all visual field examinations obtained were included if they met standard reliability criteria. Requiring a normal visual field eliminates poor test takers and tightens the confidence limits and increases the sensitivity to detect defects at the expense of specificity.

A weakness of our study is that we are comparing size III stimuli using SITA Standard and size V using a Full Threshold algorithm. However, since we are using a common database of healthy observers, there is no bias toward size V. Also, if anything, the SITA strategy might result in an increased number of visual field defects. Bengtsson and Heijl compared probability plot results between Full Threshold and SITA strategies in glaucoma patients. They found SITA showed a slightly larger number of significantly depressed points in the probability maps compared with the Full Threshold strategy. Another weakness is the small sample size of healthy observers and patients with minimal visual field damage. A larger study is necessary to determine if size V is as efficient as size III for identifying the conversion of ocular hypertension to glaucoma.

### Table 2. Summary of Total Deviations for Sizes III, V, and VI

<table>
<thead>
<tr>
<th></th>
<th>III</th>
<th>V</th>
<th>VI</th>
<th>III–V</th>
<th>III–VI</th>
<th>V–VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>−10.1</td>
<td>−8.5</td>
<td>−6.7</td>
<td>−1.6</td>
<td>−3.4</td>
<td>−1.7</td>
</tr>
<tr>
<td>SD</td>
<td>10.7</td>
<td>9.7</td>
<td>8.1</td>
<td>5.4</td>
<td>6.2</td>
<td>4.0</td>
</tr>
<tr>
<td>Median</td>
<td>−5.4</td>
<td>−4.9</td>
<td>−4.1</td>
<td>−0.5</td>
<td>−1.6</td>
<td>−1.1</td>
</tr>
<tr>
<td>Minimum</td>
<td>−53.8</td>
<td>−57.6</td>
<td>−58.4</td>
<td>−30.5</td>
<td>−31.1</td>
<td>−30.5</td>
</tr>
<tr>
<td>Maximum</td>
<td>6.0</td>
<td>4.3</td>
<td>5.4</td>
<td>30.4</td>
<td>29.2</td>
<td>28.2</td>
</tr>
<tr>
<td>IQR 25</td>
<td>−16.5</td>
<td>−9.8</td>
<td>−8.2</td>
<td>−3.2</td>
<td>−5.4</td>
<td>−2.9</td>
</tr>
<tr>
<td>IQR 75</td>
<td>−2.2</td>
<td>−2.3</td>
<td>−1.6</td>
<td>1.6</td>
<td>0.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Note the magnitude of the deviations is greatest for size III, and the SD and IQR is less for sizes V and VI.

### Figure 4.

(A) Results comparing the four tests at the 42 common test locations. Division of the sample into 12 bins of 10 after sorting by mean deviation shows little effect on counts of total abnormal test locations at the 5% level per visual field examination related to amount of visual field damage except for mean deviation less than −4.15 dB where size VI testing shows fewer abnormal test locations. (B) Shows similar results when comparing test with sizes III, V, and VI at the 52 test locations these tests have in common. Pattern deviation totals are shown for the 42 common test locations in (C) and the 52 common locations in (D).
Size V and STP provide favorable stimulus methodology for detection of mild to moderate glaucoma. Size VI appears slightly less sensitive for glaucoma with mild loss. Future studies should investigate glaucoma with minimal visual field damage to determine if these findings are confirmed. If they do, size V stimuli may be preferable to size III stimuli for glaucoma detection and follow-up.

Acknowledgments
Supported by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Rehabilitation Research and Development (RX000140-03), the Center for the Prevention and Treatment of Visual Loss, Department of Veterans Affairs Health Care System, Iowa City, Iowa, and by an unrestricted grant to the Department of Ophthalmology from Research to Prevent Blindness.

Disclosure: M. Wall, None; C.K. Doyle, None; T. Eden, None; K.D. Zamba, None; C.A. Johnson, None

References


