Test–Retest Variability of Frequency-Doubling Perimetry and Conventional Perimetry in Glaucoma Patients and Normal Subjects

Balwantray C. Chauban¹,² and Chris A. Johnson³,⁴

PURPOSE. To compare the test–retest variability characteristics of frequency-doubling perimetry, a new perimetric test, with those of conventional perimetry in glaucoma patients and normal control subjects.

METHODS. The study sample contained 64 patients and 47 normal subjects aged 66.16 ± 11.86 and 64.26 ± 7.99 years (mean ± SD), respectively. All subjects underwent frequency-doubling perimetry (using the threshold mode) and conventional perimetry (using program 30-2 of the Humphrey Field Analyzer; Humphrey Instruments, San Leandro, CA) in one randomly selected eye. Each test was repeated at 1-week intervals for five tests with each technique over 4 weeks. Empirical 5th and 95th percentiles of the distribution of threshold deviations at retest were determined for all combinations of single tests and mean of two tests, stratified by threshold deviation. The influence of visual field eccentricity and overall visual field loss on variability also were examined.

RESULTS. Mean test time with frequency-doubling perimetry in patients and normal control subjects was 5.90 and 5.25 minutes, respectively, and with conventional perimetry was 17.20 and 14.01 minutes, respectively. In patients, there was a significant correlation between the results of the two techniques, in the full field and in quadrants, whereas in normal subjects there was no such correlation. In patients, the retest variability of conventional perimetry in locations with 20-dB loss was 120% (single tests) and 127% (mean tests) higher compared with that in locations with 0-dB loss. Comparative figures for frequency-doubling perimetry were 40% and 47%, respectively. Variability also increased more with threshold deviation in normal subjects tested with conventional perimetry. In both patients and normal subjects, variability increased with visual field eccentricity in conventional perimetry, but not in frequency-doubling perimetry. Both techniques showed an increase in variability with overall visual field damage.

CONCLUSIONS. Frequency-doubling perimetry has different test–retest variability characteristics than conventional perimetry and may have potential for monitoring glaucomatous field damage. (Invest Ophthalmol Vis Sci. 1999;40:648-656)

Assessment of visual function in glaucoma has traditionally involved visual field testing with a white target on a white background. Over the past 30 years, many investigators have questioned whether conventional perimetry is the most sensitive and appropriate technique for detecting early retinal ganglion cell loss, which is the hallmark of open-angle glaucoma.

There is considerable evidence in human glaucoma¹–³ and in experimental models of pressure-related optic nerve damage⁴–⁷ that larger retinal ganglion cells may be preferentially damaged early in the disease process. These findings have led to the notion that testing visual functions mediated by magnocellular retinal ganglion cells (M-cells), which tend to have larger soma and thicker axons and mediate low spatial–high temporal frequency visual information,⁸ may detect the earliest functional loss in glaucoma. In this view, perimetric tests based on flicker,⁹–¹³ motion,¹⁴–¹⁶ and peripheral displacement sensitivity¹⁷,¹⁸ have been evaluated.

The frequency-doubling effect is a phenomenon that occurs when a low spatial-frequency (<4 cyc/deg) grating that undergoes high temporal-frequency (>15 Hz) counterphase flicker appears as a shimmering grating pattern with double the original spatial frequency.¹⁹,²⁰ The basis of frequency doubling is thought to reside in a subset of M-cells, called My-cells, which account for only 15% to 25% of the M-cell population, or 1.5% to 2.5% of the total ganglion cell population. Perimetric tests based on frequency doubling have been described by Maddess and Henry²¹ and more recently by Johnson and Samuels.²²
Psychophysical tests that are used for clinical purposes should have a minimum set of attributes before they are introduced for clinical testing. Among these are a sound psychophysical basis and rationale for being affected by a given ocular disorder; good sensitivity, specificity, and validity; large dynamic range to measure the clinical spectrum of disease; good patient acceptance (including short test time and ergonomic design); and threshold variability that does not change with the level of damage and visual field eccentricity. Unfortunately, with conventional perimetry, test-retest variability increases with defect depth and eccentricity. Practically, this means that it becomes increasingly more difficult to detect statistically significant change with moderate to advanced damage, because the test-retest confidence intervals can span almost the entire dynamic range of the instrument.

To date, only preliminary test-retest variability characteristics of frequency-doubling perimetry have been published. The purpose of the study was to investigate extensively the variability characteristics of frequency-doubling perimetry and conventional perimetry in a group of glaucoma patients with mild to advanced visual field damage and in normal control subjects.

**MATERIALS AND METHODS**

**Subjects**

The study sample contained patients with open-angle glaucoma and normal control subjects. Patients were recruited on a consecutive basis from the Nova Scotia Eye Centre and the Eye Clinic of the University of California Davis. Patients were considered stable and had no treatment changes during the study. Normal subjects were recruited from a pool of subjects undergoing psychophysical testing or were spouses of patients. The study protocol was approved by the Research Ethics Committee of the Queen Elizabeth II Health Sciences Center and the University of California Davis Institutional Review Board. Informed consent was obtained from all subjects, and the tenets of the Declaration of Helsinki were followed.

Inclusion criteria common to patients and normal subjects were best corrected visual acuity of 6/12 or higher and willingness to provide informed consent. Common exclusion criteria were chronic ocular disease (except glaucoma in the patient group), distance refraction of more than 6.00 D (equivalent sphere) or astigmatism of more than 2.00 D, and systemic disease or medication known to affect the visual field or the ability to participate in the study. The inclusion criteria for glaucoma patients were diagnosis of stable chronic open-angle glaucoma (including glaucoma with normal pressure and pseudoexfoliative glaucoma) made by glaucomatologists, glaucomatous optic disc damage including local or generalized loss of neuroretinal rim tissue, and repeatable glaucomatous visual field loss of at least 2 dB (mean deviation) and/or a Glaucoma Hemifield Test result outside normal limits on STATPAC analysis with the Humphrey Field Analyzer (Humphrey Instruments, San Leandro, CA). The inclusion criteria for normal subjects were normal results in an ocular examination, including intact neuroretinal rims with a cup-to-disc ratio asymmetry of 0.2 or less; intraocular pressure of 21 mm Hg or less; and a negative family history of glaucoma.

**Frequency-Doubling Perimetry**

Frequency-doubling perimetry was performed on a prototype device that was a precursor to the commercially available device (Welch Allyn Inc., Skaneateles Falls, NY; Humphrey Instruments, San Leandro, CA). The prototype contained a video monitor enclosed in a self-contained unit and had the same hardware and software compared with the commercially available device. The monitor was set at optical infinity and subtended an angle of 40° horizontally and 40° vertically. Subjects wore their distance correction when tested, and pupils were at least 3 mm in diameter.

The test contained 17 stimuli: 4 in each quadrant and 1 in the center. The 4 paracentral and 12 peripheral square locations subtended 10° × 10°, whereas the central circular stimulus subtended a radius of 5° (Fig. 1). The stimuli contained sinusoidal gratings with a spatial frequency of 0.25 c/deg. The gratings were counterphase flickered at 25 Hz. These spatio-temporal parameters are well within the limits at which normal subjects can observe the frequency-doubling effect throughout the central visual field.

![Figure 1. Stimulus pattern and eccentricity zones for frequency-doubling perimetry (left) and the conventional perimetry (30-2 program of the Humphrey Field Analyzer (Humphrey Instruments, San Leandro, CA); right).](image-url)
The frequency-doubling perimeter was operated in the threshold mode, in which thresholds are estimated by varying spatial contrast using a modified binary search staircase technique with four reversals and an upper/lower boundary constraint of 3 dB or less. The modified binary search procedure has been reported to have better accuracy and efficiency than conventional staircase procedures. Subjects were asked to respond when a “shimmering” grating was seen after a demonstration program was run showing the stimulus presentation at high spatial contrast. The total stimulus duration was 640 msec with a 160-msec linear on-ramp from 0% contrast to the tested contrast, followed by 400 msec at the tested contrast and a 160-msec linear off-ramp from the tested contrast to 0% contrast. The stimulus was presented in this manner to avoid temporal transience at the onset and offset of stimulus presentation. The interstimulus duration was randomly varied from 0.75 to 1.25 seconds (mean, 1.00 second) to avoid rhythmic responses. In the commercial device, the stimulus presentation was terminated if the response button was pressed (i.e., stimulus seen), whereas in the prototype the whole 640-msec duration was used. Because this difference existed only for stimuli reported as seen, we expected no difference in the performance characteristics between the two devices. Fixation was monitored using the Heijl-Krakau technique. Full details of the testing procedure and threshold estimation are published elsewhere.

Conventional Perimetry

Conventional perimetry was performed using program 30-2 of the Humphrey Field Analyzer, which tests 76 locations placed on a 6° grid in the central 30° of the visual field. The conventional bracketing strategy was used. Careful standardized instructions were provided to all subjects concerning when they were required to press the response button. The appropriate correction for the testing distance was used, and pupils were at least 3 mm in diameter.

Testing Protocol

All subjects underwent an ophthalmologic examination including slit lamp biomicroscopy, direct ophthalmoscopy, and refraction. If both eyes were eligible for the study, one eye was selected as the study eye and was the only eye that was tested. The order of testing was randomized (i.e., frequency-doubling perimetry or conventional perimetry first). Rest periods were allowed as required. Both tests were then performed in random order in four weekly sessions thereafter for five examinations using each technique over a 4-week period.

Data Analysis

Threshold deviations (difference between measured and expected age-corrected values) were determined for each examination with frequency-doubling perimetry and conventional perimetry. Left eye data were converted to a right eye format, and the two blind-spot locations from program 30-2 were removed. The correlations between threshold deviations for the whole visual field and quadrants were determined.

The empiric 5th and 95th percentiles (to compute the 90% confidence interval) of the distribution of pointwise threshold deviations at retest were then calculated for each combination of single tests and mean of two tests, stratified by threshold deviation. For example, for retest deviations of locations initially with 4-dB loss (threshold deviation, −4 dB), all locations with a threshold deviation of −4 dB in a given single baseline test, say the first, were pooled, and then the percentiles for the second, third, fourth, and fifth tests were determined. This procedure was then repeated with the second, third, fourth, and fifth tests as baseline; thus, all possible combinations of single tests were analyzed, and the mean 5th and 95th percentiles were determined. The same approach was used for all combinations with mean of two tests after a mean of two baseline tests.

To study the effects of visual field eccentricity on the loss-stratified percentiles for retest variability, the frequency-doubling perimetry field was divided into two zones: zone 1, the central and 4 paracentral stimuli, and zone 2, the 12 peripheral stimuli (Fig. 1). The 30-2 field was divided into three zones: zone 1, stimuli within 10° eccentricity; zone 2, stimuli between 10° and 20° eccentricity; and zone 3, stimuli outside 20° eccentricity (Fig. 1). Differences in variability (90% confidence interval) between the zones within a given technique were tested with the Wilcoxon (2 zones) or Friedman (3 zones) test. To study the effect of the overall extent of visual field...
damage (measured by the mean deviation index) on retest variability, the glaucoma group was divided according to three categories of severity with frequency-doubling perimetry: mild, moderate, and advanced with mean deviations of more than -3.5 units, between -3.5 and -6.0 units, and less than -6.00 units, respectively. Similarly, for conventional perimetry, the three severity categories were a mean deviation of more than -6.00 dB, between -6.00 and -12.00 dB, and less than -12.00 dB, respectively.

RESULTS

The sample contained 64 patients and 47 normal control subjects. Twenty-three of the patients and 25 of the normal control subjects were from Halifax, and the remaining 41 patients and 22 normal control subjects were from Davis. The mean (± SD) age of the patients and normal subjects was 66.16 ± 11.86 years and 64.26 ± 7.98 years, respectively (P = 0.315; not significant). Additionally, there were no age differences between the two clinics in either group (P > 0.320).

The distributions of the average mean deviation of the five tests in the patients and control subjects for frequency-doubling perimetry and conventional perimetry are shown in Figures 2 and 3, respectively. There were 22, 21, and 21 patients, respectively, in the mild, moderate, and advanced severity categories, determined by frequency-doubling perimetry. The frequency of patients in each severity category was identical with that determined by conventional perimetry. Although there was a strong association among the severity categories with the two techniques (P < 0.001; Pearson's chi-square test; Table 1), there were notable exceptions, particularly visual fields in the moderate damage category, with one technique classified in all three categories with the other technique and vice versa. Similarly, there was a strong correlation in the average mean deviations between the three techniques (r = 0.778; P < 0.001; Fig. 4); however, there were several patients, particularly in the midrange of mean deviations in whom the results of the two techniques did not correlate. There were also strong correlations (P < 0.001) in mean deviation in the superior nasal (r = 0.735), inferior nasal (r = 0.682), superior temporal (r = 0.475), and inferior temporal (r = 0.725) quadrants. We did not observe the same correlations in normal subjects in the full field (r = 0.072; P = 0.629; Fig. 4) or in the quadrants (P > 0.200), probably because of the limited response range for normal sensitivity, particularly for frequency-doubling perimetry.

The mean test time (±SD) with frequency-doubling perimetry was 5.90 ± 0.37 minutes in patients and 5.26 ± 0.49 minutes in normal subjects. The corresponding test times with conventional perimetry were 17.20 ± 2.09 and 14.01 ± 1.04 minutes, respectively.

In glaucoma patients, retest variability (estimated by the 90% retest confidence intervals) of conventional perimetry showed a strong dependence on the baseline threshold deviation (Fig. 5). Variability was reduced substantially when the mean of two tests was used for baseline and retest compared with that of single tests; however, the relative dependence of variability on threshold deviation remained unchanged. Because of the limitations of the dynamic range of the perimeter, variability artifactually reduced in locations with defect depths of more than 20 dB (threshold deviation less than -20 dB). With frequency-doubling perimetry, retest variability also reduced when the mean of two tests was used; however, the reduction seemed to be more modest compared with variability in conventional perimetry. Variability seemed to decrease artifactually at defect depths greater than 9 units (threshold deviations less than -9 units). With conventional perimetry, the 90% retest confidence interval for a baseline deviation of 0 dB was 9.60 dB in single tests and 6.4 dB in mean

<table>
<thead>
<tr>
<th>Table 1. Frequency of Severity Grades with Frequency-Doubling Perimetry and Conventional Perimetry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency-Doubling Perimetry</td>
</tr>
<tr>
<td>Mild</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>10.9</td>
</tr>
<tr>
<td>25.0</td>
</tr>
</tbody>
</table>

Values in parentheses are percentages.
tests. With a baseline deviation of $-20$ dB, the confidence intervals were 21.10 dB (120% increase compared with the variability with a baseline deviation of 0 dB) and 14.51 (127% increase) for single and mean tests, respectively (Fig. 5). With frequency-doubling perimetry, the retest confidence interval for a baseline deviation of 0 units was 5.45 and 4.19 units for single and mean tests, respectively, whereas with a baseline deviation of $-9$ units the respective confidence intervals were 7.62 units (40% increase, Fig. 5) and 6.14 units, respectively (47% increase, Fig. 5).

Although the range of threshold deviations was substantially less in normal control subjects, a similar analysis was possible (Fig. 6). In frequency-doubling perimetry and conventional perimetry, the retest variability decreased only modestly when the mean of a pair of tests, rather than that of single tests, was used. The worst threshold deviation at which it was possible to estimate variability in single and mean test results was $-2$ units and $-5$ dB, respectively. For conventional perimetry, the 90% confidence interval for retest variability when the baseline deviation was 0 dB was 7.00 dB in single tests and 5.70 dB in mean tests. With a baseline deviation of $-5$ dB, the retest confidence interval was 11.00 dB (44% increase compared with the variability for a baseline deviation of 0 dB) in single tests and 8.90 dB (56% increase) in mean tests. In frequency-doubling perimetry, the retest confidence interval for a baseline deviation of 0 units was 3.98 and 3.63 units in single tests and mean tests, respectively, whereas with a baseline deviation of $-2$ units, the respective confidence intervals were 4.83 units (21% increase) and 4.19 units (15% increase).

To examine the effect of visual field eccentricity on retest variability, we repeated the analysis just described but divided the visual field into zones when using either technique (Fig. 1). At baseline threshold deviations at which we could estimate the 90% confidence intervals for variability for all three zones in conventional perimetry, there was an increase in variability ($P < 0.001$; Friedman test) with eccentricity in glaucoma.

![Graph showing the correlation in mean deviation in five tests between frequency-doubling perimetry and conventional perimetry in glaucoma patients (left) and normal control subjects (right).](image)

**FIGURE 4.** Scattergram showing the correlation in mean deviation in five tests between frequency-doubling perimetry and conventional perimetry in glaucoma patients (left) and normal control subjects (right).

**FIGURE 5.** Retest variability distributions (5th and 95th percentiles) of threshold deviation as a function of baseline threshold deviation in glaucoma patients for single test and the mean of two tests with frequency-doubling perimetry (left) and conventional perimetry (right).
Frequency-Doubling and Conventional Perimetry Variability

FIGURE 6. Retest variability distributions (5th and 95th percentiles) of threshold deviation as a function of baseline threshold deviation in normal control subjects for single tests and mean of two tests with frequency-doubling perimetry (left) and conventional perimetry (right).

FIGURE 7. Retest variability distributions (5th and 95th percentiles) of threshold deviation as a function of baseline threshold deviation in glaucoma patients for a mean of two tests stratified by eccentricity (see Fig. 1) with frequency-doubling perimetry (left) and conventional perimetry (right).

patients (Fig. 7). Although the maximum horizontal or vertical eccentricity of frequency-doubling stimuli was 20° (compared with 27° in conventional perimetry), there seemed to be a minimal change in variability (P = 0.214; Wilcoxon test) when locations with similar threshold deviations in the two zones were compared (Fig. 7). In normal subjects, similar results were observed with conventional perimetry in which visual field eccentricity had a notable effect on variability (P < 0.001; Friedman test) in locations with similar threshold deviations (Fig. 8). With frequency-doubling perimetry, the effects of eccentricity were muted (P = 0.225; Wilcoxon test; Fig. 8). Variability with conventional perimetry increased with eccentricity in glaucoma patients and control subjects, even when zone 3 was removed from the analysis (P = 0.231 and 0.180, respectively; Wilcoxon test).

DISCUSSION

There is considerable debate on whether damage in early glaucoma affects M-retinal ganglion cells preferentially1-5 or all ganglion cell subtypes nonpreferentially.28-30 Despite this controversy, it is important to point out the distinction between preferential anatomic loss and preferential psychophysical loss believed to reflect the functional properties of ganglion cell subtypes. In the primate retina, parvocellular ganglion cells (P-cells) compared when using either technique, we noted a shift in the distribution of all percentiles toward values that indicated more damage (Fig. 9). In other words, locations with similar baseline threshold deviations tended to have progressively lower threshold deviations (lower sensitivity) on retest as the overall visual field sensitivity decreased.
are the most predominant ganglion cell subtype accounting for approximately 80% of ganglion cells, whereas M-cells account for approximately 10%. Testing M-cell function would be expected to be more sensitive in detecting preferential M-cell loss or a nonpreferential loss of all subtypes, because M-cells are sparse, and even a small percentage loss in this case would stress a mechanism that is already sparsely represented or has reduced redundancy. Because frequency doubling is thought to be mediated by a subset of M-cells, namely My cells, which are extremely sparse, psychophysical tests that invoke their function should theoretically be exquisitely sensitive in detecting the earliest functional glaucomatous loss.

In this study, we compared the test-retest variability characteristics of frequency-doubling perimetry with that of conventional automated perimetry. Because the testing interval between the five tests was short (approximately 4 weeks) and the glaucoma patients were considered stable, the variability estimates probably reflected true physiological variations from one test to the next in the absence of progression, although over a longer period the estimates may be higher. This component of variability is of paramount clinical importance in the detection of progression, because if any measured change from one test to the next exceeds the confidence intervals of the expected physiological test-retest variability, there is an increased likelihood of true change. The results of this study confirm those in previous reports showing that in conventional perimetry, the extent of local variability increases with the extent of corresponding local damage. In practical terms this finding implies that it is easier to detect changes in locations that are mildly damaged than in ones that have more extensive damage. Furthermore, because visual field instruments have a finite dynamic range, the measured variability in locations with advanced damage is artifactually truncated, adding a further limitation.

**Figure 8.** Retest variability distributions (5th and 95th percentiles) of threshold deviation as a function of baseline threshold deviation in normal control subjects for a mean of two tests stratified by eccentricity (see Fig. 1) with frequency-doubling perimetry (left) and conventional perimetry (right).

**Figure 9.** Retest variability distributions (5th and 95th percentiles) of threshold deviation as a function of baseline threshold deviation in glaucoma patients for a mean of two tests stratified by overall visual field damage (see the Methods section) with frequency-doubling perimetry (left) and conventional perimetry (right).
The source of increased threshold variability with defect severity is not fully understood. Some investigators have suggested that small fixation errors and therefore errors of stimulus projection on the retina may cause increased variability, especially if the stimulus projects near a scotoma edge.35–37 Recent studies with eye trackers suggest, however, that clinically acceptable fixation errors probably contribute only minimally to increased variability.38,39 That variability increases with defect severity suggests that the likely cause is a decreased signal-to-noise ratio due to stimulus undersampling caused by ganglion cell loss.40 Recently, Wall et al.41 showed that variability with a size V Goldmann stimulus was considerably less than that with either size III or size I stimuli, adding support to the notion of a pathophysiological source of increased variability in glaucoma patients. Fatigue in subjects also probably plays an important role in determining the level of variability and would presumably be reduced with techniques offering shorter test times.

We found that visual field locations in normal subjects, which have lower sensitivities than the age-corrected normal values, tended to have higher retest variability (Fig. 6). This finding suggests a physiological source of increased variability possibly related to reduced ganglion cell density in areas with reduced sensitivity. We also found an eccentricity-dependent increase in variability with conventional perimetry, again suggesting that the reduced ganglion cell count with increasing eccentricity42 leads to an increased retest variability in normal subjects (Fig. 8).

Our study shows that test-retest variability in glaucoma patients with frequency-doubling perimetry does not increase as much with defect severity as it does with conventional perimetry. The increase in retest variability from a baseline defect with a deviation of 0 dB (normal compared with age-matched control subjects) to one with a deviation of −20 dB was approximately 120%, whereas for a similar comparison with frequency-doubling perimetry the increase was approximately 40%. In practical terms, these observations suggest that with frequency-doubling perimetry, the change required from one test to another to measure statistically significant change is more similar across the range of defect depths, than with conventional perimetry. The effects of eccentricity on variability are also less pronounced with frequency-doubling perimetry compared with conventional perimetry, though it should be pointed out that because of the current stimulus configuration with frequency-doubling perimetry, the eccentricity effects beyond 20° are not known. From a physiological perspective, therefore, our data suggest that frequency-doubling perimetry is less prone to undersampling than is conventional perimetry. The large target size used by frequency-doubling perimetry is probably the major factor responsible for the smaller effects of defect depth and eccentricity on variability. A potential disadvantage, however, of the large stimulus size is oversampling of retinal ganglion cell receptive fields, which may mask changes occurring over a small area.

Because the two tests have fundamentally different psychophysical bases, it is impossible to make a direct comparison of the variability between the two techniques. Although there was a statistically significant correlation between the results obtained with the two techniques in patients (Fig. 4), this observation was not surprising. Visual field defects ranged from early to advanced encompassing a good portion of the dynamic range of the perimeters. Therefore, the correlation coefficient would be expected to be high despite the notable spread of the data points in Figure 4. The discordance between the results of the techniques was especially evident in moderately damaged eyes (Fig. 4; Table 1), in which the two techniques may be measuring different mechanisms of the manifestations of the glaucomatous process. Because of these findings and the likelihood that the correlation between the results of the two techniques may be fundamentally different in health and disease, caution should be used in attempting to develop a conversion scale of one technique to the other.

In summary, our study shows that frequency-doubling perimetry has considerably different test-retest variability characteristics compared with conventional perimetry because of its different psychophysical characteristics. These characteristics may be more useful in the detection of glaucomatous progression. Frequency-doubling perimetry has been shown to have high specificity and sensitivity for detecting glaucomatous visual field loss.22 Our study has shown that it has a significantly reduced test time (by a factor of 2.9 in patients and 2.7 in control subjects) compared with the standard thresholding procedure with conventional perimetry and good patient acceptability. Further research is now necessary to determine whether frequency-doubling perimetry can detect glaucomatous changes earlier than conventional perimetry in prospective trials.

References


