



Published in final edited form as:

Invest Ophthalmol Vis Sci. 2009 March ; 50(3): 1234–1240. doi:10.1167/iovs.08-2535.

Visual Function-Specific Perimetry to Identify Glaucomatous Visual Loss Using Three Different Definitions of Visual Field Abnormality

Ali Tafreshi¹, Pamela A. Sample¹, Jeffrey M. Liebmann², Christopher A. Girkin³, Linda M. Zangwill¹, Robert N. Weinreb¹, Maziar Lalezary¹, and Lyne Racette¹

¹Hamilton Glaucoma Center and Department of Ophthalmology, University of California at San Diego, La Jolla, California

²New York Eye and Ear Infirmary, New York, New York

³Department of Ophthalmology, University of Alabama at Birmingham, Birmingham, Alabama

Abstract

Purpose—To compare the most recent versions of standard automated perimetry (SAP), short-wavelength automated perimetry (SWAP), and frequency-doubling technology (FDT) using three definitions of visual field (VF) abnormality: single-test abnormality, abnormality confirmed by the same test, and abnormality confirmed by a different test.

Methods—Data obtained from one eye of each of 174 patients with glaucoma and 164 age-matched healthy control subjects from the Diagnostic Innovations in Glaucoma Study and African Descent and Glaucoma Evaluation Study were included, based on the appearance of the optic disc on stereophotographs. Each participant had two reliable 24-2 SAP-SITA, SWAP-SITA, and Matrix FDT tests. Receiver operating characteristic (ROC) curves were generated for the PSD of each test to equate the tests at 90% and 95% specificity. SAP, SWAP, and FDT were compared under each definition of VF abnormality by assessing the sensitivities, the agreement between tests and the overlap in deficit location at these set specificities. The tests were also compared using the machine-derived PSD.

Results—At a set specificity of 95%, single-test sensitivities of 30% (SAP), 29% (SWAP), and 28% (FDT) were observed (all $P > 0.05$). Sensitivities ranged from 24% to 27% (all $P > 0.05$) when same-test confirmation was used and from 20% to 23% (all $P > 0.05$) when different-test confirmation was used. SAP/SAP sensitivity was higher than all different-test combinations (all $P < 0.05$), and SWAP/FDT sensitivity was lower than all same-test combinations (all $P < 0.05$).

Conclusions—Confirming VF abnormality is important and optimal when an abnormal SAP is confirmed by a subsequent SAP or SWAP test.

Glaucoma is a progressive optic neuropathy involving a degeneration of retinal ganglion cells and their axons and resulting in a distinct appearance of the optic disc associated with a

Copyright © Association for Research in Vision and Ophthalmology

Corresponding author: Lyne Racette, Hamilton Glaucoma Center, Department of Ophthalmology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0946; lracette@glaucoma.ucsd.edu.

Disclosure: **A. Tafreshi**, None; **P.A. Sample**, Carl Zeiss Meditec (F), Haag-Streit (F), Welch-Allyn (F); **J.M. Liebmann**, Carl Zeiss Meditec (F); **C.A. Girkin**, Carl Zeiss Meditec (F), Heidelberg Engineering (F), Optovue (F); **L.M. Zangwill**, Heidelberg Engineering (F), Carl Zeiss Meditec (F), Optovue (F), Allergan (F); **R.N. Weinreb**, Heidelberg Engineering (F), Carl Zeiss Meditec (C, F); **M. Lalezary**, None; **L. Racette**, None

characteristic pattern of visual field loss.¹ Standard automated perimetry (SAP) is the current clinical standard for visual field testing.² SAP is a relatively nonselective test in that all subtypes of retinal ganglion cells are sensitive to the white target presented on a white background.³ Function-specific tests have been developed to preferentially, though not exclusively, target subpopulations of retinal ganglion cells.⁴ Short-wavelength automated perimetry (SWAP) presents a blue light on a yellow background to emphasize the response characteristics of the blue–yellow pathway,^{5–8} whereas frequency-doubling technology (FDT) perimetry uses a rapidly reversing contrast grating to emphasize the response characteristics of the magnocellular pathway.^{9–11} Some function-specific tests have been shown to be more sensitive to glaucoma than SAP.^{3,12,13} This enhanced sensitivity was believed to be the result of the reduced redundancy within the visual system achieved by targeting subpopulations of retinal ganglion cells.¹⁴

There is no consensus on how visual function testing should be implemented in clinical practice, and there are considerable differences among studies on how this is done. As an example, the Early Manifest Glaucoma Trial (EMGT)¹⁵ and the Ocular Hypertension Treatment Study (OHTS)¹⁶ both required that participants undergo two baseline visual field tests. In the OHTS, 85.9% of initial visual field abnormalities were not confirmed on a subsequent visual field test, powerfully highlighting the importance of confirmation.¹⁶ Other studies have also shown the importance of confirming visual field defects.^{17,18} Only a few studies^{3,12} compared SAP, SWAP, and FDT to each other without requiring the confirmation of visual field abnormalities. This drawback is directly addressed in the present study. In addition, we assessed the impact of confirming a visual field defect using the same test versus using a different test of visual function. The goal of the present study was to compare the most recent versions of SAP, SWAP, and FDT tests using three different definitions of visual field abnormality: (1) single-test abnormality, (2) abnormality confirmed by the same test, and (3) abnormality confirmed by a different test.

METHODS

Participants

One eye of each of 174 patients with glaucoma and 164 age-matched healthy control subjects were included ($n = 338$). These participants were selected from the Diagnostic Innovations in Glaucoma Study (DIGS; $n = 58$) and from the African Descent and Glaucoma Evaluation Study (ADAGES; $n = 280$). The DIGS is being conducted at the Hamilton Glaucoma Center at the University of California at San Diego (UCSD), whereas the ADAGES is a multicenter study being conducted at UCSD, at the University of Alabama at Birmingham, and at the New York Eye and Ear Infirmary. These ongoing studies are prospectively designed to assess structure and function in glaucoma. Participants are selected based on the inclusion–exclusion criteria specified in the next section and are followed up annually. Patient participants were referred to the study through the Glaucoma Services at the Departments of Ophthalmology of each site. Healthy participants were recruited from the general population through advertisement, from private practices, and from the staff and employees at each of the study institutions. Informed consent was obtained from each participant, and all methods adhered to the Declaration of Helsinki for research involving human subjects and complied with the Health Insurance Portability and Accountability Act (HIPAA) standards. The ethics review board at each institution approved all methods. Each participant underwent a complete ophthalmic examination that included review of relevant medical history, best-corrected visual acuity, slit lamp biomicroscopy (including gonioscopy), applanation tonometry, dilated funduscopy, stereoscopic ophthalmoscopy of the optic disc with a 78-D lens, and stereoscopic fundus photography.

Inclusion Criteria for DIGS and ADAGES

Simultaneous stereoscopic photographs were obtained in all subjects and were of adequate quality for the subjects to be included. All subjects had open angles, best corrected acuity of 20/40 or better, spherical refraction within ± 5.0 D, and cylinder correction within ± 3.0 D. All participants had reliable visual field results on all three tests, defined as 33% or fewer false-positive results, false-negative results, and fixation losses. One eye was selected from each participant. Candidates with a family history of glaucoma were included.

Exclusion Criteria for DIGS and ADAGES

Normal and ocular hypertensive subjects were excluded if they had a history of intraocular surgery (except for uncomplicated cataract surgery). We also excluded all subjects with nonglaucomatous secondary causes of elevated intraocular pressure (IOP; e.g., iridocyclitis, trauma), coexisting retinal disease (e.g., diabetic retinopathy), other diseases affecting visual field (e.g., pituitary lesions, demyelinating diseases, HIV or AIDS, or diabetes), with medications known to affect visual field sensitivity, or problems affecting color vision other than glaucoma.

The DIGS and ADAGES participants included in this study had two SAP, SWAP, and FDT tests with reliable results performed within a 3-month period. They also had stereophotographs taken within 6 months of the visual field tests.

Stereoscopic Optic Disc Photographs

Subjective evaluation of structural damage to the optic nerve was based on clinical assessment of stereoscopic optic disc photographs. Simultaneous stereophotographs were obtained after maximum pupil dilation (TRC-SS camera; Topcon Instrument Corp of America, Paramus, NJ). Each photograph was graded by two experienced graders using a stereoscopic viewer (Asahi Pentax StereoViewer II; Asahi Optical Co, Tokyo, Japan) and a standard fluorescent light box. Each grader was masked to the subject's identity, study group classification, results from the other grader, and other test results. When the two graders disagreed, a third experienced grader adjudicated. The diagnosis of glaucomatous optic neuropathy (GON) was based on cup-to-disc asymmetry between the eyes of 0.2 or more, neuronal rim-thinning, notching, excavation, or nerve fiber layer thinning (focal or diffuse).

Study Groups

Participants were classified based on the presence of structural damage to the optic disc as assessed by simultaneous stereophotographs.¹⁹ Visual field results were not used to assign participants to the study groups.

Healthy Control Subjects—Participants ($n = 164$) included in this group had normal appearance of the optic disc on stereophotographs and fundus examination in both eyes. They also had no history of ocular hypertension defined as IOP ≥ 22 mm Hg, no history of use of glaucoma medication or glaucoma surgery in either eye. Healthy control subjects were age-matched to the glaucoma patient group.

Patients with Glaucoma—Participants ($n = 174$) included in this group had glaucomatous appearance of the optic disc on simultaneous stereophotographs.

Tests of Visual Function

Each participant was tested twice on SAP (SAP-1 and SAP-2), SWAP (SWAP-1 and SWAP-2), and FDT perimetry (FDT-1 and FDT-2) within a 3-month period. For different-test confirmation, SAP-1, SWAP-1, and FDT-1 tests were used (for 88% of participants, these three

tests were performed on the same day). For same-test confirmation, SAP-2, SWAP-2, and FDT-2 were used to confirm the results obtained on SAP-1, SWAP-1, and FDT-1 tests, respectively. We limited the time period between tests to ensure that no progression occurred between the initial and confirmatory tests. The confirmatory tests used in this study should therefore not be considered as longitudinal follow-up. The two locations above and below the blind spot were always excluded, leaving 52 points for analysis. All tests assessed the central 24° of the visual field and required fixation by the patient. Adequate refraction was provided for each device, and the pupils had a diameter of at least 3 mm. The pupils were dilated when this requirement was not met. The order of testing was counterbalanced across participants and remained the same for each visit of the same participant.

Standard Automated Perimetry (SAP-SITA)—SAP is a non-selective test, in that all types of retinal ganglion cells are able to detect the target. Each participant underwent SAP using the 24-2 program on the Humphrey Field Analyzer II, using the Swedish Interactive Thresholding Algorithm (SITA) version 4.1²⁰ (Carl Zeiss Meditec Inc, Dublin, CA). The target used in this achromatic test is a small (0.43°) flash of white light presented on a dim background (31.5 apostilbs) for 200 ms. Participants were asked to respond when a flash of light was detected.

Short-Wavelength Automated Perimetry—SWAP targets the short-wavelength-sensitive cones and pathway. At the ganglion cell level, the patient's response to this test is most likely mediated by the small bi-stratified blue-yellow ganglion cells that comprise approximately 9% of the total population of retinal ganglion cells. The test provides a dynamic range of approximately 35 dB and 15 dB of isolation before the next most sensitive mechanism can detect the target, most likely the middle-wavelength-sensitive pathway cells.²¹ SWAP uses a bluish (440-nm wavelength) narrow-band target of 1.8° presented for 200 ms on a bright (100 cd/m²) yellow background. Participants were tested with the 24-2 test pattern and the new SITA testing strategy (version 4.1).²²

Frequency-Doubling Technology—FDT (24-2) perimetry measures contrast sensitivity. The test is based on the frequency-doubling illusion, which was first described by Kelly²³ and later was proposed as a sensitive measure of glaucomatous visual field loss.^{11,24} This illusion occurs when a sinusoidal grating of low spatial frequency undergoes counterphase flickering at a high temporal frequency. It was originally believed that the FDT test isolates the spatially nonlinear M_y cells,^{24,25} a subset of the magnocellular retinal ganglion cells. However, more recent evidence suggests that the magnocellular pathway is isolated as a whole by the FDT stimulus and it appears increasingly likely that the FDT stimuli are detected through flicker sensitive mechanisms.^{26,27} FDT was measured with the Humphrey Matrix FDT Visual Field Instrument (Carl Zeiss Meditec), with the 24-2 test pattern and Welch-Allyn Frequency-Doubling Technology (Skaneateles Falls, NY) and the Zippy Estimation by Sequential Testing (ZEST) thresholding algorithm.²⁸ The details of the test have been described elsewhere.¹³

Data and Statistical Analyses

Equating Tests for Specificity—To target different subpopulations of retinal ganglion cells, each test included in this study uses different stimuli, backgrounds, thresholding algorithms and normative databases. By equating the tests for set specificity levels, these differences between the tests can be minimized, thus allowing for fair comparisons between the results. We generated receiver operating characteristic (ROC) curves and derived abnormality cutoffs at set specificity levels of 90% and 95%. ROC curves were generated for the following visual field parameters: mean deviation (MD), pattern standard deviation (PSD), and the number of total deviation (TD) and pattern deviation (PD) points triggered at 5% and at 1%. This procedure was used for each of the six visual field tests used in this study (initial

and confirmatory SAP, SWAP, and FDT tests). For each of the six tests, the cutoff associated with the desired specificity (90% and 95%) was applied to determine whether each result was normal or abnormal. The areas under the ROC curves were compared statistically with the method of DeLong et al.²⁹ using commercial software (Matlab; The MathWorks Inc., Natick, MA). After exploring the data, we opted to compare the tests based on the abnormality cutoffs obtained for the PSD for each test. Four reasons prompted this decision: (1) the ROC-derived PSD was the best parameter for two of the three tests (SAP and SWAP), (2) no significant differences were observed between the area under the ROC curves of the best parameter of each test (that which yielded the highest area under the ROC curve) and the ROC-derived PSD for any of the tests ($P > 0.05$), (3) PSD is a continuous variable, allowing for specificities to be equated more accurately, and (4) PSD performs better at distinguishing between normal and glaucoma subjects than MD, although MD may be better for determining progression.¹⁷ We are reporting the results based on 95% specificity level (similar results were obtained at 90% specificity), as high specificities are desirable for glaucoma. This allows a more direct comparison with the machine-derived PSDs which are based on the 95% specificity level derived from their respective internal normative databases. We compared the sensitivities across tests at set specificities using the McNemar test.

Definitions of Visual Field Abnormality—Visual field abnormality was defined in three different manners by applying the ROC-derived PSD cutoffs: (1) single-test abnormality, (2) abnormality confirmed by the same test, and (3) abnormality confirmed by a different test. The performance of the three tests was compared under each definition. Single-test abnormality was based on the results of the initial test and did not require confirmation. The ROC-derived PSD abnormality cutoffs were applied, and each test was labeled as either normal (PSD lower than the cutoff) or abnormal (PSD at or higher than the cutoff). The second definition of abnormality (confirmed with the same test) required that the results of an abnormal initial visual field test be confirmed by the results of a second test of the same type. For example, an initial SAP test result was considered abnormal only if the results of a confirmatory SAP test were also abnormal. The third definition of abnormality (confirmed with a different test) required that the results of an abnormal initial visual field test be confirmed by the results of a second test of a different type. For example, an abnormal SAP test (SAP-1) was considered abnormal if the results were confirmed by an abnormal FDT test (FDT-1). As mentioned previously, the confirmatory tests in this study were conducted within a short period and should not be considered as longitudinal follow-up testing. All test combinations were evaluated.

Agreement between Tests—The level of agreement between tests under each definition of visual field abnormality was assessed using the κ statistic.³⁰ The κ values range from 0.00 to 1.00, with values between 0.00 and 0.20 indicating slight agreement, 0.21 and 0.40 indicating fair agreement, 0.41 and 0.60 indicating moderate agreement, 0.61 and 0.80 indicating substantial agreement, and 0.81 and 1.00 indicating almost perfect agreement.

Overlap in Deficit Location—When two or all three visual field results were abnormal based on the ROC-derived PSD cutoffs, we determined which quadrants had a deficit. This was done for each definition of visual field abnormality. A defective quadrant was defined as having at least three PD points triggered at 5% or at least one PD point triggered at 1% (the points did not have to be clustered). When the same quadrant was defective on the initial and confirmatory visual field tests, this quadrant was considered to have an overlapping deficit. This method was used for each of the same-test and the different-test confirmation combinations. Our definition of overlap in deficit location required that overlap be present in one or more quadrants.

Comparison of Healthy Control Subjects with Those Included in the Normative Databases of Each Instrument—

We did not use visual field results to classify our participants into study groups. Some participants who showed visual field loss before structural deficits were therefore included in the control group. This affected the ROC-derived abnormality cutoffs. To determine the extent of the effect, we compared the results obtained with the ROC-derived cutoffs to those obtained with the machine-derived cutoffs. A PSD triggered at 5% or worse by the system was considered abnormal. We applied this abnormality criterion to each of the three visual field abnormality definitions described earlier.

RESULTS

Table 1 provides descriptive results for the participants in the healthy control and patient groups. The mean, SD, and range for MD and PSD are provided for the healthy control subjects and patients with glaucoma for descriptive purposes only, as SAP was not used to classify participants into study groups.

Equating Tests for Specificity

The areas under the ROC curves obtained for each parameter of each initial test are presented in Table 2. The areas under the ROC curves for PSD were 0.692, 0.693, and 0.685 for SAP, SWAP, and FDT, respectively. Table 2 also presents the abnormality criteria associated with 90% and 95% specificity levels for each parameter for each initial test. In our dataset, the ROC-derived PSD cutoffs at 95% specificity approximate the machine-derived PSD at 99.5%.

Definitions of Visual Field Abnormality

The left panel in the Venn diagrams presented in Figure 1 shows the number of patients with GON with abnormal visual field results when single-test abnormality was used and the PSD criteria associated with 95% specificity were applied. Of the 174 patients included in this study, 52 (30%) had an abnormal initial SAP test result, 50 (29%) had an abnormal initial SWAP test result, and 48 (28%) had an abnormal initial FDT test result. The McNemar test showed no significant differences between any of these single-test sensitivities (SAP versus SWAP: $P = 0.67$; SAP versus FDT: $P = 0.39$; SWAP versus FDT: $P = 0.71$). The sensitivities obtained using a single test were always higher than those obtained using either same- or different-test confirmation combinations ($P < 0.05$).

When visual field abnormalities were confirmed with the same test, the number of patients with abnormal visual field test results decreased to 47 (27%) for SAP, 42 (24%) for SWAP, and 44 (25%) for FDT (Fig. 1; middle). The intersections of the Venn diagrams presented in the right panel of Figure 1 show the number of patients with visual field abnormalities confirmed with different test pairings. Of the 174 patients, 40 (23%) patients had abnormal results on both SAP and SWAP, 39 (22%) on SAP and FDT, and 35 (20%) on SWAP and FDT. We compared the sensitivities obtained with each of the same- and different-test combinations by using the McNemar test and we report the associated probabilities in Table 3. No statistically significant differences were observed between the sensitivities obtained using any of the same-test confirmation combinations (SAP/SAP, SWAP/SWAP, and FDT/FDT). Similarly, no differences were observed between any of the different-test combinations (SAP/SWAP, SAP/FDT, and SWAP/FDT). The sensitivity obtained using the SAP/SAP same-test confirmation combination was the only one significantly higher than those obtained with the different-test combinations. The sensitivity obtained with the SWAP/FDT different-test confirmation combination was lower than those obtained with each of the same-test combinations.

The Venn diagrams do not show whether the same patients were identified by each of the test combinations. For example, the middle panel of Figure 1 does not indicate whether the patients with confirmed abnormal SWAP results ($n = 42$) also have confirmed abnormal SAP results ($n = 47$). These groups could be mutually inclusive or mutually exclusive. Table 4 shows that most of these participants ($n = 37$) were included in both groups. Because each of the function-specific tests targets different subsets of retinal ganglion cells, confirming with a different test may have identified glaucoma in different patients. A large proportion of patients who were identified by each of the test combinations were also identified by other test combinations (72%–100%; Table 4). However, a small subset of patients was identified by each test combination uniquely (12%–28%).

The percentage of abnormal initial visual field test results that were not confirmed by a subsequent visual field test was also derived from the data included in Figure 1. When visual field abnormalities were confirmed using the same test, 10% (5/52) of SAP tests, 16% (8/50) of SWAP tests, and 8% (4/48) of FDT results were not confirmed. Of the 52 abnormal initial SAP results, 13 were not confirmed with FDT (25%) and 12 were not confirmed with SWAP (23%). Of the 50 abnormal initial SWAP results, 10 were not confirmed with SAP (20%) and 15 were not confirmed with FDT (30%). Of the 48 abnormal initial FDT tests, 9 were not confirmed with SAP (19%), and 13 were not confirmed with SWAP (27%).

Agreement between Tests—The agreement observed between the results of the visual field tests is presented in Table 5. At 95% specificity levels, substantial agreement between test results was obtained when the same test was performed twice. When the results of one test were compared with the results of a different test, moderate to substantial agreement was achieved.

Overlap in Deficit Location—In participants with two or more abnormal field types, visual field defects overlapped in at least one quadrant in all participants in all test combinations, both when visual field abnormalities were confirmed with the same test and when they were confirmed with a different test.

Comparing Healthy Control Subjects to Those Included in the Normative Databases of Each Instrument—Using the machine-derived PSD, we calculated the sensitivity and specificity of each visual function test under the three definitions of visual field abnormality. The results are shown in Table 6. Overall, higher specificity levels were obtained when the FDT test was used. Higher specificity and lower sensitivity levels were obtained when the results of each visual function test were confirmed by the same or by a different test compared with when the results were unconfirmed. Figure 2 shows that confirmation results are proportionally similar to those obtained with the ROC-derived abnormality cutoffs shown in Figure 1. We compared the sensitivities obtained with each of the same- and different-test combinations by using the McNemar test. The results show that the sensitivity obtained using the SAP/SAP combination was significantly higher than FDT/FDT ($P = 0.04$), SAP/FDT ($P = 0.02$), and SWAP/FDT ($P = 0.01$). The sensitivity obtained with the SAP/SWAP combination was significantly higher than that obtained with SAP/FDT ($P = 0.03$) and SWAP/FDT ($P = 0.002$). No significant difference was found between the sensitivities obtained with SAP/SAP and SAP/SWAP ($P = 0.83$).

DISCUSSION

Using older versions of the tests (SWAP full-threshold and FDT N-30), our group has previously shown that at equal specificity levels, patients could show deficits on any perimetric procedure, whereas remaining normal on the others. FDT had the highest sensitivity overall.³ However, because this study did not require confirmation of visual field abnormalities, the

results may have been affected by false-positive test results. In the present study, we compared the most recent versions of SAP, SWAP, and FDT by using three definitions of visual field abnormality: single-test abnormality, abnormality confirmed by the same test, and abnormality confirmed by a different test. When confirmation was not required (single-test abnormality), we obtained results comparable to those in the earlier study,³ in that no one test was better than all others in identifying glaucomatous visual field defects.

Consistent with the finding reported by the Ocular Hypertension Treatment Study (OHTS),¹⁶ our results indicate that not all initially abnormal visual field test results are confirmed on a subsequent test. In our study, 10% of initial abnormal SAP visual field results were not confirmed by a subsequent SAP test, compared with the 85.9% observed in OHTS.¹⁶ This smaller percentage is probably due to the different populations included in each study. Our population included patients with GON evident on stereophotographs, whereas the OHTS included participants without detectable glaucomatous optic disc damage. This smaller percentage could also be attributed to differences in visual field test-taking experience between these two populations. All patients in our study had prior experience with visual field testing, most likely resulting in more consistent test results. In our study, a higher percentage of initially abnormal visual field test results were confirmed when the same test (between 84% and 92%) was used, compared with when a different test (between 70% and 81%) was used to confirm. This finding is consistent with the agreement results that show a trend toward better agreement for same-test confirmation compared with different-test confirmation (Table 5).

The results of the present study are consistent with the recommendation that confirmation is necessary to avoid overcalling visual field abnormalities. With ROC-derived cutoffs, significantly higher sensitivities were always obtained with single-test abnormality compared with abnormalities confirmed with the same or a different test. When confirmation was used, similar sensitivities were achieved for all same test combinations. Similarly, none of the sensitivities obtained using the different test confirmation combinations were significantly different from each other. However, when same-test combinations were compared with different-test combinations, only the SAP/SAP same-test combination yielded significantly higher sensitivities than did the different-test combinations. It should be noted, however, that each of the test combinations identified a small but unique subset of patients (12%–28%).

The lack of a truly independent gold standard for glaucoma is a limitation common to all glaucoma studies. To minimize classification bias, we did not use visual fields to classify our participants into study groups, and instead required that patients show evidence of glaucomatous optic neuropathy on stereophotographs. If the healthy control subjects included in this study were similar to those included in the normative databases of each instrument, we would expect 95% specificity levels when defining abnormalities based on the machine-derived PSD triggered at 5%. However, specificity levels obtained with the machine-derived PSD were considerably lower as shown in Table 6, indicating that our control subjects were different from those of the normative databases. Although the specificity levels were lower and the sensitivity levels were higher when abnormality was defined based on machine-derived PSD, the results were proportionately similar and the necessity to confirm visual field abnormalities remained evident. Significantly higher sensitivities were obtained for single-test abnormalities compared with those obtained when same-and different-test confirmation was used. The optimal combination to confirm visual field abnormalities using the machine-derived analysis was SAP/SAP or SAP/SWAP. Using a different test provides additional information on the nature of the losses in visual function. Although using a surrogate gold standard (GON on stereophotographs) affects the overall sensitivity and specificity levels, the necessity of confirming visual field abnormalities remains.

Previous studies have reported better diagnostic performance (area under the ROC curve and/or sensitivity) for FDT compared with SAP and SWAP. In this study however, the area under the ROC curve for FDT was similar to that of the other tests. A possible explanation is that the parameter yielding the highest area under the ROC curve was used for SAP and SWAP (PSD), but not for FDT (number of TD points triggered at 5%). Although the area under the ROC curves was not significantly different between the best parameter (number of TD points triggered at 5%) and the PSD for FDT, it is possible that this resulted in underestimating the FDT performance. We therefore reanalyzed the data using the best parameter for each test. The results were similar to those reported in this study, with comparable sensitivities for all definitions of visual field abnormality. The previously reported advantage of FDT (N-30 and 24-2) over SAP and SWAP was small,^{3,13} and it may not be consistently repeatable in different populations.

In conclusion, confirming visual field abnormalities increases confidence in the functional status of patients. Our results verify previous findings, showing that visual field deficits, when present, occur in the same location across tests. Clinically, this provides further evidence that these defects represent real disease. Confirming visual field deficits is imperative to avoid overcalling abnormalities. Overall, an abnormal SAP result confirmed with either a subsequent abnormal SAP or SWAP result offers an optimal combination of sensitivity and specificity.

Acknowledgments

The authors thank Madhusudhanan Balasubramanian for performing the statistical comparisons between the areas under the ROC curves.

Supported by Grants U10 EY14267 and EY 08208 (PAS), EY 11008 (LMZ), and EY13959 (CAG) from the National Eye Institute and by the Eyesight Foundation of Alabama (CAG). Participant retention incentive grants in the form of glaucoma medication at no cost: Alcon Laboratories, Inc.; Allergan; Pfizer, Inc.; Santen, Inc.; and Merck.

References

1. Weinreb RN, Khaw PT. Primary open-angle glaucoma. *Lancet* 2004;363:1711–1720. [PubMed: 15158634]
2. San Laureano J. When is glaucoma really glaucoma? *Clin Exp Optom* 2007;90:376–3. 85. [PubMed: 17697185]
3. Sample PA, Medeiros FA, Racette L, et al. Identifying glaucomatous vision loss with visual-function-specific perimetry in the diagnostic innovations in glaucoma study. *Invest Ophthalmol Vis Sci* 2006;47:3381–339. [PubMed: 16877406]
4. Anderson RS. The psychophysics of glaucoma:improving the structure/function relationship. *Prog Retin Eye Res* 2006;25:79–97. [PubMed: 16081311]
5. Sample PA, Weinreb RN. Color perimetry for assessment of primary open-angle glaucoma. *Invest Ophthalmol Vis Sci* 1990;31:1869–1875. [PubMed: 2211033]
6. Racette L, Sample PA. Short-wavelength automated perimetry. *Ophthalmol Clin North Am* 2003;16:227–236. vi–vii. [PubMed: 12809160]
7. Johnson CA, Brandt JD, Khong AM, Adams AJ. Short-wavelength automated perimetry in low-, medium-, and high-risk ocular hypertensive eyes:initial baseline results. *Arch Ophthalmol* 1995;113:70–76. [PubMed: 7826296]
8. Johnson CA, Adams AJ, Casson EJ, Brandt JD. Blue-on-yellow perimetry can predict the development of glaucomatous visual field loss. *Arch Ophthalmol* 1993;111:645–650. [PubMed: 8489447]
9. Quigley HA. Identification of glaucoma-related visual field abnormality with the screening protocol of frequency doubling technology. *Am J Ophthalmol* 1998;125:819–829. [PubMed: 9645719]
10. Maddess T, Goldberg I, Dobinson J, Wine S, Welsh AH, James AC. Testing for glaucoma with the spatial frequency doubling illusion. *Vision Res* 1999;39:4258–4273. [PubMed: 10755162]

11. Johnson CA, Samuels SJ. Screening for glaucomatous visual field loss with frequency-doubling perimetry. *Invest Ophthalmol Vis Sci* 1997;38:413–425. [PubMed: 9040475]
12. Sample PA, Bosworth CF, Blumenthal EZ, Girkin C, Weinreb RN. Visual function-specific perimetry for indirect comparison of different ganglion cell populations in glaucoma. *Invest Ophthalmol Vis Sci* 2000;41:1783–1790. [PubMed: 10845599]
13. Racette L, Medeiros FA, Zangwill LM, Ng D, Weinreb RN, Sample PA. Diagnostic accuracy of the Matrix 24-2 and original N-30 frequency-doubling technology tests compared with standard automated perimetry. *Invest Ophthalmol Vis Sci* 2008;49:954–960. [PubMed: 18326718]
14. Johnson CA. Selective versus nonselective losses in glaucoma. *J Glaucoma* 1994;3(suppl):S32–S44. [PubMed: 19920586]
15. Leske MC, Heijl A, Hyman L, Bengtsson B. Early Manifest Glaucoma Trial: design and baseline data. *Ophthalmology* 1999;106:2144–2153. [PubMed: 10571351]
16. Keltner JL, Johnson CA, Quigg JM, Cello KE, Kass MA, Gordon MO. Confirmation of visual field abnormalities in the Ocular Hypertension Treatment Study. Ocular Hypertension Treatment Study Group. *Arch Ophthalmol* 2000;118:1187–1194. [PubMed: 10980763]
17. Artes PH, Chauhan BC. Longitudinal changes in the visual field and optic disc in glaucoma. *Prog Retin Eye Res* 2005;24:333–354. [PubMed: 15708832]
18. Gardiner SK, Anderson DR, Fingeret M, McSoley JJ, Johnson CA. Evaluation of decision rules for frequency-doubling technology screening tests. *Optom Vis Sci* 2006;83:432–437. [PubMed: 16840859]
19. Johnson CA, Sample PA, Zangwill LM, et al. Structure and function evaluation (SAFE):II. Comparison of optic disk and visual field characteristics. *Am J Ophthalmol* 2003;135:148–154. [PubMed: 12566017]
20. Bengtsson B, Olsson J, Heijl A, Rootzen H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand* 1997;75:368–375. [PubMed: 9374242]
21. Sample PA, Johnson CA, Haegerstrom-Portnoy G, Adams AJ. Optimum parameters for short-wavelength automated perimetry. *J Glaucoma* 1996;5:375–383. [PubMed: 8946293]
22. Bengtsson B. A new rapid threshold algorithm for short-wavelength automated perimetry. *Invest Ophthalmol Vis Sci* 2003;44:1388–94. [PubMed: 12601072]
23. Kelly DH. Frequency doubling in visual responses. *J Opt Soc Am* 1966;56:1628–1633.
24. Maddess T. Performance of nonlinear visual units in ocular hypertension and glaucoma. *Clin Vision Sci* 1992;7:371–383.
25. Maddess T, Hemmi JM, James AC. Evidence for spatial aliasing effects in the Y-like cells of the magnocellular visual pathway. *Vision Res* 1998;38:1843–1859. [PubMed: 9797962]
26. Anderson AJ, Johnson CA. Mechanisms isolated by frequency-doubling technology perimetry. *Invest Ophthalmol Vis Sci* 2002;43:398–401. [PubMed: 11818383]
27. White AJ, Sun H, Swanson WH, Lee BB. An examination of physiological mechanisms underlying the frequency-doubling illusion. *Invest Ophthalmol Vis Sci* 2002;43:3590–3599. [PubMed: 12407172]
28. Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Performance of efficient test procedures for frequency-doubling technology perimetry in normal and glaucomatous eyes. *Invest Ophthalmol Vis Sci* 2002;43:709–715. [PubMed: 11867588]
29. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845. [PubMed: 3203132]
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174. [PubMed: 843571]

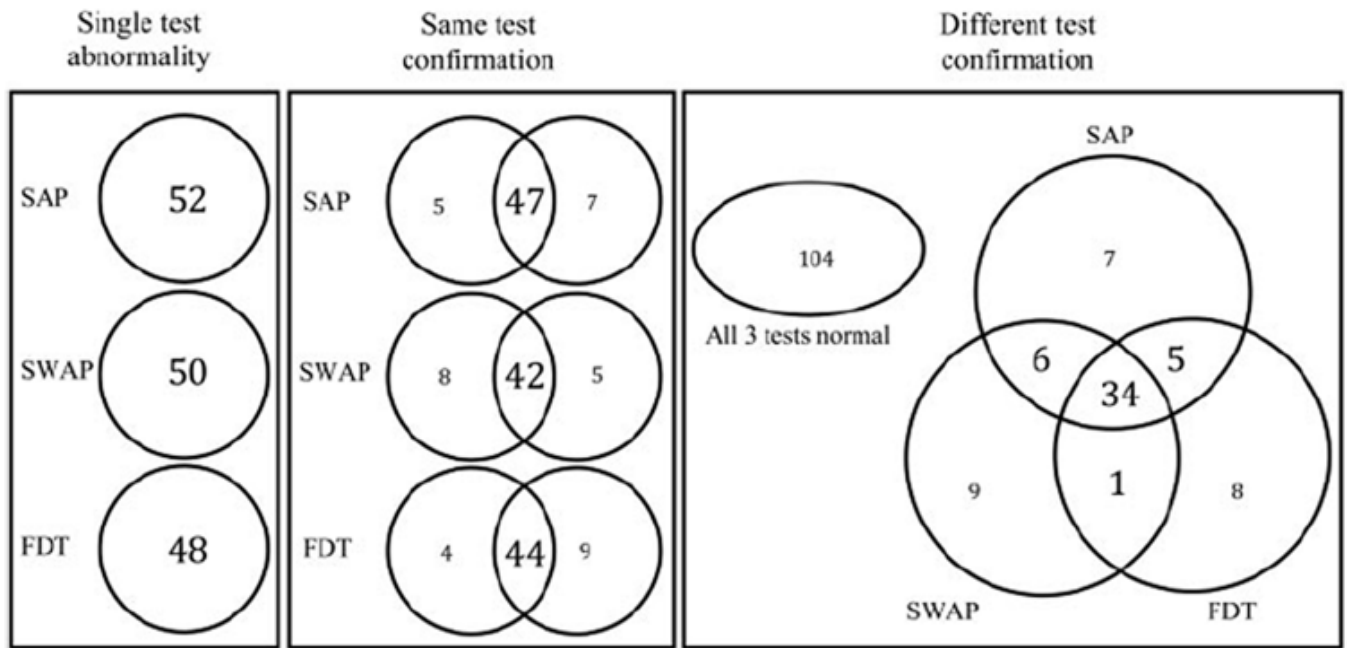


Figure 1.

Venn diagrams show the number of patients with GON ($n = 174$) with abnormal visual field results for single-test abnormalities (*left*), same-test confirmation (*middle*), and different-test confirmation (*right*). The numbers presented for SAP, SWAP, and FDT are based on the results obtained at 95% specificity using the ROC-derived PSD abnormality cutoffs.

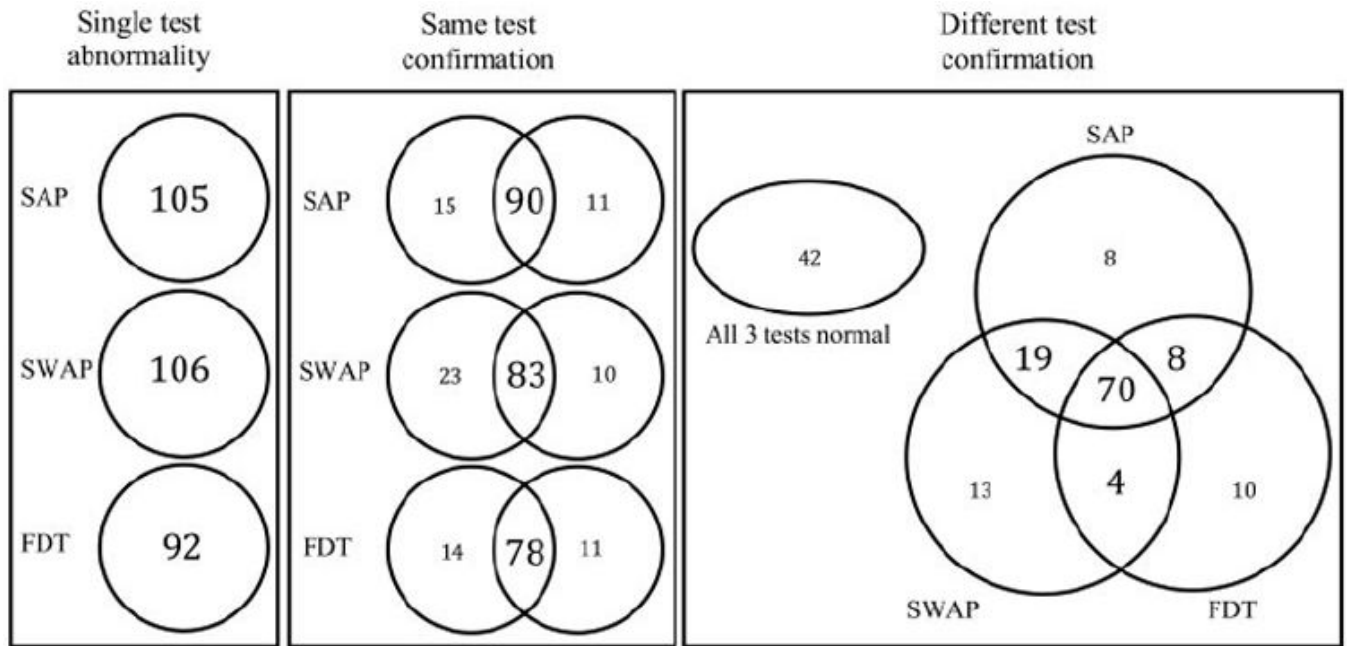


Figure 2. Venn diagrams show the number of patients with GON ($n = 174$) with abnormal visual field results for single-test abnormalities (*left*), same-test confirmation (*middle*), and different-test confirmation (*right*). The numbers presented for SAP, SWAP, and FDT are based on the machine-derived PSD triggered at 5%.

Table 1

Descriptive Measures for Each Study Group

	Controls (<i>n</i> = 164)	Glaucoma (<i>n</i> = 174)	<i>P</i>
Age (y) *	59.6 ± 13.5	56.9 ± 11.3	0.05
Eye (% right eye) †	84.2	71.8	0.01
Sex (% male) †	36.0	46.6	0.05
SAP MD (dB) *	-1.09 ± 0.39	-4.40 ± 0.38	<0.0001
SAP MD range (dB)	-8.07 to 2.26	-30.16 to 2.10	
SAP PSD (dB) *	2.16 ± 0.23	4.43 ± 0.22	<0.0001
SAP PSD range (dB)	0.95 to 11.92	1.13 to 17.0	

Data are the mean ± SD, except as noted.

* By *t*-test.

† By χ^2 test.

Table 2

The Area under the ROC Curves (AUC) and the Associated SE for Each Parameter for Each Initial Test

	AUC	SE	Sens/Spec (%)	Criterion for 90% Spec	Sens/Spec (%)	Criterion for 95% Spec
SAP-SITA						
MD	0.674	0.029	37/90	-3.45	22/95	-6.05
PSD	0.692	0.028	37/90	3.71	30/95	4.85
TD < 5%	0.662	0.029	30/90	38	16/95	38
TD < 1%	0.659	0.029	33/90	12	18/95	24
PD < 5%	0.675	0.029	26/90	20	22/95	22
PD < 1%	0.672	0.029	34/91	9	26/95	13
SWAP-SITA						
MD	0.662	0.029	35/90	-9.34	28/95	-11.06
PSD	0.693	0.028	37/90	4.56	29/95	5.6
TD < 5%	0.648	0.030	28/90	42	17/95	49
TD < 1%	0.660	0.029	31/90	24	22/95	32
PD < 5%	0.670	0.029	28/90	19	24/95	21
PD < 1%	0.675	0.029	30/90	10	22/95	14
FDT 24-2						
MD	0.714	0.028	39/90	-6.32	32/95	-7.53
PSD	0.685	0.029	34/90	4.26	28/95	5.02
TD < 5%	0.716	0.028	36/90	24	26/95	31
TD < 1%	0.691	0.028	43/90	9	31/95	15
PD < 5%	0.702	0.028	37/90	12	28/95	16
PD < 1%	0.673	0.028	33/91	7	30/95	8

The sensitivities obtained for the initial test at approximately 90% and 95% specificities for each parameter of each test are also presented. The criteria presented in this table are those associated with the initial test.

Table 3

Probabilities Associated with the Comparisons of Sensitivities at a Set Specificity of 95% (ROC-Derived) across All Combinations of Tests

	SAP/SAP (27%)	SWAP/SWAP (24%)	FDT/FDT (25%)	SAP/SWAP (23%)	SAP/FDT (22%)
SWAP/SWAP (24%)	0.20				
FDT/FDT (25%)	0.47	0.64			
SAP/SWAP (23%)	0.01	0.48	0.32		
SAP/FDT (22%)	0.01	0.44	0.06	0.76	
SWAP/FDT (20%)	0.001	0.02	0.003	0.06	0.10

Significant results are in bold. The sensitivity for each test combination is shown in parentheses in the column headings.

Table 4

The Number of Patients with Glaucoma with Confirmed Visual Field Abnormalities Who Were Identified by More Than One Combination of Tests

	SAP/SAP (n = 47)	SWAP/SWAP (n = 42)	FDT/FDT (n = 44)	SAP/SWAP (n = 40)	SAP/FDT (n = 39)
SWAP/SWAP (n = 42)	37				
FDT/FDT (n = 44)	37	34			
SAP/SWAP (n = 40)	40	37	34		
SAP/FDT (n = 39)	38	33	38	34	
SWAP/FDT (n = 35)	34	34	35	34	34

Table 5
The Agreement Observed between Each of the Visual Field Tests at 95% Specificity

	SAP		SWAP		FDT	
	κ (Strength of Agreement)	Proportion of Agreement	κ (Strength of Agreement)	Proportion of Agreement	κ (Strength of Agreement)	Proportion of Agreement
SAP	0.799 (S)	0.94				
SWAP	0.671 (S)	0.90	0.746 (S)	0.92		
FDT	0.646 (S)	0.89	0.556 (M)	0.87	0.741 (S)	0.92

M, moderate agreement; S, substantial agreement.

Table 6

The Sensitivity/Specificity According to the Machine-Derived PSD for Each Definition of Visual Field Abnormality

	SAP (%)	SWAP (%)	FDT (%)
Single-test	60/60	61/63	53/76
SAP	52/74	51/76	45/85
SWAP		48/78	43/84
FDT			45/87