Journal of Glaucoma 11:511–516

The Learning Effect in Visual Field Testing of Healthy Subjects Using Frequency Doubling Technology

Amjad Horani, BMSc, Shahar Frenkel, MD, PhD, Claudia Yahalom, MD, Marilyn D. Farber, DrPH, Uriel Ticho, MD, and Eytan Z. Blumenthal, MD

Department of Ophthalmology, Hadassah University Hospital and Hebrew University/Hadassah Medical School, Jerusalem, Israel

Purpose: To evaluate the presence, duration and magnitude of a learning effect in serial visual field (VF) testing, using the commercially available frequency doubling technology (FDT) instrument.

Patients and Methods: 21 healthy adults with no prior VF experience underwent 6 serial VF tests, using the full-threshold C-20 program of the Zeiss-Humphrey FDT analyzer, on one randomly chosen eye. Tests were spaced at least two days apart.

Results: The average mean sensitivity was 32.37 ± 2.6 dB; the average mean deviation (MD) was 1.22 ± 1.8 dB. The MD at the first examination (0.28 ± 2.1 dB) was significantly poorer than at any of the other testing sessions ($p<0.003$). Similarly, the mean sensitivity at the first examination (31.16 ± 3.0 dB) was significantly lower than any other testing session ($p<0.004$). The proportion of improvement from the first to the second session was 63% and 65% of the total improvement, for mean sensitivity and MD, respectively. Mean test duration showed a modest reduction, from 4.40 ± 0.3 minutes in the first session to 4.17 ± 0.4 minutes in the last session ($p = 0.023$). A sub-analysis comparison of the different VF segments showed a more prominent learning effect in the peripheral and nasal visual segments ($p<0.0001$).

Conclusion: Baseline measurements should best rely on the second testing session, since MD and mean sensitivity are somewhat poorer when subjects with no prior VF experience are first tested on the FDT instrument. This may be especially true for the purpose of following patients over time.

Key Words: Frequency doubling technology—Glaucoma—Glaucoma-diagnosis—Learning effect—Visual fields.

Glaucoma is second only to cataract as the leading cause of preventable blindness in the world.[1] It is estimated that over 65 million people throughout the world are affected by this disease.[2] Because glaucoma is largely asymptomatic until the very late stages of the disease, early diagnosis, relying on simple cost-effective screening methods, is highly desirable.

Automated, achromatic, static perimetry is currently considered the gold standard in testing the visual field (VF). However, this technique is rather time consuming and necessitates a relatively expensive instrument, usually placed in a dedicated room. Furthermore, it provides a limited sampling of the VF.[3]

Newer VF techniques are continuously being developed to improve the diagnostic yield and shorten the test duration. The FDT is one such technique, targeting a subpopulation of retinal ganglion cells, referred to as M-type cells, that may be preferentially damaged in early glaucoma.[4–6] FDT has been reported to detect early glaucomatous VF loss.[7,8] The frequency-doubling phenomenon, initially described by Kelly,[9,10] constitutes a counter flicker of rapidly alternating black and white

Received May 7, 2002, sent for revision May 29, 2002; accepted July 22, 2002.

Address correspondence and reprint requests to Eytan Z. Blumenthal, MD, Department of Ophthalmology, Hadassah University Hospital, P.O. Box 12000, Jerusalem 91120, Israel. E-mail: eblumenthal@md.huji.ac.il

bars, thus giving an illusion of seeing twice as many bars. Maddess[11] suggested that this stimulus preferentially stimulates a subset of the large retinal ganglion cells known as My cells. These My cells, part of a larger ganglion cell population called magnocellular cells, constitute approximately 3% of the total ganglion cell population. The My cells (and probably the larger population of the M-cells[12]) are thought to be sensitive to low-spatial and high-temporal frequency.[7,11] Their scarcity and presumed vulnerability makes them ideal for detecting early dropout of ganglion cells. Thus, damage due to glaucoma may be identifiable earlier than with other perimetry techniques. In contrast, Anderson and Johnson[12] have recently shown that at a temporal frequency of 25 Hz or greater, there is evidence that the FD stimulus-isolated mechanisms are similar to those isolated by a spatially uniform flickering patch.

Wood et al.[13] showed that healthy inexperienced subjects tend to present artificial field defects, when tested using standard visual field perimetry, which in a glaucomatous patient may be added to their true VF defects. These artificial defects tend to disappear in subsequent tests, as the subject acquires experience in performing the examination. The learning effect for automated static perimetry was extensively studied[14–16] showing an increase in the absolute mean sensitivity with subsequent examinations, especially between the first and second examination.

Earlier studies on the FDT have shown this technique to be highly sensitive and specific to glaucomatous loss.[17–19] This high specificity and sensitivity could be undermined by a learning effect that, if present, could offset the baseline examination, thus minimizing the possibility of detecting small changes in the VF. In addition, and perhaps just as important, a learning effect would tend to increase the false-positive rates for inexperienced examinees, and therefore be detrimental to the implementation of any technology used as a screening device.

Iester et al.[20] and Fujimoto et al.[21] have previously addressed the learning effect in FDT VF testing, and their findings are discussed. We set out to determine the extent of the learning effect phenomenon in VF testing using the FDT, and to seek characteristic topographic patterns of improvement throughout the FDT VF grid.

PATIENTS AND METHODS

Twenty-two healthy individuals were recruited from the general population, including hospital staff and their family members, but patients or family members of patients were not recruited. Prior to commencing the study, each subject underwent a full eye examination that included intraocular pressure (IOP) measurement, slit-lamp evaluation, and funduscopy. Excluded from the study were those with chronic or persistent ocular condition, apart from refractive errors; a history of glaucoma or ocular hypertension (IOP > 21 mmHg); those who underwent ocular surgery, including refractive surgery; a history of chronic corticosteroid treatment; those who had an active migraine or had suffered a migraine attack within the last 12 months prior to the study[20]; or those with any abnormal findings upon entry eye examination. Additional exclusion criteria were refraction errors outside ±7 D,[22] best-corrected visual acuity of worse then 6/9, or any previous VF testing using standard static perimetry, kinetic perimetry, FDT, or any other perimetric testing. VF testing was not used as an inclusion-exclusion criteria to avoid any bias on the aim of this study. Informed consent was obtained from all participants and the Hadassah Hospital Human Subject Committee approved the methodology.

One eye was chosen at random for each volunteer, and only that eye was tested throughout the study. Each subject received a brief explanation about testing VFs, the device, and the examination procedure. The subjects were not informed about the purpose of the study to avoid any bias. They were merely and repeatedly asked to do their best at each and every examination. All subjects were examined six times, twice weekly, using the commercially available FDT perimeter (Welch Allyn, Skaneateles, NY, and Zeiss-Humphrey, Dublin, CA). Subjects could choose whether to perform the test with either their habitual glasses or without them, as myopia of up to −12 D was shown to have no effect on the FDT test results.[22] Of the 22 subjects who began the study, 1 subject dropped out after undergoing 3 VFs (claiming boredom and disinterest in continuing the study) and was not included in the analysis.

The C-20 full-threshold program (FDT/VF software version 2.60/1.00) was used for all examinations. The grid pattern is made of 16 square test locations, each 10° across, encompassing 20° in each direction, and a central (foveal) circular grid location. Each test area shows a black-and-white sinusoidal grating (0.25 cycles/deg) flickering at 25 Hz. The contrast between the dark and white bars is varied throughout the test according to the subject's response. Threshold values are then determined at each location from the log contrast sensitivity, and expressed in decibels. Paired *t* tests, Pearson product-moment correlations, and ANOVA were used for statistical analysis. A second-degree polynomial fit was used in the graphs.

TABLE 1. *Mean values for the six test sessions*

	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Total
MD ± SD	0.28 ± 2.1	1.12 ± 16	1.12 ± 1.6	1.26 ± 1.6	1.86 ± 2.0	1.58 ± 1.7	1.22 ± 1.8
Mean sensitivity ± SD	31.16 ± 3.0	32.24 ± 2.4	32.38 ± 2.3	32.41 ± 2.6	33.17 ± 2.8	32.88 ± 2.4	32.45 ± 2.6
PSD ± SD	3.72 ± 0.8	3.25 ± 0.7	3.42 ± 0.6	3.60 ± 0.9	4.17 ± 2.1	3.42 ± 0.9	3.61 ± 1.2

MD, mean deviation; PSD, pattern standard deviation.

RESULTS

Twenty-one subjects (8 females and 13 males) completed all six VF examinations. The mean age was 32.8 ± 13.3 years (range, 23–60 years). Ten eyes were emmetropic and the remaining 11 had a mean refractive error of −2.07 D (range, −0.75 to −4.5 D). A single technician collected data for the entire group within a 4-week period using one FDT unit.

The mean false-positive error rate was 2%; no false-negative errors were encountered. The mean fixation loss error rate was 8%. Data for the MD, mean sensitivity, and pattern standard deviation (PSD) for each of the six sessions is shown in Table 1.

To test for a learning effect, the MD was plotted against each test session (Fig. 1). The MD, lowest at the first examination (0.28 ± 2.1 dB), showed a tendency toward a plateau in the last three tests. There was a statistically significant difference between the MD between the first and second sessions (0.28 ± 2.1 vs 1.13 ± 1.6, respectively; $P = 0.0032$). In contrast, the differences in MD for all other consecutive sessions were not statistically significant. An ANOVA performed on the six test sessions was found to be statistically significant ($P = 0.0072$), and ceased to be significant once the first session was omitted.

Mean sensitivity was calculated as the average of all 17 threshold values. Mean sensitivity represents the raw threshold decibel. This is in contrast to the mean deviation value, which is a comparison against the age-adjusted expected value, as defined by a normative database. When a learning effect was sought using the mean sensitivity data, a significant difference was found only for the first pair (first vs second session, $P = 0.0031$). None of the other consecutive pairs was statistically significant (Fig. 2). Again, an ANOVA performed on the six test sessions was found to be statistically significant ($P = 0.017$), and ceased to be significant once the first session was omitted. No learning effect could be found for the PSD, which fluctuated randomly between sessions. The PSD was lowest on the second session and highest on the fifth session.

On average, each examination was performed faster than the preceding one ($P = 0.023$). As far as test duration is concerned, we found a clear, but mild, learning effect throughout the study (from 4.40 ± 0.3 minutes for the first examination, to 4.17 ± 0.4 for the sixth examination) (Fig 3). The older age group (>40 years) showed a slightly larger improvement between the first and the last session duration when compared with the younger age group (<40 years), an improvement of 0.34 minutes in the older group compared with 0.17 minutes in the younger group.

In order to verify whether the learning effect was of

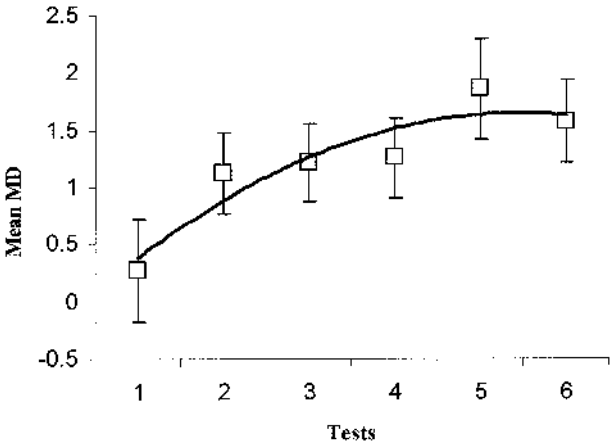


FIG. 1. Averaged MD ± SE plotted for each of the six sessions (second-degree polynomial fit). A significant difference was found only for the first pair (first vs second session, $P = 0.0032$).

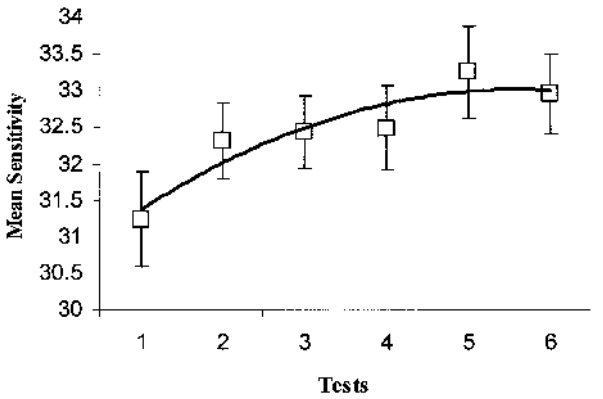


FIG. 2. Mean sensitivity ± SE plotted against session number. A significant difference was found only for the first pair (first vs second session, $P = 0.0031$).

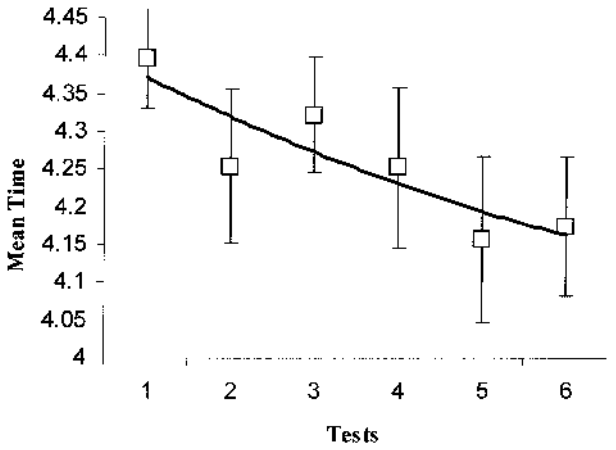


FIG. 3. Mean test duration ± SE plotted against session number. A trend for shorter test duration continued throughout the six sessions ($P = 0.023$).

clinical implication, we also analyzed the average number of depressed locations on the total deviation probability plot, for each subject, in each session. Depressed locations were counted as <5% probability level or worse, <2% probability level or worse, or <1% probability level or worse (Fig. 4). No locations reached the <0.5% probability level. The improvement seen between the first and second sessions almost reached statistical significance ($P = 0.054$) for the "<5% probability level or worse" counts.

Comparing the four paracentral VF grid locations (excluding the fovea) to the 12 peripheral VF grid locations, we found a significantly higher mean sensitivity in the paracentral VF ($P = 0.0048$). Both segments of the VFs

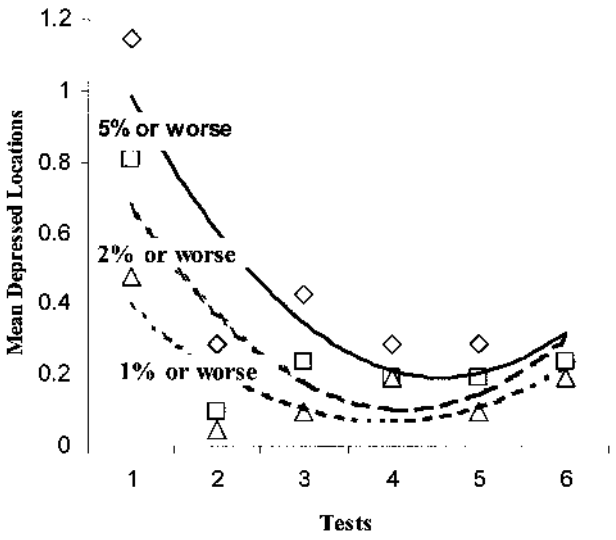


FIG. 4. Mean number of depressed locations on the total deviation plot in each session.

improved with each successive session and were highly correlated (Pearson product-moment correlation = 0.976). The superior hemifield showed a higher sensitivity than the inferior hemifield ($P = 0.042$), and the nasal hemifield showed a higher sensitivity than the temporal hemifield ($P = 0.0048$) (Fig. 5). For the first and second test, a statistically significant steeper learning effect was found in the peripheral and nasal visual segments compared with the paracentral and temporal segments, respectively ($P < 0.0001$) (Fig. 5).

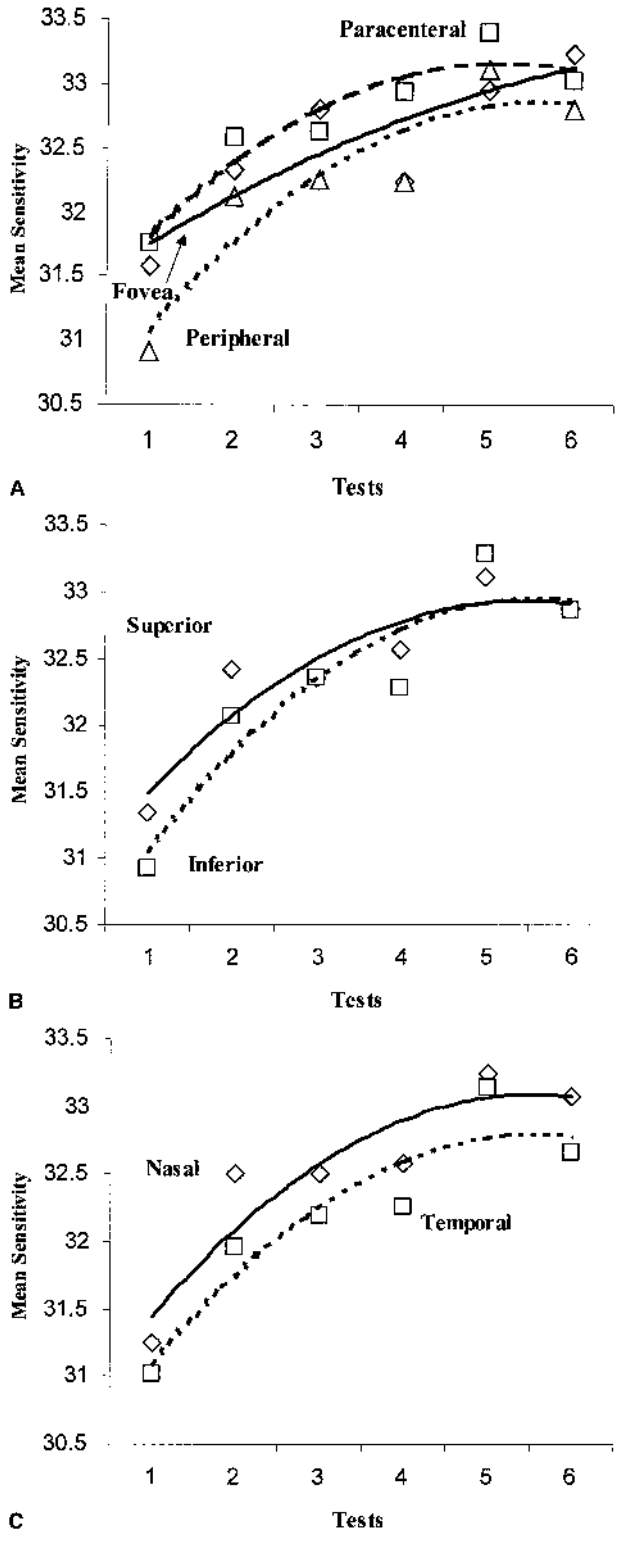
When examining the foveal VF grid location, a continuous learning effect was found, such that the mean sensitivity of the fovea continued to improve up to the sixth session (Fig. 5). There was a statistically significant difference between the foveal mean sensitivity in the younger age group compared with the older age group (33.16 ± 2.4 vs 30.94 ± 4.8 dB, respectively; $P = 0.0013$).

DISCUSSION

A prerequisite for any new VF technology aimed at screening for glaucoma is a short, predictable learning curve, preferably of small magnitude. Without this, it may be difficult to distinguish true-positive from false-positive results. Standard automated perimetry is known to harbor a learning effect that has been extensively studied.[14–16,23] FDT, a relatively new approach for quantifying the VF, is gaining popularity owing to its short test duration, patient acceptance, and the initial promising results for diagnosing glaucoma.[24,25]

Iester et al.[20] performed six FDT VFs on 20 healthy subjects in an attempt to establish the short- and long-term fluctuation, as well as the learning effect. Consequently, examination intervals were clustered such that examinations 2, 3, and 4 were performed on the same day. This clustering might have influenced the learning effect. In addition, the authors used as an inclusion criterion a normal standard automated VF. This might have biased the results in two ways. First, all patients, by definition, must have already performed two VF examinations. Although it may be claimed that standard automated perimetry does not influence the learning effect of FDT, the similarity in the VF tasks suggests otherwise. Second, provided that a normal standard automated VF is an entry criterion, one must conclude that any subject with an abnormal standard VF, due to a learning effect, would not have been included in the study.

Fujimoto et al.[21] presented a short report evaluating the learning effect in a group of glaucoma patients, and concluded that a learning effect was found between the first and second, but not between the second and third,



tests. Again, these patients had considerable experience (at least three examinations of each eye) with standard automated perimetry. A total of three FDT examinations were performed at 1- to 3-month intervals. The investigators concluded that a learning effect was found for glaucoma patients experienced with conventional automated perimetry.

Prior work in the learning-effect field suggests that a learning effect for VFs may be present during the first few examinations. The learning effect for standard (white-on-white) perimetry was estimated to be of negligible importance beyond the second examination.[14–16] Short-wavelength automated perimetry (SWAP) was shown to have a learning effect that probably exceeds that of standard automated perimetry.[26] In this study, we chose to collect data for six repetitions of the test to get a better perspective on the presence, magnitude, and duration of the learning effect.

Our study shows a significant improvement of the MD as well as the mean sensitivity of the FDT VFs over time. This improvement, though present throughout the study, ceased to be of statistical significance beyond the second session. The MD ± SD improved from 0.27 ± 2.1 dB in the first session to 1.58 ± 1.7 dB in the sixth session, a total of 1.3 dB. The mean sensitivity ± SD improved from 31.24 ± 3.0 dB in the first session to 32.96 ± 2.5 dB in the sixth session, a total of 1.7 dB. Of the total improvement seen throughout the study, the proportion of improvement seen from the first to the second session was 63% and 65% for mean sensitivity and MD, respectively. The lack of a trend in the PSD values suggests that the learning effect found is spread out across the VF, keeping the overall shape of the hill of vision rather similar.

Heijl et al.[27] and Wood et al.[13] demonstrated a more prominent learning effect in the peripheral VF when compared with more central locations for standard automated perimetry. Likewise, we found a significant difference in the FDT learning effect between the peripheral and paracentral VFs for the first two sessions (Fig. 5A).

In conclusion, the data presented strongly suggest a learning effect. This improvement, though modest in magnitude (MD improvement = 1.3 dB; mean sensitivity

FIG. 5. (A) Each session plotted against mean sensitivity, shown separately for foveal, paracentral, and peripheral grid locations. **(B)** Each session plotted against mean sensitivity, shown separately for the upper eight locations and the lower eight locations. The foveal location was not included in this analysis. The superior hemifield shows a higher sensitivity than the inferior hemifield. **(C)** Each session is plotted against mean sensitivity, graphed separately for the nasal and temporal hemifields. The nasal hemifield shows a higher sensitivity than the temporal hemifield.

improvement = 1.7 dB), was shown to be statistically significant when comparing the first session with subsequent sessions. Likewise, the number of depressed locations on the total deviation probability plot was shown to decrease in a statistically significant manner when comparing the first test to the second test. In respect to test duration, we found a clear but mild improvement trend throughout the study ($P = 0.023$).

The results of this study imply that baseline measurements should rely on the second testing session since MD and mean sensitivity are somewhat poorer when subjects with no prior VF experience are first tested on the FDT instrument. This may be especially true for the purpose of following-up patients over time. Therefore, it may be justified to recommend one practice FDT session prior to establishing a VF as a baseline against which future tests will be compared. A relatively short-lived learning effect is a reassuring finding for any subjective test. The extent of the learning effect found in our study does not appear longer than that found in other studies testing subjects experienced with standard VFs.[20,21]

REFERENCES

1. Vaughan D, Riordan-eva P. Glaucoma. In: Vaughan D, Asbury T, Riordan-eva P, eds. *General Ophthalmology,* 14th ed. Lange; 1999:208–225.
2. Quigley HA. Number of people with glaucoma worldwide. *Br J Ophthalmol* 1996;80:389–393.
3. Spry P, Johnson CA. Assessing visual function in clinical practice. *Ophthalmol Clin North Am* 2000;13:361–382.
4. Quigley HA, Dunkelberger GR, Green WR. Chronic human glaucoma causing selectively greater loss of large optic nerve fibers. *Ophthalmology* 1988;95:357–363.
5. Quigley HA, Dunkelberger GR, Green WR. Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *Am J Ophthalmol* 1989;107:453–464.
6. Quigley HA, Sanchez RM, Dunkelberger GR, et al. Chronic glaucoma selectively damages large optic nerve fibers. *Invest Ophthalmol Vis Sci* 1987;28:913–920.
7. Johnson CA, Samuels SJ. Screening for glaucomatous visual field loss with frequency-doubling perimetry. *Invest Ophthalmol Vis Sci* 1997;38:413–425.
8. Alward WL. Frequency doubling technology perimetry for the detection of glaucomatous visual field loss. *Am J Ophthalmol* 2000;129:376–378.
9. Kelly DH. Frequency doubling in visual responses. *J Opt Soc Am* 1966;56:1628–1633.
10. Kelly DH. Nonlinear visual responses to flickering sinusoidal gratings. *J Opt Soc Am* 1981;71:1051–1055.
11. Maddess T, Henry G. Performance of nonlinear visual units in ocular hypertension and glaucoma. *Clin Vis Sci* 1992;7:371–383.
12. Anderson AJ, Johnson CA. Mechanisms isolated by frequency-doubling technology perimetry. *Invest Ophthalmol Vis Sci* 2002; 43:398–401.
13. Wood JM, Wild JM, Hussey MK, et al. Serial examination of the normal visual field using Octopus automated projection perimetry. Evidence for a learning effect. *Acta Ophthalmol (Copenh)* 1987; 65:326–333.
14. Kulze JC, Stewart WC, Sutherland SE. Factors associated with a learning effect in glaucoma patients using automated perimetry. *Acta Ophthalmol (Copenh)* 1990;68:681–686.
15. Wild JM, Dengler-Harles M, Searle AE, et al. The influence of the learning effect on automated perimetry in patients with suspected glaucoma. *Acta Ophthalmol (Copenh)* 1989;67:537–545.
16. Werner EB, Krupin T, Adelson A, et al. Effect of patient experience on the results of automated perimetry in glaucoma suspect patients. *Ophthalmology* 1990;97:44–48.
17. Quigley HA. Identification of glaucoma-related visual field abnormality with the screening protocol of frequency doubling technology. *Am J Ophthalmol* 1998;125:819–829.
18. Casson R, James B, Rubinstein A, et al. Clinical comparison of frequency doubling technology perimetry and Humphrey perimetry. *Br J Ophthalmol* 2001;85:360–362.
19. Burnstein Y, Ellish NJ, Magbalon M, et al. Comparison of frequency doubling perimetry with Humphrey visual field analysis in a glaucoma practice. *Am J Ophthalmol* 2000;129:328–333.
20. Iester M, Capris P, Pandolfo A, et al. Learning effect, short-term fluctuation, and long-term fluctuation in frequency doubling technique. *Am J Ophthalmol* 2000;130:160–164.
21. Fujimoto N, Minowa K, Miyauchi O, et al. Learning effect for frequency doubling perimetry in patients with glaucoma. *Am J Ophthalmol* 2002;133:269–270.
22. Ito A, Kawabata H, Fujimoto N, et al. Effect of myopia on frequency-doubling perimetry. *Invest Ophthalmol Vis Sci* 2001;42: 1107–1110.
23. Heijl A, Lindgren G, Olsson J. The effect of perimetric experience in normal subjects. *Arch Ophthalmol* 1989;107:81–86.
24. Wadood AC, Azuara-Blanco A, Aspinall P, et al. Sensitivity and specificity of frequency-doubling technology, tendency-oriented perimetry, and Humphrey Swedish interactive threshold algorithm-fast perimetry in a glaucoma practice. *Am J Ophthalmol* 2002;133:327–332.
25. Sample PA, Bosworth CF, Blumenthal EZ, et al. Visual function-specific perimetry for indirect comparison of different ganglion cell populations in glaucoma. *Invest Ophthalmol Vis Sci* 2000;41: 1783–1790.
26. Wild JM, Moss ID. Baseline alterations in blue-on-yellow normal perimetric sensitivity. *Graefes Arch Clin Exp Ophthalmol* 1996; 234:141–149.
27. Heijl A, Bengtsson B. The effect of perimetric experience in patients with glaucoma. *Arch Ophthalmol* 1996;114:19–22.