Threshold and Variability Properties of Matrix Frequency-Doubling Technology and Standard Automated Perimetry in Glaucoma

Paul H. Artes, Donna M. Hutchison, Marcelo T. Nicolela, Raymond P. LeBlanc, and Balwantray C. Chauhan

PURPOSE. To compare test results from second-generation Frequency-Doubling Technology perimetry (FDT2, Humphrey Matrix; Carl-Zeiss Meditec, Dublin, CA) and standard automated perimetry (SAP) in patients with glaucoma. Specifically, to examine the relationship between visual field sensitivity and test-retest variability and to compare total and pattern deviation probability maps between both techniques.

METHODS. Fifteen patients with glaucoma who had early to moderately advanced visual field loss with SAP (mean MD, −4.0 dB; range, +0.2 to −16.1) were enrolled in the study. Patients attended three sessions. During each session, one eye was examined twice with FDT2 (24-2 threshold test) and twice with SAP (Swedish Interactive Threshold Algorithm [SITA] Standard 24-2 test), in random order. We compared threshold values between FDT2 and SAP at test locations with similar visual field coordinates. Test-retest variability, established in terms of test-retest intervals and standard deviations (SDs), was investigated as a function of visual field sensitivity (estimated by baseline threshold and mean threshold, respectively). The magnitude of visual field defects apparent in total and pattern deviation probability maps were compared between both techniques by ordinal scoring.

RESULTS. The global visual field indices mean deviation (MD) and pattern standard deviation (PSD) of FDT2 and SAP correlated highly (r > 0.8; P < 0.001). At test locations with high sensitivity (>25 dB with SAP), threshold estimates from FDT2 and SAP exhibited a close, linear relationship, with a slope of approximately 2.0. However, at test locations with lower sensitivity, the relationship was much weaker and ceased to be linear. In comparison with FDT2, SAP showed a slightly larger proportion of test locations with absolute defects (3.0% vs. 2.2% with SAP and FDT2, respectively, P < 0.001). Whereas SAP showed a significant increase in test-retest variability at test locations with lower sensitivity (P < 0.001), there was no relationship between variability and sensitivity with FDT2 (P = 0.46). In comparison with SAP, FDT2 exhibited narrower test-retest intervals at test locations with lower sensitivity (SAP thresholds <25 dB). A comparison of the total and pattern deviation maps between both techniques showed that the total deviation analyses of FDT2 may slightly underestimate the visual field loss apparent with SAP. However, the pattern-deviation maps of both instruments agreed well with each other.

CONCLUSIONS. The test-retest variability of FDT2 is uniform over the measurement range of the instrument. These properties may provide advantages for the monitoring of patients with glaucoma that should be investigated in longitudinal studies.

The large variability in damaged visual fields with standard automated perimetry (SAP) is one of the principal impediments in determining visual field progression in patients with glaucoma. Even with newer testing strategies such as the Swedish Interactive Threshold Algorithm (SITA), repeated threshold estimates from damaged areas of the field can fall throughout almost the entire measurement range, making it difficult to detect further change. Visual field tests are often repeated when there is evidence of change, and the protocols of most glaucoma-related clinical trials stipulate that endpoints for visual field progression be confirmed by several examinations. Variability can delay or prevent the detection of visual field progression with trend analyses (e.g., those based on regression of a visual field parameter over time) as well as with event analyses (such as the Glaucoma Progression Analysis of the Statpac software [Carl Zeiss Meditec, Dublin CA] which perform a point-by-point comparison between baseline and follow-up examinations). Despite its large test-retest variability, SAP remains the standard for clinical visual field examinations in glaucoma.

Frequency-doubling technology (FDT) perimetry is one of the newer psychophysical techniques for visual field examination. Empiric investigations have shown the first-generation FDT device to perform well at detecting glaucomatous visual field loss. Importantly, the variability characteristics of FDT perimetry appear more uniform across the measurement range than those of SAP. Response variability, measured by the slope of the psychometric function, was shown to be independent of visual field loss, and the test-retest intervals of the FDT1 were more uniform over the instrument’s dynamic range compared with SAP.

A second-generation FDT instrument (FDT2; Humphrey Matrix, Carl-Zeiss Meditec) has recently become available. Its stimuli, smaller than those of the original FDT device, permit a larger number of visual field locations to be examined, providing greater detail of the spatial distribution of visual field loss. In addition, thresholds are estimated by a maximum-likelihood strategy, which, by presenting a constant number of four stimuli at each location, ensures uniform test duration independent of the level of field loss.
The main objective of this study was to compare threshold estimates and their test–retest variability between FDT2 and SAP. To address the methodological issues that arise when tests are compared to an imperfect gold standard, we reduced the effects of variability by examining each patient six times with both techniques. Principal curve analysis\(^1\)\(^2\) (which accounts for variability in both the dependent and independent variables) was then performed to determine the relationship between threshold values obtained with both techniques. The test–retest variability of both techniques was investigated as a function of visual field sensitivity and established in terms of test–retest intervals and standard deviations (SDs). Total and pattern deviation probability maps obtained with both techniques were compared by calculating ordinal defect scores.

**Methods**

**Description of Techniques**

**Frequency-Doubling Technology Perimetry.** The Humphrey Matrix (FDT2) perimeter presents FDT stimuli on a cathode ray tube with a nominal background luminance of 100 cd/m\(^2\), mounted at optical infinity from the observer’s eye. With the exception of the foveal stimulus, the stimuli of FDT2 (A) are 5°-square windows of a vertical cosine wave grating with a spatial frequency of 0.5 cyc/deg (B), counterphase flickered at 18 Hz (C). The foveal stimulus is a circular patch with a diameter of 5°. The presentation time is 500 ms, including ramped onsets and offsets of 100 ms. The background luminance is 100 cd/m\(^2\). The SAP stimuli are circular luminance increments with a diameter of 0.45°, presented for 200 ms with sudden on- and offsets. The background luminance is 10 cd/m\(^2\). Grayscale levels (A) and amplitudes (B, C) are not to scale.

**Comparison of Threshold Estimates.** To establish the relationship between the threshold estimates of FDT2 and SAP, we compared the mean result of the six tests at those locations at which maximum and minimum luminance of the stimulus, respectively. The test locations of the 24-2 program of the instrument are shown in Figure 1.

**Standard Automated Perimetry.** SAP was performed with a Humphrey Field Analyzer (HFA; Carl-Zeiss Meditec), using the SITA-Standard program 24-2 with the standard Goldmann size III stimulus (diameter, 0.45°, see Fig. 1). The principles and properties of the SITA Standard strategy have been described previously.\(^1\)\(^6\)\(^–\)\(^18\) In summary, stimulus intensities are varied in steps of 4 and 2 db, and the final threshold estimates are obtained after maximum-likelihood calculations based on the patient’s responses and the prior PDFs. In contrast to the threshold strategy of the Humphrey Matrix, the SITA Standard strategy terminates once the threshold of each location has been estimated with a given confidence, after a variable number of presentations. The definition of the db scale in SAP relates to the brightest stimulus that the instrument is capable of displaying [dB = 10 log(maximum intensity/threshold stimulus intensity)], which, with the HFA, is 3183 cd/m\(^2\). The test locations of the HFA 24-2 program are shown in Figure 2, alongside those of the corresponding Humphrey Matrix program.

**Study Sample and Testing**

Fifteen patients with glaucoma (mean age, 66.3 years; range, 56.1–80.6) with early to moderate visual field loss (mean MD [SITA Standard], −4.0 db; range, 0.2 to −16.1) were recruited from the glaucoma clinics of the QEII Health Sciences Centre (Halifax, Nova Scotia, Canada). Criteria for inclusion in the study were a clinical diagnosis of open-angle glaucoma, refractive error within 5 D equivalent sphere or 3 D astigmatism, best-corrected visual acuity ≥6/12 (+0.3 logMAR), and prior experience with FDT1 perimetry and SAP. Patients were examined over three sessions within a period of 4 weeks. Within each session, the randomly selected study eye was examined twice with FDT2 (24-2 threshold test) and twice with SAP (SITA Standard; 24-2 test). The order of the tests was randomized, and a mandatory break of 6 minutes was given between examinations. All participants wore the appropriate refractive correction for each test. The study adhered to the tenets of the Declaration of Helsinki. The protocol was approved by the Queen Elizabeth II Health Science Centre Research Ethics Committee, and all participants gave written informed consent.

**Analyses**

**Comparison of Threshold Estimates.** To establish the relationship between the threshold estimates of FDT2 and SAP, we compared the mean result of the six tests at those locations at which

---

**Figure 1.** Spatial and temporal characteristics of the stimuli of FDT2 (left) and SAP (right). With the exception of the foveal stimulus, the stimuli of FDT2 (A) are 5°-square windows of a vertical cosine wave grating with a spatial frequency of 0.5 cyc/deg (B), counterphase flickered at 18 Hz (C). The foveal stimulus is a circular patch with a diameter of 5°. The presentation time is 500 ms, including ramped onsets and offsets of 100 ms. The background luminance is 100 cd/m\(^2\). The SAP stimuli are circular luminance increments with a diameter of 0.45°, presented for 200 ms with sudden on- and offsets. The background luminance is 10 cd/m\(^2\). Grayscale levels (A) and amplitudes (B, C) are not to scale.

**Figure 2.** Stimulus locations (right eye) of the 24-2 programs of the HFA (small circles) and the Humphrey Matrix perimeter (squares and large central circle). For the comparison between the threshold estimates, locations were excluded if the stimulus centers of the two techniques were not within 2° of each other (filled circles, shaded squares). Also excluded were the two locations in the vicinity of the blind spot.
the stimulus centers were within 2° of each other (Fig. 2, closed circles and unshaded squares). In addition to the test locations on either side of the vertical meridian, we excluded the two locations near the blind spot.

A principal curve implementation available in the open-source statistical environment R was used to derive the relationship between the two perimetric techniques. This mathematical method finds a best fit between two variables by minimizing the residuals perpendicular to the fitted curve in both the x and y variables. In contrast to the more familiar least-squares type of regression (which minimizes the residuals in the dependent variable only), the principal curve algorithm is not based on the assumption that the independent variable is measured without error, and it is therefore immaterial which one of the two perimetric techniques is represented on the x-axis. To avoid floor effects that occur when the lower limit of the dynamic range of either technique has been reached, we excluded the data from test locations at which one or more of the threshold estimates were 0 dB with either FDT2 or SAP.

**Test–Retest Variability.** Test-retest intervals describe the range within which the central 90% of follow-up thresholds are likely to fall, for each level of baseline threshold, if no real change has taken place. We derived such intervals for combinations of two baseline and two follow-up tests with FDT2 and SAP. Since there were no significant learning or fatigue effects within or between the three sessions with either of the two techniques (repeated-measures ANOVA of MD and PSD, P > 0.1), we treated the order of the six examinations as interchangeable, thus using all 90 possible combinations of independent baseline and follow-up pairs of tests. For example, a baseline threshold calculated from tests 1 and 2 was compared to all six possible follow-up thresholds obtained by pair-wise combination of tests 3, 4, 5, and 6. The test–retest intervals were then established by calculating the empirical 5th and 95th percentiles of the distribution of follow-up thresholds, stratified for the baseline value.

As a quantitative comparison of the relationships between threshold and test–retest variability with FDT2 and SAP, we investigated the SD of the six repeated threshold estimates as a function of their mean value at each test location, using linear regression analyses of log SD versus mean. Because the dB scales of FDT2 and SAP are based on two distinct definitions (provided in Description of Techniques), we first transformed all thresholds into instrument-independent units of log Weber contrast sensitivity to render the data numerically comparable. As the dynamic range of both perimeters is limited, the sensitivity of severely damaged visual field locations may not always be truly measurable. Some thresholds may be estimated at 0 dB, since the maximum stimulus contrast has been reached. Because this floor effect can lead to an artifactual decrease in the test–retest variability, the regression analyses were confined to locations at which all six threshold estimates were >0 dB. Similarly, if all six threshold estimates had the same value, the resultant SDs (0) may unduly bias the estimation of the underlying variability, and data from such locations were also excluded from the regression analyses of both techniques.

**Comparison of Total and Pattern Deviation Probability Maps.** For the comparison between the probability maps of FDT2 and SAP, we derived an ordinal defect score for each visual field test. Each test location was assigned a value ranging from 0 to 4 according to its probability (P > 5%, P < 5%, P < 2%, P < 1%, P < 0.5%, respectively) in the printout. These scores were then summed across the 52 test locations of the entire visual field (excluding the foveal test point and the two locations at and above the blind spot), and the global defect sums were compared between total and pattern deviation probability maps of FDT2 and SAP.

**RESULTS**

The global indices mean deviation (MD) and pattern standard deviation (PSD) obtained with FDT2 and SAP were closely related (r = 0.86, P < 0.001; r = 0.95, P < 0.001; respectively). The mean test time of FDT2 was 313 seconds (range, 302–321), independent of visual field damage as measured by MD (r = 0.95, P = 0.44). In comparison, the mean test time for SAP was 319 seconds (range, 259–413), increasing significantly with visual field damage (r = 0.81, P < 0.001).

**Comparison of Threshold Estimates**

The measurement scales of FDT2 and SAP appeared numerically similar; 90% of threshold estimates in our study were between 3 and 32 dB with FDT and between 5 and 32 dB with SAP. At test locations with high sensitivity (mean SAP threshold >25 dB), there was a close and approximately linear relationship between the mean thresholds of FDT2 and SAP (Fig. 3). In this range of sensitivities, the principal curve had a slope of approximately 2.0. At locations with lower sensitivity, however, the spread of the data points increased considerably and the curve became progressively shallower. In comparison to FDT2, SAP estimated a slightly larger proportion of absolute defects (proportion of threshold estimates at 0 dB, 3.0% vs. 2.2%, respectively with SAP and FDT2; P < 0.001).

**Test–Retest Variability**

**Test–Retest Intervals of Threshold Values.** With FDT2, the width of the test–retest intervals appeared nearly constant (~8 dB) across virtually the entire measurement range. In contrast, the intervals of SAP were narrow (~3 dB) with thresholds near 30 dB, but broadened considerably with lower thresholds, ranging over nearly 15 dB at locations with baseline values near 10 dB (Fig. 4).

**Parametric Analyses of Test–Retest Variability.** In contrast to SAP, which showed a substantial increase in test–retest variability with decreasing levels of visual field sensitivity, there was essentially no such relationship with FDT2. Whereas the relationship between sensitivity and test–retest variability accounted for approximately 40% of the observed variance with SAP, virtually none of the variance observed in the FDT2 data was explained by sensitivity (Fig. 5).
Comparison of Total and Pattern-Deviation Probability Maps

A comparison of the defect scores from the total and pattern deviation probability maps of FDT2 and SAP showed good overall agreement between both techniques (Fig. 6). In patients with relatively early SAP visual field loss (SAP total deviation defect scores <30), the total deviation probability maps of FDT2 appeared slightly less abnormal than those of SAP, but this finding did not reach statistical significance ($P = 0.06$, Wilcoxon test). No such systematic differences were seen with the pattern deviation analyses ($P = 0.78$, Wilcoxon test).

Case Examples

The examples given in Figures 7 and 8 illustrate some of the findings of our study. For clarity, only single test results are shown, but they are representative of all six examinations with both perimetric techniques. The first example (Fig. 7) illustrates discrepancies between the threshold estimates of FDT2 and SAP where the decibel values of the former were consistently higher than those of the latter. In the second example (Fig. 8), discrepancies in the opposite direction were observed in the superior nasal quadrant. At a location in the inferior temporal quadrant, however, the decibel values of FDT2 were consistently higher than those of SAP. In both examples, the probability maps of both techniques agreed closely with each other, although SAP flagged a slightly larger number of locations as outside normal limits.

DISCUSSION

Previous investigations with FDT perimetry have shown that the technique may possess several potential advantages over SAP. The second generation of instruments using this technology has now become available, and the objective of this study was to compare test results from FDT2 and SAP in patients with glaucoma. To gain precise estimates, both of thresholds and of test-retest variability, we examined a small group of patients with a rigorous protocol of six examinations with each technique.

For visual field locations with high sensitivity (>25 dB with SAP), our data showed a close association between the pointwise mean thresholds of FDT2 and SAP. Within this range, the relationship appeared nearly linear, with a gradient of 2, conforming to what would be expected from consideration of the techniques’ distinct definitions of the dB scale. With FDT2, a 1-dB change in threshold refers to a 0.05-log unit change in stimulus contrast, whereas with SAP, a 1-dB change refers to a contrast change of 0.1 log units. This means that, at test locations with early damage, visual field changes over time should be numerically twice as large with FDT2 compared with SAP. At visual field locations with lower sensitivity (<25 dB with SAP), the relationship was less linear, with a gradient of 0.88, conforming to a 0.05-log unit change in stimulus contrast with a 0.1-log unit change in threshold.

The parameters of the fitted function $\log(\text{CS}) = A \cdot \log(\text{threshold}) + \text{constant}$ are shown in the graph, along with the proportion of variance explained by the relationship ($R^2$). Test locations with floor effects (threshold values of $0 \text{ dB}$, all mean thresholds $<20$ and $<23 \text{ dB}$ with FDT2 and SAP, respectively) were excluded from the regression analyses (solid lines). The italic values above the abscissa give the number of data points with SDs of 0, which were also excluded from the regression. A small amount of noise was added to both plots to improve the visibility of overlapping data points.
dB with SAP), however, the dispersion of the data increased considerably, and the principal curve fit showed a markedly nonlinear relationship between the mean thresholds obtained with FDT2 and SAP. In several patients, we observed visual field locations with near-absolute losses with one technique but not the other. Of interest, the proportion of absolute defects (threshold estimates of 0 dB) was slightly lower with FDT2 compared to SAP (2.2% vs. 3%, \(P < 0.001\)), indicating that FDT2, despite its smaller contrast range, does not suffer from more substantial floor effects than SAP. Because each observation is the average of six tests, the large dispersion of the data from damaged visual field locations is more likely to reflect genuine differences caused by the psychophysical properties of the techniques’ stimuli (see Fig. 1) rather than random measurement variability. Owing to the large dispersion of the data, a simple conversion factor to translate from one scale to the other would offer limited practical utility outside the narrow range of near-normal threshold values. An important focus of future work will be to investigate systematically the sources of discrepancy—for example, the size and location of the visual field loss. Specifically, further work is needed to establish how FDT2 characterizes the small yet deep losses that may occur near the fixation point (Fig. 7).

The nearly uniform test-retest intervals of FDT2 contrasted with those of SAP, which expanded markedly at lower threshold values. These results, in conjunction with the regression analyses, show that the test-retest variability of FDT2 is remarkably constant over the entire measurement range of the instrument. These findings are in agreement with previous investigations on frequency-of-seeing curves obtained with stimuli similar to those of FDT2 and SAP. At test locations with low sensitivity (<25 dB with SAP), FDT2 showed narrower test-retest intervals than SAP, but the opposite was true at test locations with high sensitivity (>25 dB with SAP). Because the test-retest intervals are dependent on the measurement scales, the relationship between the threshold estimates of both techniques must be taken into account to interpret these findings. The relatively larger test-retest intervals of FDT2 in areas of high visual field sensitivity, for example, may be offset if changes over time are also larger with this technique, as is suggested by the relationship between the threshold estimates (Fig. 3). Longitudinal studies are needed to establish whether, and how, the variability characteristics of FDT2 translate into tangible benefits to the follow-up of patients with glaucoma. However, the nearly uniform test-retest variability characteristics of FDT2 make this test a promising candidate for such trials. Since our data were established from a small sample of patients experienced with perimetry, it is possible that the test-retest variability of less experienced patients is somewhat higher. This is unlikely, however, to influence the fundamental differences in the relationship between visual field sensitivity and variability with FDT2 and SAP observed in our study.
In contrast to the total deviation analyses, the defect scores are only in patients with very early visual field loss. Total deviation maps of both techniques were small and clinically relevant. However, the overall differences between the normative databases are unlikely to be the sole explanation for the latter explanation by recomputing normative values for the FDT2 and SITA-Standard had been established from the pattern deviation maps agreed closely between FDT2 and SAP.

In conclusion, our study showed that the test–retest variability of FDT2 perimetry is uniform over the entire measurement range of the instrument. In combination with the relatively lower variability of the threshold estimates in areas of visual field damage, these properties may allow an earlier and more accurate detection of visual field progression and therefore prove advantageous for the monitoring of patients with glaucoma. These benefits, however, can only be conclusively demonstrated in prospective longitudinal studies.

**Acknowledgments**

The authors thank William H. Swanson (SUNY, New York, NY) for critical comments on an earlier draft of the manuscript, Chris A. Johnson and colleagues (Discoveries in Sight, Portland, OR) for providing data from healthy control subjects examined with SITA-Standard and FDT2, and David B. Henson (University of Manchester, UK) for his software to digitize Statpac printouts.

**References**


**Figure 8.** Example 2: A patient with superior nasal step in the right visual field. A test location in the superior arcuate area had six repeated thresholds of 0 dB with FDT2, whereas the mean of the SAP thresholds was 10 dB (short arrows). Another location inferior to the blind spot had an FDT2 threshold of 11 dB (mean value of the six tests, 7 dB), whereas all but two of the six repeated SAP thresholds were 0 dB (long arrows).


