



Published in final edited form as:

Ophthalmology. 2008 December ; 115(12): 2127–2131. doi:10.1016/j.ophtha.2008.08.013.

Assessment of Patient Opinions of Different Clinical Tests Used in the Management of Glaucoma

Stuart K. Gardiner, PhD and Shaban Demirel, PhD

Devers Eye Institute, Discoveries In Sight Laboratories, Legacy Health System, Portland, Oregon

Abstract

Purpose—To compare different tests used in the clinical management of glaucoma, with respect to the testing experience for patients undergoing each test.

Design—Evaluation of diagnostic tests.

Participants—A total of 101 subjects with high-risk ocular hypertension or early glaucoma.

Methods—Subjects were asked to give their opinion on 7 tests used clinically in glaucoma management by assigning each a score between 0 (absolute dislike) and 10 (perfect satisfaction).

Main Outcome Measure—Tests were ranked for each subject from 1 (favorite test) to 7 (least favorite test) on the basis of patient-assigned scores.

Results—Goldmann applanation tonometry for measurement of intraocular pressure was ranked significantly better than any other test (median rank 2.5, P 0.01). This was followed by confocal scanning laser ophthalmoscopy using a Heidelberg Retina Tomograph (median rank 3.3); frequency doubling technology perimetry (4.0); multifocal visual evoked potential (4.0); optic nerve photography (4.3); and standard automated perimetry (4.8). Short-wavelength automated perimetry was ranked significantly worse than any other test (median rank 5.3, P 0.04).

Conclusions—In many cases, statistically significant differences were found between the patients' opinions of the tests. Information on this issue has to date largely been anecdotal or subjective. To our knowledge, this is the most comprehensive study to assess and compare the patient experience when undergoing these tests.

Financial Disclosure(s)—The authors have no proprietary or commercial interest in any materials discussed in this article.

When a decision is being made as to which clinical test to use in any given situation, the most important consideration is the clinical information that will be gained. Therefore, much of the clinical research literature is devoted to examining the relevance, reliability, variability, and predictive power of the various available tests for detection or prognostic determination. Other factors considered could include the cost of performing the test (to the clinic or the patient) and the length of time necessary to complete the test, both of which are straightforward to assess.

However, there is another factor that needs to be considered, namely, the testing experience for the patient. Tests that place a larger burden on the patient, caused by time, frustration, or

any other factors, may result in a reduced willingness to return for follow-up. In glaucoma management, this is particularly important, because testing over a prolonged period of time is necessary to detect progression and thus predict long-term prognosis.¹ Reduced subject retention would also adversely affect longitudinal clinical studies. For functional tests of the visual field, which require that the subject actively respond to a stimulus, lower motivation could reduce the reliability of the results, because it could adversely affect attention and concentration.

In the ophthalmic research literature, the issue of the patient's opinion of the testing experience has largely gone unreported, possibly because it is harder to quantify in a manner that allows direct comparisons to be made. Alternatively, this may be due to the subjective nature of the question; opinions will vary markedly between individuals and in some cases between visits for the same individual. However, the issue deserves to be explored in a scientific manner to avoid decisions being made on the basis of only anecdotal evidence.

Surveys have been used for other purposes in the ophthalmic literature. Quality-of-life studies typically use patient questionnaires as their primary data source.²⁻⁵ However, patient opinions of the methods currently used in a clinic to test for visual deficits have rarely been quantified. Factors affecting compliance with follow-up visits in glaucoma have been surveyed, but the type of test used was not a major factor considered in those studies.⁶⁻⁸ Surveys comparing different tests have largely been confined to the clinicians' point of view.^{9,10} Bjerre et al¹¹ found that most patients in their study preferred multifocal visual evoked potential (mfVEP) over Swedish Interactive Thresholding Algorithm (SITA) standard perimetry, but other tests were not evaluated in that study.

In this study, we aim to quantify the overall patient experience for a number of tests used in the clinical management of glaucoma. To the best of our knowledge, this study provides the most comprehensive objective evaluation to date of the patient's perceived burden when exposed to these tests. We believe that this study provides a useful tool to address the issue of patient preferences between tests in an objective manner and that this is important information to be considered in clinical practice.

Materials and Methods

Subjects were taken from an ongoing longitudinal study of glaucomatous progression at Discoveries in Sight Laboratories, Devers Eye Institute, Portland, Oregon. Subjects with early glaucoma and suspected glaucoma were recruited and have been followed annually for up to 10 years. Detailed inclusion and exclusion criteria, including clinical characteristics, are tangential to this study but have been described elsewhere.¹²⁻¹⁴ The study adheres to the tenets of the Declaration of Helsinki and complies with the Health Insurance Portability and Accountability Act of 1996, and the Legacy Health System Institutional Review Board approved the protocol. All subjects signed an informed consent form before participation in the study after all foreseeable risks and benefits had been explained to them.

As part of their annual testing, participants were asked to give their opinion on 7 tests performed within the study. Subjects were given a survey on which they were asked to indicate their satisfaction with each test by circling an integer from 0 (absolute dislike) to 10 (perfect satisfaction). Discrete scoring was used rather than a continuous visual analog scale for easier understanding by the subjects. Informally, subjects' reasons for their opinions were also sought, although by their nature these data are purely anecdotal (a more thorough questionnaire of this qualitative information was not undertaken to avoid prolonging the data collection and biasing the primary quantitative results by asking "leading" questions). In most cases, the survey was administered before performing the first test for the current study

year, based on memory from previous years, to eliminate risk of bias caused by the order in which the tests had been carried out on that visit. Subjects who indicated that they could not remember sufficiently well, and the 3 subjects in their first year in the study, completed the survey after carrying out the tests for the current study year. In total, 101 subjects completed the survey. Ages ranged from 30 to 87 years (mean 63.7, standard deviation 10.8), with 59 female and 42 male subjects. Figure 1 shows histograms of the number of years in the study and the mean deviation (MD) from standard automated perimetry (SAP) for those subjects when the survey was undertaken (mean -0.51 dB).

The 7 tests included were as follows:

1. SAP: white-on-white perimetry on a 10 cd/m^2 background, carried out on a Humphrey Field Analyzer II¹⁵ (Carl Zeiss Meditec Inc., Dublin, CA), using the SITA standard algorithm.¹⁶ Test duration is approximately 5 minutes per eye.
2. Short wavelength automated perimetry (SWAP): short wavelength stimulus presented on a 100 cd/m^2 yellow background, also carried out on a Humphrey Field Analyzer II, using the SITA SWAP algorithm.^{17,18} Test duration is approximately 4 minutes per eye.
3. Frequency doubling perimetry (FDT): a counterphase flickering sinusoidal stimulus on a 100 cd/m^2 background, carried out on the Humphrey Matrix perimeter.^{19,20} Test duration is approximately 5 minutes per eye.
4. mfVEP: mfVEPs were obtained using one of the standard stimulus options available within the VERIS software package (Dart Board 60 with Pattern, VERIS 4; Electro-Diagnostic Imaging, San Mateo, CA).²¹ Test duration is approximately 16 minutes per eye divided into 30-second intervals.
5. Heidelberg retinal tomograph (HRT): Confocal scanning laser ophthalmoscopy images of the optic nerve head were obtained using the Heidelberg Retina Tomograph (versions 2.01/3.04 HRT; Heidelberg Engineering, Heidelberg, Germany).²¹ Test duration is approximately 10 minutes per eye, including setup and preparation time.
6. Optic nerve photography: Optic disc photographs were obtained in all patients by using a simultaneous stereoscopic camera (3-Dx; Nidek Co., Ltd., Gamagori, Japan) after maximum pupil dilation.²¹ Up to 12 photographs are taken per eye depending on image quality. Test duration is approximately 5 minutes per eye after maximal pupil dilation has been achieved.
7. Intraocular pressure (IOP): Goldmann Applanation Tonometry at a slit-lamp. Test duration is less than 1 minute per eye after instillation of anesthetic drops.

Scores for each subject were converted into ranks to remove the effect of some subjects being generally more positive than others. The 7 tests were ranked from 1 (highest score) to 7 (lowest score). Tests receiving the same score were assigned the same rank based on averaging; for example, if 2 tests both received a raw score of 10, they would each be assigned a rank of 1.5. The raw scores and the ranks given to each test by all subjects were then summarized. To compare the scores between 2 tests, the nonparametric Wilcoxon signed-rank test was used.

The first 3 tests (SAP, SWAP, and FDT) require the subject to respond when a stimulus is detected. Therefore, it could be hypothesized that subjects with more advanced damage would have a different opinion of the test, because they are likely to detect a lower proportion of the stimuli or be subjected to longer test durations in the case of SAP and SWAP. To test this, scores given by each subject for these 3 functional tests were correlated

against their MD for that test. Spearman's rho statistic was used because it is more robust to non-Gaussian data distributions than the standard Pearson correlation coefficient.

Results

Figure 2 shows histograms of the raw scores for each test. Table 1 summarizes the raw scores for each of the 7 tests and the by-subject ranks for the 7 tests. IOP was the highest-rated test, followed by HRT and FDT. SWAP received the lowest average score, with 48 subjects ranking SWAP as their least favorite or joint least favorite test. Table 2 shows the 7 tests in order of mean rank, along with *P* values from pairwise comparisons of ranks.

Discussion

The patient's opinion of a particular test is not the only factor when choosing which tests to perform. The expected clinical utility of the results will generally be the primary consideration. However, the patient's experience is an important factor to consider. For some patients, it is possible that it may affect their willingness to return for follow-up visits and the frequency of visits they will tolerate. In a disease such as glaucoma, which generally progresses slowly, this may be critical to clinical care.

For some patients, a lack of motivation may conceivably adversely affect the quality of the results, because many functional tests of the visual field (including SAP, SWAP, and FDT) require the patient to concentrate throughout and respond to stimuli. In such cases, it would be important to explain and repeatedly reaffirm the importance of the test so that motivation will be maintained. Steps may also be taken to improve the testing experience, for example, allowing breaks partway through a test or altering the order in which tests are taken. Information about the relative patient experiences with different tests can also be useful for improving current tests and devising new ones. This renders the results of this survey useful beyond the lifespan of the 7 tests used here. Determining the reasons underlying preferences relies on the use of anecdotal evidence, and some reasons may remain undetected. However, some of this anecdotal evidence is sufficiently common for conclusions to be drawn.

IOP received the most favorable opinions of the 7 tests in the study, followed by HRT. Both are quick to perform, and no demands are placed on the subject beyond keeping the eye open and maintaining steady eye and head position. HRT also produces an image of the retina and optic nerve head that can be viewed by the subject, providing interest and increasing motivation. The other structural test was stereophotos (optic nerve photography), which was rated less favorably than IOP or HRT, partly because some subjects found the flash illumination to be uncomfortably bright. A more significant reason may be that optic nerve photography requires the subject's pupils to be dilated, requiring placing uncomfortable drops into the eye, a delay while the eye drops take effect, and a subsequent period of altered vision until the eye drops wear off.

Of the functional tests of the visual field, SAP, SWAP, and FDT all required the subject to respond to a stimulus by pressing a button. It may have been expected that this would adversely affect subjects' opinions of those tests, because concentration and attention are required throughout the test. Furthermore, approximately half of the stimuli are not detected during the thresholding algorithm, possibly causing frustration. For this reason, mfVEP has previously been reported to be "easier and less stressful" than SAP.¹¹ However, FDT was rated highly (receiving scores not significantly different from those for HRT). Many of the subjects have participated in the study for several years, and so it may be the case that they are keener and better motivated than the average patient in a clinical setting.

The correlations between the MD of the worst eye and the survey score for that same test were -0.010 for SAP, -0.045 for SWAP, and 0.033 for FDT. None of these correlations were significant. It is possible that significant correlations would appear in a group of patients with more advanced damage; although the MDs of some subjects were as low as -14 dB (Fig 1), the majority of subjects had at most very early visual field damage. Therefore, although evidence was not found to support the hypothesis that a worse MD would negatively affect the subject's opinion of these tests, the hypothesis cannot be refuted on the basis of this cohort.

Anecdotally, many subjects report that SWAP is more fatiguing than SAP, despite the test durations being similar (because both use the newer SITA thresholding algorithms).^{16,17} It may be beneficial for some patients to be allowed to take breaks partway through SWAP testing, even with the shorter test durations of approximately 5 minutes per eye achieved with the SITA SWAP algorithm (although care would be needed to ensure that adaptation to the yellow background was maintained). The background may contribute to the fatigue; SWAP uses a brighter background than SAP (100 cd/m² instead of 10 cd/m²), and yellowing of the aging lens further increases glare from the yellow background used in SWAP compared with the white background used in SAP. Many subjects also report increased mental fatigue with SWAP and to a lesser extent in SAP when compared with FDT because of frustration caused by uncertainty about whether they saw a stimulus or not. Frequency-of-seeing curves are flatter (greater short-term variability) with SWAP than with SAP^{22,23} and steeper (lower variability, more certainty) with FDT.²⁴ This may be a major factor in determining the relative patient experience of these 3 tests. For some time it has been assumed on the basis of anecdotal data that subjects prefer SAP or FDT to SWAP, and this has been one of the factors preventing a more widespread adoption of SWAP testing in clinics, but to our knowledge this is the first study to provide quantitative support for that assumption.

As with any clinically oriented study, there are caveats attached to these findings. The study participants are experienced with each test and in many cases have been in the study for 10 years. Therefore, the results may not be perfectly representative of a typical clinical population. Subjects who have remained in the study for multiple years are probably better motivated than the average patient, causing the mean raw score given (7.25) to be higher than would be expected in subjects who were not somewhat self-selected. Any subjects with an intense dislike of one or more of the tests would likely have dropped out of the study, and so these individuals will not be represented. A repeat of this study with subjects who had not previously undergone any of the tests would be of interest. However, this also means that the subjects are less liable to bias caused by one unusual experience, and their opinions have had time to crystallize, so day-to-day variability in the results from one subject should be reduced. In addition, the results should be less affected by novelty and unfamiliarity with any of the tests.

The patient experience when undergoing clinical tests needs to be taken into consideration. Knowledge that a particular test is less pleasant to perform warrants a more detailed explanation to the patient of the benefits of that particular test, both for retaining patients for follow-up visits and for improving the reliability of test results. In some cases, it may affect the decision of which test(s) to use for a particular patient or when designing a research study. The results presented in this study provide valuable information for objectively comparing different tests in terms of the satisfaction levels of subjects.

Acknowledgments

Supported in part by the National Institutes of Health (Bethesda, Maryland) grant EY03424. The funding organization had no role in the design or conduct of this research.

References

1. Chauhan BC, Garway-Heath DF, Goni FJ, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol*. 2008; 92:569–73. [PubMed: 18211935]
2. Parrish RK II, Gedde SJ, Scott IU, et al. Visual function and quality of life among patients with glaucoma. *Arch Ophthalmol*. 1997; 115:1447–55. [PubMed: 9366678]
3. Nelson P, Aspinall P, O'Brien C. Patients' perception of visual impairment in glaucoma: a pilot study. *Br J Ophthalmol*. 1999; 83:546–52. [PubMed: 10216052]
4. Janz N, Wren P, Lichter P, et al. CIGTS Study Group. The Collaborative Initial Glaucoma Treatment Study: interim quality of life findings after initial medical or surgical treatment of glaucoma. *Ophthalmology*. 2001; 108:1954–65. [PubMed: 11713062]
5. Freeman E, Munoz B, West S, et al. Glaucoma and quality of life: the Salisbury Eye Evaluation. *Ophthalmology*. 2008; 115:233–8. [PubMed: 17655930]
6. Mansberger SL, Edmunds B, Johnson CA, et al. Community visual field screening: prevalence of follow-up and factors associated with follow-up of participants with abnormal frequency doubling perimetry technology results. *Ophthalmic Epidemiol*. 2007; 14:134–40. [PubMed: 17613848]
7. Kosoko O, Quigley HA, Vitale S, et al. Risk factors for noncompliance with glaucoma follow-up visits in a residents' eye clinic. *Ophthalmology*. 1998; 105:2105–11. [PubMed: 9818613]
8. Ngan R, Lam DL, Mudumbai RC, Chen PP. Risk factors for noncompliance with follow-up among normal-tension glaucoma suspects. *Am J Ophthalmol*. 2007; 144:310–1. [PubMed: 17659968]
9. Strong NP. How optometrists screen for glaucoma: a survey. *Ophthalmic Physiol Opt*. 1992; 12:3–7. [PubMed: 1584613]
10. Tuck MW, Crick RP. Use of visual field tests in glaucoma detection by optometrists in England and Wales. *Ophthalmic Physiol Opt*. 1994; 14:227–31. [PubMed: 7970736]
11. Bjerre A, Grigg JR, Parry NR, Henson DB. Test-retest variability of multifocal visual evoked potential and SITA standard perimetry in glaucoma. *Invest Ophthalmol Vis Sci*. 2004; 45:4035–40. [PubMed: 15505053]
12. Gardiner SK, Johnson CA, Spry PG. Normal age-related sensitivity loss for a variety of visual functions throughout the visual field. *Optom Vis Sci*. 2006; 83:438–43. [PubMed: 16840869]
13. Spry PG, Johnson CA, Mansberger SL, Cioffi GA. Psychophysical investigation of ganglion cell loss in early glaucoma. *J Glaucoma*. 2005; 14:11–9. [PubMed: 15650598]
14. Gardiner SK, Johnson CA, Cioffi GA. Evaluation of the structure-function relationship in glaucoma. *Invest Ophthalmol Vis Sci*. 2005; 46:3712–7. [PubMed: 16186353]
15. Anderson, DR.; Patella, VM. *Automated Static Perimetry*. 2nd ed.. Mosby; St. Louis, MO: 1999. p. 147-59.
16. Bengtsson B, Olsson J, Heijl A, Rootzen H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand*. 1997; 75:368–75. [PubMed: 9374242]
17. Bengtsson B. A new rapid threshold algorithm for short-wavelength automated perimetry. *Invest Ophthalmol Vis Sci*. 2003; 44:1388–94. [PubMed: 12601072]
18. Bengtsson B, Heijl A. Normal intersubject threshold variability and normal limits of the SITA SWAP and full threshold SWAP perimetric programs. *Invest Ophthalmol Vis Sci*. 2003; 44:5029–34. [PubMed: 14578431]
19. Anderson AJ, Johnson CA, Fingeret M, et al. Characteristics of the normative database for the Humphrey matrix perimeter. *Invest Ophthalmol Vis Sci*. 2005; 46:1540–8. [PubMed: 15790927]
20. Artes PH, Hutchison DM, Nicolela MT, et al. Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma. *Invest Ophthalmol Vis Sci*. 2005; 46:2451–7. [PubMed: 15980235]

21. Fortune B, Demirel S, Zhang X, et al. Comparing multifocal VEP and standard automated perimetry in high-risk ocular hypertension and early glaucoma. *Invest Ophthalmol Vis Sci.* 2007; 48:1173–80. [PubMed: 17325161]
22. Kwon YH, Park HJ, Jap A, et al. Test-retest variability of blue-on-yellow perimetry is greater than white-on-white perimetry in normal subjects. *Am J Ophthalmol.* 1998; 126:29–36. [PubMed: 9683146]
23. Gilmore ED, Hudson C, Nrusimhadevara RK, Harvey PT. Frequency of seeing characteristics of the short wavelength sensitive visual pathway in clinically normal subjects and diabetic patients with focal sensitivity loss. *Br J Ophthalmol.* 2005; 89:1462–7. [PubMed: 16234454]
24. Spry PG, Johnson CA, McKendrick AM, Turpin A. Variability components of standard automated perimetry and frequency-doubling technology perimetry. *Invest Ophthalmol Vis Sci.* 2001; 42:1404–10. [PubMed: 11328758]

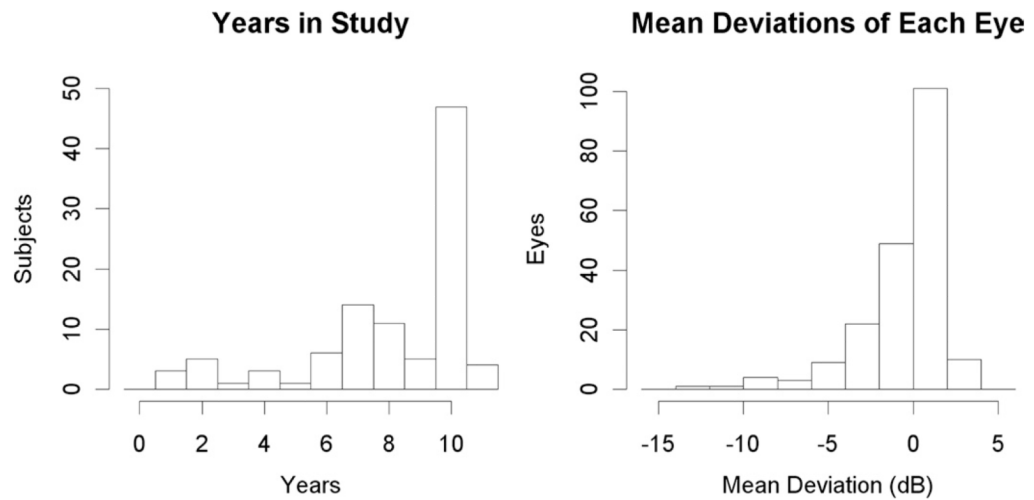


Figure 1. Population characteristics for the study. The first histogram shows the number of years each subject has been in the study and receiving annual testing. The second histogram shows the mean deviations of each eye, according to SAP, measured in decibels. Note that both eyes are included in the second histogram, so there are twice as many data points.

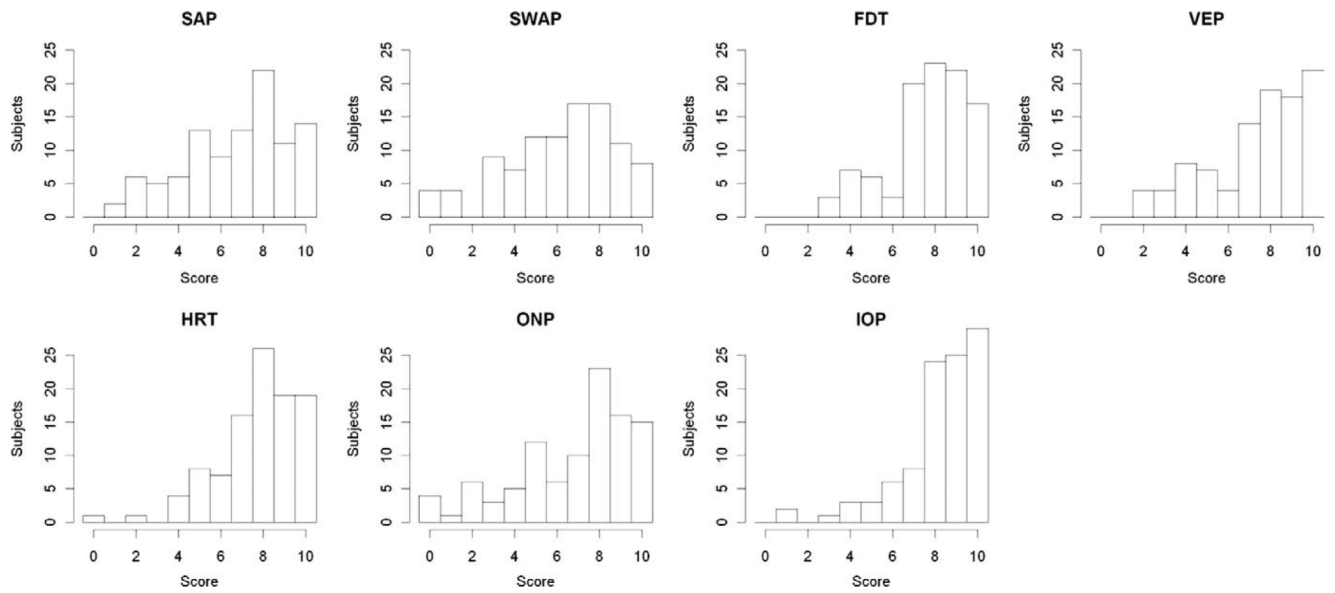


Figure 2.

Histograms of the survey scores for each test, from 0 (absolute dislike) to 10 (perfect satisfaction), for the 101 subjects in the study. The tests included were SAP, SWAP, FDT, VEPs, HRT, ONP, and IOP. SAP = standard automated perimetry; SWAP = short wavelength automated perimetry; FDT = frequency doubling perimetry; VEP = visual evoked potential; HRT = Heidelberg retina tomograph; ONP = optic nerve photography; IOP = intraocular pressure.

Table 1

Survey Results

Test	Raw Scores			Ranks		
	Mean	Median	SD	Mean	Median	SD
SAP	6.7	7.0	2.5	4.6	4.8	1.8
SWAP	6.2	7.0	2.6	5.1	5.3	1.6
FDT	7.7	8.0	1.9	3.6	4.0	1.6
mVEP	7.4	8.0	2.3	3.8	4.0	1.8
HRT	7.7	8.0	1.9	3.5	3.3	1.5
ONP	6.8	8.0	2.7	4.3	4.3	1.8
IOP	8.2	9.0	1.9	3.0	2.5	1.6

SD = standard deviation; SAP = short wavelength automated perimetry; SWAP = frequency doubling perimetry; mVEP = multifocal visual evoked potential; HRT = Heidelberg retina tomograph; ONP = optic nerve photography; IOP = intraocular pressure.

Survey results, summarized according to raw scores (0 = absolute dislike, 10 = complete satisfaction) and rankings (1 = favorite test, 7 = least favorite test). For each test, the mean, median, and standard deviation are given.

Table 2

Statistical Comparisons Among the Seven Tests

Test	IOP	HRT	FDT	mVEP	ONP	SAP	SWAP
IOP	—	0.011	0.007	0.001	<0.001	<0.001	<0.001
HRT		—	0.856	0.214	<0.001	<0.001	<0.001
FDT			—	0.400	0.003	<0.001	<0.001
mVEP				—	0.027	0.027	<0.001
ONP					—	0.635	0.042
SAP						—	0.016
SWAP							—

IOP = intraocular pressure; HRT = Heidelberg retina tomograph; FDT = frequency doubling perimetry; mVEP = multifocal visual evoked potential; ONP = optic nerve photography; SAP = standard automated perimetry; SWAP = short wavelength automated perimetry.

Comparison among the 7 tests. The tests are ordered according to mean per-subject rank as shown in Table 1. The *P* values represent the results of Wilcoxon rank-sum tests to determine whether the scores were significantly different between the 2 tests.