Accuracy of Scanning Laser Polarimetry in the Diagnosis of Glaucoma

John R. Trible, MD; Richard O. Schultz, MD; James C. Robinson, MD; Terri L. Rothe

Objective: To determine the diagnostic accuracy of scanning laser polarimetry.

Subjects and Methods: A total of 95 healthy subjects and 102 patients with glaucoma met all inclusion criteria. Data collected on each participant included an automated visual field examination, stereoview optic nerve head photographs, intraocular pressure measurement, and a screening and full scanning laser polarimetry study. Each participant was classified as “normal,” “glaucoma,” or “uncertain” by each of 3 ophthalmologists based on all available clinical information, with the exception of the scanning laser polarimetry results. Before data analysis, 4 diagnostic algorithms for the full-test mode and 2 for the screening mode were chosen to be evaluated for their sensitivity and specificity in detecting glaucoma.

Results: Of the 4 algorithms tested for the full-test mode, “the number” (abnormal test score, ≥35) had sensitivities of 57%, 71%, and 81% for early, moderate, and severe glaucoma, respectively. Specificity was 89%. For the screening test, sensitivities were much lower, particularly for those with severe glaucoma damage.

Conclusions and Clinical Relevance: Scanning laser polarimetry can help to differentiate subjects with normal findings from patients with glaucomatous damage. Even the best algorithm tested, however, failed to detect a substantial number of subjects with severe damage. Further study is needed before scanning laser polarimetry can be recommended as a screening method for glaucoma.
SUBJECTS AND METHODS

The study protocol was approved by the Medical College of Wisconsin, Milwaukee, Institutional Review Committee. Subjects with normal findings or glaucoma on ocular examination were recruited from various sources, including our comprehensive ophthalmology service and the glaucoma referral service, or were volunteers from the Medical College of Wisconsin campus and surrounding communities. For inclusion, an age of 45 years or older with visual acuity of 20/60 or better was required with no evidence of ocular disease other than glaucoma, mild drusen, or cataract. Persons with previous intraocular surgery other than glaucoma surgery or uncomplicated cataract surgery were excluded. A total of 225 subjects who fulfilled all of the inclusion criteria were prospectively enrolled in this study.

After giving informed consent, subjects underwent an evaluation that included an abbreviated medical history, full ocular history, Snellen visual acuity assessment, autorefration if visual acuity was below 20/30, tonometry (Goldmann applanation tonometry or a handheld tonometric measure [Tonopen; Mentor Ophthalmics Inc, Santa Barbara, Calif]), and dilation with stereoview optic nerve head photographs. Tonopen measurements were typically performed by the technician (T.L.R.) when subjects were tested on days when they were not seeing their ophthalmologist. A visual field 24-2 program with a full-threshold strategy (Humphrey Systems, San Leandro, Calif) was required to be both reliable and performed within 12 months of the scanning laser polarimetry examination. In rare instances, a previous 30-2 program was allowed when the patient was unable to return for the standard 24-2 test.

Three clinicians (J.C.R., R.O.S., and J.R.T.) then evaluated all available information. Optic nerve assessment was performed by a careful analysis of the stereoview optic nerve head photographs. The results for each eye were then categorized separately by each clinician as “normal,” “glaucoma,” or “uncertain.” In addition to the information acquired from the standard evaluation, most subjects were patients of the Department of Ophthalmology, Medical College of Wisconsin, and, thus, had additional information available to maximize the ability to correctly assess their clinical status. The result for an eye was categorized as normal or glaucoma, for data analysis purposes, if all 3 clinicians categorized the result for an eye as normal or glaucoma. Subjects who had other permutation results categorized as uncertain were excluded from further analysis.

Although actual classification was based on all available information, each clinician followed some general guidelines. Glaucoma, as defined in this study, required either direct evidence of damage to the optic nerve or a strong suggestion of damage with associated glaucomatous visual field changes. In this study, an elevated intraocular pressure alone was insufficient for a classification of glaucoma. Furthermore, an elevated intraocular pressure was not required to classify an eye as glaucomatous, which is consistent with our current understanding of this optic neuropathy. If the optic nerve showed clear glaucomatous changes, then the eye was classified as glaucomatous. If the optic nerve had possible damage that was associated with glaucomatous visual field changes, such as the presence of 1 or more clusters of significantly depressed points on the pattern deviation plot or glaucoma hemifield test results that were “outside normal limits,” then the eye would typically be classified as glaucomatous. Conversely, if the optic nerve had suggestive features but the result of visual field testing was normal, then the eye was typically classified as “uncertain.” If an optic nerve appeared to be normal, the eye was not classified as glaucomatous regardless of visual field findings or intraocular pressure. In such a subject, if an eye had abnormal findings on visual field testing, then classification would typically be uncertain, whereas if normal findings on visual field testing were associated with a normal optic nerve, the eye would be considered normal.

Scanning laser polarimetry was performed on eyes that were undilated. The technique and image acquisition has been previously described.10 In brief, a polarized 780-nm diode-laser illuminating light source is directed at the retina and the birefringent nerve fiber layer. Changes in the polarization state of the reflected laser light are quantitated by a detection unit. An anterior segment compensating device is used to neutralize the polarization effects of the cornea and lens. The retardation of polarized light is measured at 65 536 pixels in 0.7 seconds. The screening test was done first, and then the full test was done. For the full test, an ellipse is placed around the optic nerve. A measurement ellipse is then generated for the peripapillary retina. In the screening mode, the placement of an ellipse is not necessary, fewer measurements are taken, results are not saved, and a more basic results printout is generated, so the test time is shorter. The test time for the combined screening and full test was typically 4 to 6 minutes for both eyes. The full-test mode allows combining images to form a mean image. Three well-centered images with good resolution were combined to form a mean composite. If 1 of the 3 images caused a clear degradation in quality of the mean composite, only 2 images were used. The image was then analyzed with standard software (GDX, Laser Diagnostic Technologies, San Diego, Calif). Study participants also typically had frequency-doubling perimetry performed to evaluate glaucoma screening devices.

Eyes diagnosed as having glaucoma were graded according to the level of associated visual field damage. Damage was graded as “early” if the visual field failed to meet “moderate” damage criteria. Damage was graded as “moderate” if the probability value associated with the Humphrey global indices, corrected-pattern SD, was P<.01, and mean deviation was less than –13 dB. Damage was graded as “severe” if the mean deviation was –13 dB or worse.

Continued on next page
The classification of subjects resulted in 102 patients with glaucoma, 95 normal subjects, and 28 subjects whose eyes were graded as uncertain. Characteristics of the group are shown in Table 1. One eye from each subject was included in the analysis of sensitivity and specificity for early, moderate, or severe glaucoma. If both eyes of a subject with glaucoma met the inclusion criteria but had different levels of glaucomatous damage (23 of 102 subjects), each eye was included in the respective subgroup analysis. For example, a subject with glaucoma might have 1 eye involved in the analysis of sensitivity and specificity for early glaucoma, and the other eye might be included in the analysis of moderate glaucoma. Because the analysis for early glaucoma includes only subjects with eyes with early glaucoma and normal eyes and because only 1 eye is contributed to the analysis from a given person, we maintain independence of observations. A total of 51 eyes with early glaucoma, 42 eyes with moderate glaucoma, and 95 eyes classified as normal were included in the analysis. When both eyes of a person met the inclusion criteria, preference was given to eyes graded as normal or showing glaucoma, rather than those graded as uncertain. If both eyes were eligible for entry, then the eye with the best-corrected visual acuity was chosen. The determination of eyes included in the analysis was done before analysis and without knowledge of the results of scanning laser polarimetry.

The scanning laser polarimetry software calculates a large number of values for assessment. Before analysis, 4 algorithms for the full-test mode and 2 algorithms for the screening test mode were chosen to evaluate the diagnostic accuracy of this procedure.

**Full-Test Mode**

**Number of Abnormalities.** Thirteen parameters graded as “within normal limits,” “borderline,” or “outside normal limits,” based on a comparison with an internal normative database, are provided by the scanning laser polarimetry software (Table 2).9 We allotted 1 unit for a grade of outside normal limits, 0.5 units for a grade of borderline, and 0 units for a grade of within normal limits. The total number of units for 13 parameters was then determined and entered as the “number of abnormalities.”

**Logistic Regression.** Seven parameters were chosen that were most likely to differentiate between subjects with and without glaucoma. These were selected based on published reports or abstracts1,11,12 evaluating the diagnostic accuracy of scanning laser polarimetry. These included symmetry, superior ratio, inferior ratio, superior-nasal ratio, maximum modulation, ellipse modulation, and average thickness. Logistic regression was performed to determine the best combination of parameters to explain the variance of the dependent variable “classification status” (normal or glaucoma). By including the grade of borderline, as well as within normal limits and outside normal limits, for each parameter, the regression equation was strengthened, as demonstrated by larger F values and greater statistical significance.

**Number.** Ranging between 0 and 100, a value of each eye is calculated that is derived from a neural network analysis of multiple parameters. The polarimetry manual9 states that in early evaluations, eyes that score 0 to 30 are normal, but those scoring above 70 tend to be glaucomatous.

**TSNIT Image Abnormality.** The TSNIT (temporal, superior, nasal, inferior, temporal quadrants) image abnormality is defined as nerve fiber layer thickness that drops below the lower limit of the 95% confidence interval for normal eyes in either the superior or inferior quadrants. The 95% confidence interval for normal eyes is calculated based on the internal normative database of the polarimetry software.9 The deviation below the lower limit is easily appreciated from the TSNIT image (also called nerve fiber layer analysis graph). Sensitivity and specificity are calculated by simply determining the percentage of the subjects of each group with this abnormality.

**Screening Test Mode**

**Number of Abnormalities.** The number of abnormalities is identical to that of the full-test mode, with the exception that the screening mode provides 7 parameters: symmetry, superior ratio, inferior ratio, maximum modulation, superior maximum, inferior maximum, and superior integral.

**Logistic Regression.** Logistic regression is identical to that of the full-test mode, with the exception that, to our knowledge, there are no published data suggesting the most promising parameters for the screening study. For this reason, we have included all 7 parameters in the logistic regression.

Analysis was based on the statistical comparison of each parameter with the polarimetry internal normative database, rather than the raw measurement. The normal subjects in this study were not part of that database. The area under the receiver operating characteristic (ROC) curve was calculated for each algorithm. Commercial software (Statistica 3.1; StatSoft, Inc, Tulsa, Okla; and Stata 4.0; Stata, College Station, Tex) was used for all statistical calculations, including descriptive statistics, logistic regression, ROC curves, and 2-sample t tests.
Glaucomatous damage is shown in the ROC curves for each algorithm in early, moderate, and severe glaucoma, respectively. The TSNIT image was not present when the measured nerve fiber layer thickness dropped below the range of normal in the superior or inferior region.

The area under the ROC curve for the "Number of Abnormalities" algorithm was 45%, 50%, and 59% for early, moderate, and severe glaucoma, respectively.

Specificity measured 77%, and sensitivity was 66% for early, 74% for moderate, and 81% for severe glaucoma, respectively. The number was not available in the screening mode (Table 4).

A TSNIT image abnormality was determined to be present when the measured nerve fiber layer thickness was 1.0 abnormalities and "borderline" = 0.5 abnormalities. The gray shaded lines indicate the cutoff for sensitivities associated with specificity of approximately 90%.

We wished to assess the possibility that the misclassification of glaucoma status was related to confounding factors such as vision, astigmatism, or pseudophakia. Misclassification was defined as an incorrect classification of a patient's eye by the number algorithm (cutoff, 35) compared with our clinical assessment (gold standard) described earlier. We then evaluated whether misclassification errors were more likely with decreased vision, greater astigmatism, or pseudophakia. For each parameter, frequencies were compared to determine if there was a statistically significant difference between correctly classified (CC, n = 157) and incorrectly classified (IC, n = 40) groups. There were no significant differences found for any of the variables: vision (quintile of Snellen acuity), CC = 0.83 ± 0.20, IC = 0.83 ± 0.17;
astigmatic error (in diopters), CC = 0.86 ± 0.87, IC = 0.83 ± 0.85; pseudophakia, CC = 0.18, IC = 0.18. This suggests that misclassification errors by laser scanning polarimetry were not related to these variables.

This study was done to measure the diagnostic accuracy of scanning laser polarimetry. When measuring the diagnostic accuracy of a new test, care must be taken in choosing the standard against which it is compared. This standard should be the best available method of differentiating between those with and without disease. It should effectively discriminate these groups across the full spectrum of disease. For example, an experienced observer can often determine whether a patient has glaucoma before visual field defects occur by performing a careful stereoscopic optic nerve evaluation. Therefore, to require visual field defects as part of the definition of glaucoma artificially limits our ability to assess the test in earlier cases of glaucoma. This is particularly important when the test may be good at detecting early glaucoma. Scanning laser polarimetry has been reported to measure nerve fiber layer thickness, and because defects of the retinal nerve fiber layer have been reported to be present long before visual defects occur, it could be inferred that this technique has promise in the early detection of glaucoma. The current standard for the diagnosis of glaucoma must at least include a careful examination of the optic nerve by an experienced observer, evaluation of a reliable threshold perimeter examination, and the measurement of the intraocular pressure. In this study, this information was provided to 3 separate experienced observers. To further maximize the likelihood of proper classification, a unanimous agreement was required to place an eye in the “glaucoma” or “normal” group. If there was not unanimous agreement, the result was categorized as uncertain, and the eye was not analyzed further.

Sensitivity and specificity of a test will depend on the cutoff used to judge the result as abnormal. We arbitrarily highlighted cutoff values that result in a specificity in the 90% range. With comparable specificity between the algorithms, the sensitivities can be easily compared to determine the more accurate algorithms, at least at that cutoff level. Algorithm results detailing both higher and lower specificity and the associated sensitivities are also provided in Tables 3 through 5 to allow a more comprehensive assessment of the performance of each algorithm. Because the prevalence of glaucoma is low, screening studies should have cutoff values that result in at least 90% specificity to avoid an excessive ratio of false-positive to true-positive test results. If a specificity of about 90% is required, then sensitivity in this study typically ranged between 39% and 57% for early glaucoma, 62% and 71% for moderate glaucoma, and 68% and 81% for severe glaucoma with the full-test mode. The TSNIT image abnormality algorithm did not perform as well. The screening test mode demonstrated sensitivities of 41% to 57% for early glaucoma, 50% to 67% for moderate glaucoma, and 44% to 59% for severe glaucoma. The number provided the greatest accuracy of the algorithms tested. When a cutoff score of 35 or higher was chosen as abnormal, specificity was 89%, but sensitivities were 57%, 71%, and 81% for early, moderate, and severe glaucoma, respectively. Preliminary work by the manufacturer has suggested that scores of 0 to 30 are normal, whereas scores above 70 tend to indicate the presence of glaucoma. In our study, if we use these cutoff values, we find that 85% of subjects we classified as normal have scores of less than 30, but only 17% to 41% of glaucomatous eyes, depending on severity, scored 70 or more.

Two scenarios wherein a more accurate and objective method of diagnosing glaucoma is needed include population screening and when clinically it is uncertain whether a patient has early glaucoma. Objective, accurate, and rapid testing is ideal for screening purposes. It is particularly important for a glaucoma screening test to detect patients who have severe manifestation of disease because substantial morbidity may ensue. Because of the large number of subjects tested and the relatively low prevalence of glaucoma, even in a selected population, specificity needs to be high to avoid overdiagnosing the disease. In this study, even when specificity was required to be only about 90%, scanning laser polarimetry detected only 81% of subjects with severe glaucoma with the full-test mode and 59% with the screening mode. This does not compare favorably with the Henson perimetry test. With the 26-point test algorithm, which requires about 3 minutes per eye, Sponsel et al reported 100% sensitivity and 93.9% specificity for
those with severe glaucoma. Frequency-doubling perimetry has shown promise as a screening method for glaucoma. Johnson and Samuels \cite{16} reported 93% sensitivity and 100% specificity for 15 normal subjects and 15 age-matched patients with early or moderate glaucomatous visual field damage.

There is little difficulty in diagnosing a patient as having glaucoma who demonstrates moderate neural rim loss on stereoscopic examination. It is not uncommon, however, to be faced with a patient who may have some early optic nerve changes, but the findings are somewhat equivocal. Typically, definitive visual field defects will not be present at this early stage of disease. In such a patient, the use of this procedure might be considered because of the theoretical potential for detecting defects of the retinal nerve fiber layer. For example, suppose that there is a 50% likelihood (50% pretest probability) that the patient has glaucoma based on the optic nerve appearance on examination. In this study, the best algorithm tested was the number, with 89% specificity and 37% sensitivity for early glaucoma, as in this example. If the number in this patient is 35 or higher (a positive test result), then the probability that the patient has glaucoma would increase from 50% to 84%. This is also called the positive predictive power of the test and is defined as the ratio of true-positive results to the sum of true- and false-positive results. If the test score is less than 35, then the probability that the patient does not have glaucoma would increase from 50% to 67%. This is the negative predictive value of the test and is defined as the ratio of true-negative results to the sum of true- and false-negative results. The positive and negative predictive values are dependent on sensitivity, specificity, and pretest probability. So if the test result is positive, there is a sufficient likelihood of glaucoma being present to alter follow-up frequency, and some physicians may feel comfortable initiating treatment. If the test result is negative, then there is only a marginal increase in confidence that the patient does not have glaucoma.

We were unable to directly compare our results with those of Tjon-Fo-Sang and Lemij, \cite{1} who reported 96% sensitivity and 93% specificity for detecting glaucoma. They used 3 ratios in their algorithm for classifying a subject. They used the “squares calculation method” to determine ratios. In brief, for the superior, inferior, and nasal segments, they chose 6 squares of 256 pixels each that were devoid of blood vessels. These were handpicked by the investigator. The values were averaged for each segment, and the superior-nasal, inferior-nasal, and superior-inferior ratios were calculated. The current software \cite{1} does not provide an inferior-nasal ratio with normative database comparison. In their study, 56.5% of glaucomatous eyes had superior-inferior ratios outside normal limits. We found that this ratio (also called “symmetry” by the software) identified only 6 of 51 eyes with early, 5 of 42 eyes with moderate, and 0 of 32 eyes with severe glaucoma. However, the superior-nasal ratio was more accurate, detecting 11 of 51 eyes with early, 19 of 42 eyes with moderate, and 15 of 32 eyes with severe glaucoma. Other differences between the studies include a blood vessel removal algorithm for our data and their requirement of the glaucoma hemifield test measuring outside normal limits. As would be expected, a number of eyes in our study showing early glaucoma would not meet their requirements for entry. However, 70 of 73 eyes with moderate or severe glaucoma did have results of a glaucoma hemifield test that were outside normal limits. Therefore, despite having similar levels of visual field damage, the diagnostic accuracy for moderate and severe glaucoma in this study was still well below their reported levels. The average visual field depression appears to be mildly worse in their study (mean deviation, −10.33 dB [95% confidence interval, −31.5 to 0.76 dB] vs −8.88 dB [95% confidence interval, −27.4 to 0.66 dB]), although they used a 30-2 program, as opposed to a 24-2 program, which may account for some of the difference. Our group was older than their group (mean age of subjects with normal eyes, 65.8 vs 49 years, and of those with glaucoma, 73.0 vs 67.1 years). As with our study, they tested a largely white population. In summary, the reason for poorer accuracy in our study is unclear. It does not appear to clearly relate to the fact that some glaucomatous eyes had normal visual fields because the diagnostic accuracy was also much lower for moderate and severe glaucoma.

A limitation of this or any similar study involves the need to restrict the number of diagnostic algorithms tested. Although we may be able to derive a more accurate algorithm with further analysis, the possibility that this will be the result of chance alone also increases. Nevertheless, an exploration of the data, with testing of a large number of possible diagnostic algorithms, might prove interesting, particularly if then applied to an independent data set for validation. For example, in our number of abnormalities algorithm, we arbitrarily weighted a borderline result 50% less than a test result outside normal limits. If we weight borderline results equal to outside normal limits, then accuracy improves modestly. For 3 or more abnormalities (3 combined borderline or outside normal limits parameters) with a specificity of 88%, sensitivity equals 57%, 62%, and 74% for early, moderate, and severe glaucoma, respectively. Further work is needed to optimize our use of the information provided.

This study evaluated the diagnostic accuracy of this procedure in glaucoma. We did not attempt to validate or disprove its ability to accurately measure nerve fiber layer thickness.

CONCLUSIONS

The best algorithm tested for scanning laser polarimetry demonstrated sensitivities of 57%, 71%, and 81% for early, moderate, and severe glaucoma, respectively. This was associated with a specificity of 89%. The best algorithm for the screening test was less sensitive, particularly for those with severe glaucoma. The demonstrated accuracy in detecting severe glaucoma with scanning laser polarimetry appears to be much lower than with existing perimetric screening devices, but studies involving direct comparisons of the techniques are needed.
Because of the particular importance in detecting severe cases, further study will be necessary before scanning laser polarimetry can be recommended as a screening method for glaucoma.

Accepted for publication March 22, 1999.

This study was supported in part by grants from Ann and Joseph F. Heil, Jr, the R. D. and Linda Peters Foundation, Milwaukee, Wis, and Research to Prevent Blindness, Inc, New York, NY.

Reprints: John R. Trible, MD, Wolfe Clinic, 309 E Church St, Marshalltown, IA 50158.

REFERENCES


ARCHIVES Web Quiz Winner

Congratulations to our August Web quiz winner, Norman C. Charles, MD, clinical professor of Ophthalmology at New York University Medical Center, New York. The answer to the August quiz was primary intraocular lymphoma. For a complete discussion of this case, see the “Clinicopathologic Report” section in the September ARCHIVES (Buggage RR, Velez G, Myers-Powell B, Shen D, Whitcup SM, Chan C-C. Primary intraocular lymphoma with a low interleukin 10 to interleukin 6 ratio and heterogeneous IgH gene rearrangement. Arch Ophthalmol. 1999;117:1239-1242).

Be sure to visit the Archives of Ophthalmology World Wide Web site (http://www.archophthalmol.com) and try your hand at our new Clinical Challenge interactive quiz. We invite visitors to make a diagnosis based on selected information from a case report or other feature scheduled to be published in the following month’s print edition of the ARCHIVES. The first visitor to e-mail our Web editors with the correct answer wins an Archives of Ophthalmology CD-ROM and will be recognized in the print journal and on our Web site.

Magnetic resonance images showing generalized cortical atrophy and multiple enhancing lesions (arrows) in December 1998.