The purpose of visual field examination in clinical ophthalmology is threefold: detection of early defects, determination of specific patterns of visual field loss for differential diagnostic purposes, and monitoring of patients for evidence of progression, stability, or improvement of visual field loss. Traditionally, measurement of perimetric differential light thresholds with conventional or standard automated perimetry (SAP) has been used for these objectives. However, newer perimetric techniques using alternative stimuli and test conditions have recently become available. An example of this is frequency-doubling technology (FDT) perimetry, which assesses contrast threshold by using low-spatial-frequency and high-temporal-frequency test targets to elicit mechanisms that demonstrate nonlinear responses to contrast, which serve as the basis for the frequency-doubling effect.\(^\text{1,2}\) It has been suggested that the frequency-doubling effect is mediated by a subset of magnocellular (M) ganglion cells, referred to as M cells.\(^\text{3}\) This M-cell subset reportedly has larger soma and axon diameters than other M cells, exhibits relatively little receptive field overlap, and constitutes less than 5% of the total ganglion cell population. These physiological attributes make FDT perimetry a good psychophysical test candidate for detection of glaucoma, because of histologic data reporting that ganglion cells with larger diameter axons may be preferentially lost in early glaucoma.\(^\text{4,5}\) and also because the reduced redundancy hypothesis predicts that sparse subsets of ganglion cells will more readily detect early damage.\(^\text{6,7}\)

FDT perimetry has been shown to exhibit high discriminatory power for detection of early glaucoma\(^\text{8-11}\) and is clinically desirable, because the test is resilient to refractive errors and blur and has a large dynamic range, and because threshold test strategies are short in duration. In addition, in patients with glaucoma, it has been reported that test–retest variability that occurs with FDT perimetry does not increase as much with increases in defect severity as with SAP. Provided FDT perimetry demonstrates adequate sensitivity to change, its beneficial variability characteristics imply that it may be useful for the detection of glaucomatous visual field progression.\(^\text{12}\)

Two variability components have been described for threshold perimetry: within a single test session (intratest) and between test sessions (interetest). Tests that exhibit lower amounts of either of these variability components are likely to be of greater clinical value, because they may be better able to differentiate progressive visual function loss from pathophysiologic variability. Quantification of test–retest variability obtained from clinically applied thresholding algorithms has been reported for both SAP and FDT perimetry,\(^\text{12,13}\) but does not provide information on the described individual variability components. Although more rigorous psychophysical evaluations of intratest variability for both normal and glaucomatous visual fields have been reported for SAP,\(^\text{14-18}\) this information does not yet exist for FDT perimetry. The use of these more rigorous psychophysical test procedures has not been applied to determination of intertest variability for either SAP or FDT perimetry. Data in previous reports on intertest variability for SAP were obtained by using mathematical methods of extraction from threshold estimates derived from repeated staircase procedures,\(^\text{19-22}\) rather than by more extensive psychophysical approaches, such as longitudinal evaluation of precise threshold measurements using repeated frequency-of-seeing (FOS) curves.

The purpose of this investigation was to precisely quantify and compare intra- and intertest variability components for both SAP and FDT perimetry in a small group of normal individuals and in a group of patients with glaucoma.

---

From Discoveries in Sight, Devers Eye Institute, Portland, Oregon. Supported in part by Grant EY03424 from the National Eye Institute (CAJ). Submitted for publication June 12, 2000; revised August 31 and December 15, 2000; accepted January 12, 2001.

Commercial relationships: F (CAJ); C (CAJ); N (all others).

Corresponding author: Paul G. D. Spry, Bristol Eye Hospital, Lower Maudlin Street, Bristol BS1 2LX, UK. pgds@hotmail.com
MATERIALS AND METHODS

Subjects

Data were collected from one eye of seven patients with glaucoma (one man, six women; age range, 45–86 years) and eight normal individuals (four men, four women; age range, 21–50 years). Patients with glaucoma were recruited from Devers Eye Institute and normal individuals from staff members. All participants were selected on the basis of good reliability (<20% fixation losses, <35% false-positive errors, and <33% false-negative errors) on previous Humphrey Field Analyzer (HFA; Humphrey Systems, Dublin, CA) examinations.

A fellowship-trained glaucoma specialist diagnosed glaucoma in patients on the basis of optic nerve head assessment and visual field status. For patients with glaucoma, the range of mean deviations (MDs) on HFA program 24-2 (full threshold) examination was from −3.36 to −8.56 dB (mean, −5.77 dB). No patients with glaucoma were receiving miotic glaucoma therapy, and natural pupil size was used. Each normal individual had MD, pattern standard deviation (PSD), and corrected interquartile range to quantify intratest variability,15,16 and this variability, rather than glaucomatous progression.

Informed consent was obtained from each participant in accordance with the Declaration of Helsinki. The study was approved by the institutional review board of Legacy Health System.

Perimetric Instrumentation

Matched test locations were examined for both conventional and frequency-doubled stimuli. Subjects were appropriate refractive corrections for both tests. SAP testing was performed on an HFA (model 610; Humphrey) that was externally driven by a computer running custom software. Test conditions identical with routine testing were used: size III test target, 200-msec stimulus duration, and 31.5-astrobil (asb; 10 candelas [cd] m−2) background illumination. Frequency-doubling stimuli were presented on a 21-in. video monitor (Multiscan GS500; Sony, Tokyo, Japan) driven by a video board (VSG2/5; Cambridge Research Systems, Rochester, UK). Frequency-doubling stimuli were generated using the same spatiotemporal properties used by the commercially available FDT perimeter (0.25 cyc/deg spatial frequency sinusoidal waveforms and 25 Hz counterphase flicker). Mean luminance was 50 cd/m2. Other properties of frequency-doubling stimuli were also controlled to emulate the commercially available FDT instrument, including test target configuration (square 10° × 10°) and stimulus duration (720-msec total stimulus duration, with 160-msec linear on-ramp from 0% to tested contrast, 400-msec at test contrast, and 160-msec off-ramp returning to 0% contrast).

Procedure

The MOCS was used to examine three test locations between fixation and 20° in each subject. All subjects received MOCS training in a separate session (on different days) before data collection. In the group of normal individuals, test locations were chosen so that equal numbers of superior and inferior locations were used, with equal spread among eccentricities, thereby providing data representative of locations throughout tested areas of the visual field. In patients with glaucoma, test locations were selected in visual field areas that ranged between normal and moderately damaged (total deviation, up to −10 dB) on the basis of the most recent SAP examination. This total-deviation criterion was selected to avoid a floor effect that may have artifactually truncated variability measurements, had deeper defects been included.

Each subject was tested at these same three visual field loci weekly on five consecutive occasions. For each test location and test session, the MOCS range was centered using an average of two threshold estimates obtained with a 4- to 2-dB double-reversal staircase algorithm. For MOCS, seven stimulus luminances were examined, three each side of estimated threshold. Step sizes between stimuli were adjusted to approach both 0% and 100% seen, and ranged from 1 to 3 dB. Twenty stimuli were presented at each stimulus luminance. Stimuli were randomized between luminance levels and test location. Order of testing at each session was determined randomly for each subject. Timing of test during the day was not controlled. Subjects were given short rests at regular intervals, if required.

It is important to recognize that although SAP and FDT perimetry both make measurements of sensitivity in decibels, the two measurement scales differ conceptually and so have different ranges and number of intervals. Care should be taken to note that HFA decibels and FDT decibels are not equivalent. In this study, SAP sensitivity measurements used the proprietary logarithmic HFA scale of retinal sensitivity. Each interval on this scale represents a 0.1-log unit (1 dB) attenuation of the brightest stimulus that the HFA can present (10,000 asb). Higher numbers on this scale therefore represent dimmer stimuli (thus denoting higher sensitivity) and vice versa. This HFA decibel scale has a range of 4 log units (or 40 dB), with 0 dB being equivalent to a 10,000-asb increment stimuli and 40 dB being 1 asb on a 31.5-asb background.23 The scale used for FDT perimetry in this study is also logarithmic but is a decibels scale of FDT stimulus contrast sensitivity: 1 dB = log(1/contrat threshold) · 10. An FDT stimulus that is visible at the maximum stimulus contrast of 100% corresponds to an FDT scale measurement of 0 dB (log(1/0)) · 10, whereas a stimulus perceived at a contrast of 1% is equivalent to a FDT sensitivity of 20 dB (log(1/0.01)) · 10. The FDT scale therefore has a range of 2 log units (20-dB intervals) to represent the stimulus contrast range from 1% to 100% and uses 1-dB scale intervals. Because 1 dB on the HFA measurement scale is therefore fundamentally different from 1 dB on the FDT measurement scale, this difference precludes comparison of the instruments. In this study it was critical that comparison between the variability of the instruments be based on the number of scale intervals that characterize variability, although, regretfully, both instruments use the same decibel nomenclature.

It is also important to note that the FDT scale used in this experiment is not the same as the scale of the commercially available FDT device, which has been adjusted by a proprietary multiplicative factor to produce an HFA-equivalent range, although it uses the same number of scale intervals, as described earlier.

Data Analysis

FOS curves were constructed by fitting data using a cumulative gaussian function (Tablecurve 2D; SPSS Science, Chicago, IL). FOS curves were used to quantify threshold sensitivity (50% seen on FOS, in decibels), intratest variability (FOS interquartile range, in decibels) and intertest variability (interquartile range of five repeated threshold sensitivity determinations over 4 weeks, in decibels). Other groups have used interquartile range to quantify intratest variability,15,16 and this was selected for use in this study, because it represents a quantity that is common to the measurement scale of the instrumentation, thereby providing clinical context.

The mean threshold and mean intratest variability over five repeat visits were used in all analyses. It was assumed that by use of a relatively short duration for data collection (4 weeks) any change in threshold over the examination period could be attributed to intertest variability, rather than glaucomatous progression.

Data analysis was performed by computer (Intercooled Stata 5.0; Stat, College Station, TX, and SigmaStat 2.0; SPSS Science).

RESULTS

Figures 1 and 2 show FOS curve examples. Figure 1 shows FOS curves obtained from stimulus locations in a normal individual for SAP (right) and FDT perimetry (left). For both procedures, there was a slight reduction in the steepness of the FOS curve with increasing eccentricity. The FOS curves for FDT perimetry were also steeper than for SAP. Results for locations in a
patient with glaucomatous visual field loss are presented in Figure 2 for SAP (right) and FDT perimetry (left). The FOS curves for FDT remained steep, even in areas of visual field loss. Conversely, FOS curves for SAP demonstrate a considerable reduction in steepness for visual field locations with reduced sensitivity.

**Figure 1.** Examples of FOS curves (solid lines) from matched test locations in a normal individual. Data points (●) are shown for FDT perimetry (left) and SAP (right). Threshold was defined as the point of the curve at which 50% of stimuli were detected. Intratest variability was defined as the interquartile range of the FOS curve. For both tests, increased eccentricity was associated with sensitivity reduction and curve steepness.

**Figure 2.** Examples of FOS curves (solid lines) from matched, damaged test locations in a patient with glaucoma. Data are displayed as described in Figure 1. The curve gradients for FDT perimetry were not greatly different from those in normal individuals (Fig. 1), unlike the SAP FOS curves, which appeared shallower, indicating increased intratest variability.
For both normal individuals and patients with glaucoma, SAP exhibited significantly greater intratest variability than did FDT perimetry ($P < 0.001$, two-way ANOVA). For both SAP and FDT perimetry, intratest variability was found to be significantly greater in patients with glaucoma than in normal individuals ($P < 0.001$, Mann–Whitney rank sum test). However, the difference in mean intratest variability between normal individuals and patients with glaucoma was greater for SAP than FDT perimetry. Mean (95% confidence interval [CI]) SAP intratest variability was 1.5 dB (1.62, 1.38) in normal individuals and 6.3 dB (8.44, 4.16) in patients with glaucoma, compared with 1.0 dB (1.10, 0.90) and 1.5 dB (1.68, 1.32), respectively, for FDT perimetry (Table 1).

Figure 4 shows the relationship between sensitivity and intratest variability for both tests. For SAP (Fig. 4A) it can be seen that intratest variability appeared dependent on sensitivity. In patients with glaucoma, a strong relationship was found between sensitivity and intratest variability: Variability increased by $\approx$0.6 dB (95% CI, 0.42–0.83) for every 1 dB of sensitivity loss (Table 2). However, no such relationship was observed at normal sensitivity levels. The relationship between sensitivity and intratest variability for FDT perimetry is shown in Figure 4B (also Table 2) and demonstrates that intratest variability was independent of sensitivity for both normal individuals and patients with glaucoma.

For both perimetric test types, the amount of intratest variability was found to be significantly greater than intertest variability (SAP $P = 0.001$, FDT perimetry $P < 0.001$, two-way ANOVA). Mean SAP (95% CI) intertest variability was found to be 0.7 dB (0.84–0.56) in normal individuals and 2.5 dB (3.41–1.59) in patients with glaucoma (Table 1), compared with 0.7 dB (0.79–0.61) and 0.8 dB (0.97–0.65), respectively, for FDT perimetry. For both normal individuals and patients with glaucoma, SAP exhibited significantly greater intertest variability than FDT perimetry ($P = 0.003$, two-way ANOVA).

Figures 5A and 5B show the relationship between sensitivity and intertest variability for SAP and FDT perimetry (Table 3). From Figure 5A it can be seen that SAP intertest variability behaved in a manner similar to that described for intratest variability, in that the amount of variability was independent of sensitivity in normal individuals but was dependent on sensitivity in patients with glaucoma. This relationship between sensitivity and intertest variability in patients with glaucoma was such that intertest variability increased by $\approx$0.2 dB for every 1 dB of glaucomatous sensitivity loss. For FDT perimetry (Fig. 5B), intertest variability was independent of sensitivity and remained at $\approx$0.75 dB throughout the measured range of sensitivities.

**DISCUSSION**

At any given time point, the amount of physiologic visual system background noise against which a stimulus signal may be detected is variable. Over a short time, such as in a single

![FIGURE 3. Frequency distributions of sensitivities obtained in normal individuals (filled bars) and in patients with glaucoma (striped bars) for SAP (A) and FDT perimetry (B).](image)

**TABLE 1. Summary Data for Intra- and Intertest Variability for SAP and FDT Perimetry**

<table>
<thead>
<tr>
<th></th>
<th>Normal Individuals ($n = 24$)</th>
<th>Glaucoma Patients ($n = 21$)</th>
<th></th>
<th>Normal Individuals ($n = 24$)</th>
<th>Glaucoma Patients ($n = 21$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean intratest variability</td>
<td>1.5 ± 0.30</td>
<td>6.3 ± 5.01</td>
<td></td>
<td>1.0 ± 0.25</td>
<td>1.5 ± 0.42</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>1.62–1.38</td>
<td>8.44–4.16</td>
<td></td>
<td>1.10–0.90</td>
<td>1.68–1.32</td>
</tr>
<tr>
<td>Mean intertest variability</td>
<td>0.7 ± 0.36</td>
<td>2.5 ± 2.12</td>
<td></td>
<td>0.7 ± 0.22</td>
<td>0.8 ± 0.40</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>0.84–0.56</td>
<td>3.41–1.59</td>
<td></td>
<td>0.79–0.61</td>
<td>0.97–0.63</td>
</tr>
</tbody>
</table>

Data are mean decibels ± SD.
Test, detection threshold is therefore probabilistic, with larger signals having a higher likelihood of detection. This noise constitutes intratest patient response variability. In addition, it has been reported that longer term sensitivity changes, over hours or days, are also present that are superimposed on intratest variability. These longer term sensitivity modulations have been attributed to reversible ocular and neural sensitivity fluctuations and constitute intertest variability. Both these variability components affect the precision, or reproducibility, of visual field test results. Quantification of visual field test variability is important, because it directly influences the ability of the test to detect progression. Furthermore, this confounding influence is augmented by increases in both intratest and intertest variability in optic nerve diseases, such as glaucoma.

In clinical environments, tests with less variability are therefore more desirable because they do not require as much sensitivity shift to denote significant change or progression. In this view, previous investigators have reported that high-pass resolution perimetry (HRP) has lower variability than SAP and Chauhan et al. have reported that this results in HRP's being able to detect glaucomatous visual field progression an average of 12 to 18 months earlier than SAP. However, it is important to recognize that low variability does not ensure that a test is sensitive to change. Prospective longitudinal investigations are still needed to establish sensitivity to validate the technique as a clinically reliable means of detecting progressive loss.

The experimental design used in this study provided robust measurements of threshold and intratest variability, by using FOS curves. Our MOCS procedure was performed for both SAP

<table>
<thead>
<tr>
<th>Test</th>
<th>Group</th>
<th>Slope</th>
<th>95% CI (Slope)</th>
<th>P (Slope)</th>
<th>Y Intercept</th>
<th>95% CI (Y Intercept)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP</td>
<td>normal individuals</td>
<td>0.02</td>
<td>0.07, −0.04</td>
<td>0.56</td>
<td>0.93</td>
<td>2.92, −1.05</td>
</tr>
<tr>
<td>SAP</td>
<td>glaucoma patients</td>
<td>−0.05</td>
<td>−0.13, −0.07</td>
<td>&lt;0.001</td>
<td>1.14</td>
<td>−1.27, 1.08</td>
</tr>
<tr>
<td>FDT</td>
<td>normal individuals</td>
<td>0.04</td>
<td>0.13, 0.05</td>
<td>0.34</td>
<td>0.28</td>
<td>1.84, −1.27</td>
</tr>
<tr>
<td>FDT</td>
<td>glaucoma patients</td>
<td>−0.05</td>
<td>−0.12, 0.02</td>
<td>0.18</td>
<td>1.94</td>
<td>2.69, 1.19</td>
</tr>
</tbody>
</table>
and FDT perimetry to provide a quantitative comparison of their variability characteristics, while also yielding the first FOS curve analysis for the FDT stimulus. In addition, an alternative method of intertest variability quantification was achieved by collection of data for FOS curve construction at the same test locations on a number of visits. Measurement of intertest variability in this manner was considered preferable to the previously reported mathematical extraction techniques that were based on data gathered using staircase strategies and therefore do not have the accuracy and precision of FOS curves.12–25 It was interesting to note that these MOCS-based sensitivity estimates were, in some cases, lower than anticipated from the total-deviation inclusion criterion. This finding may be the result of error in threshold estimation by either strategy, although the rigorous nature of MOCS dictates that overestimation of sensitivity by the 4- to 2-dB double-reversal adaptive staircase strategy is the most parsimonious explanation.29

The results of this study are in agreement with reports that describe significantly increased levels of intra- and intertest variability with SAP sensitivity reductions in glaucoma.13,14,16,18,20 For FDT perimetry, although average levels of intra-and intertest variability were greater in patients with glaucoma than in normal individuals, the differences between the groups were small and were found to be independent of sensitivity. The amounts of both variability components in this sample were substantially and significantly less in FDT perimetry than in SAP. Practically, these findings imply that FDT perimetry may be a valid tool for detection of change in clinical situations, provided it can be prospectively shown to exhibit similar or better sensitivity to change than SAP.

A number of investigators have questioned the cause of increased variability in damaged visual field areas in SAP.12,18,30 Given that artifactual fixation errors have been ruled out,31,32 it may be postulated that factors responsible for increased variability include greater background variability noise—for example, atypical firing characteristics of damaged retinal ganglion cells or reduced pooling of ganglion cell response signal caused by lower ganglion cell densities resulting from glaucomatous cell death.33 Reports that SAP variability is independent of the cause of optic nerve damage34 and is reduced with larger stimuli35 are consistent with the latter of these hypotheses. For these reasons it has been suggested that the larger FDT perimetry stimulus size (10° compared with the 0.43° used in SAP) contributes to its lower variability.12,13 Despite the differences in relative density between sparse My-cells purported to produce response slope and abundant parvocellular ganglion cells that detect SAP stimuli. The rationale underlying the use of different stimulus sizes for SAP and FDT perimetry in this study was to permit generalization of the findings to clinical situations. Recent studies have shown that higher spatial resolution FDT perimetry can be performed reliably using 4° diameter stimuli in a pattern similar to the 24-2 used by the HFA.34 Studies evaluating the variability characteristics of these smaller stimuli are under way.

In this investigation, intratest variability was found to be significantly greater than intertest variability for both SAP and FDT perimetry. It is particularly important to emphasize this result, because many modern studies that quantify perimetric variability use test-retest analyses. Test-retest methodology represents a compound measurement of intra- and intertest variability, because results are obtained using thresholding strategies used by commercially available instrumentation, such as the adaptive staircase for SAP12 and the modified binary search35,36 for FDT perimetry.37 The results obtained by these strategies are therefore affected by both intra- and intertest variability components. For this reason, test-retest analysis may be mistakenly interpreted as a quantification of intertest variability alone. By demonstrating that intratest variability is a significantly larger variability component than intertest variability, intratest variability therefore contributes more to test-retest variability. It should also be noted that test-retest analyses are subject to variability induced by the thresholding strategy. Thresholding strategies that exhibit equal or less intratest variability, such as those based on maximum likelihood37 (e.g., the Swedish Interactive threshold algorithm [SITA]38 or ZEST for FDT perimetry39) may be of greater value for monitoring progression.

Although the comparison of the two perimetric techniques performed in this investigation suggests that FDT perimetry demonstrates more favorable variability characteristics than SAP, it is important to recognize that this finding may have resulted from differences in measurement scales between the two techniques. The measurement scale used for FDT perimetry was purposely designed to be equivalent to the 20-interval scale used by the commercial instrumentation, but was compared in this investigation with the 40-interval HFA scale. Therefore, although FDT perimetry appeared to exhibit less variability with this coarse measurement scale, use of such a scale may be less sensitive to progressive loss than a finer scale. The issue of whether FDT perimetry is more sensitive to change than SAP requires resolution by prospective longitudinal clinical investigation. In addition, the possibility that use of an equal number of scale intervals by both techniques may yield minimal differences in variability components should also be considered.

In summary, this study shows that under test and measurement scale conditions similar to those used in clinical situations, FDT perimetry exhibited significantly lower intra- and intertest variability than SAP. FDT perimetry may therefore be preferable for monitoring individuals with glaucoma, because less sensitivity change is required to exceed variability. However, longitudinal investigations are needed to establish the sensitivity of FDT perimetry to change. Also, intratest variability contributes more to total variability than intertest variability, implying that thresholding strategies with lower intratest variability used with either SAP or FDT perimetry will be able to differentiate significant change from variability. Thresholding strategies with lower intratest variability may have preferable discriminatory power for detection of progressive defects.

References


