Detection of Glaucoma Using Scanning Laser Polarimetry with Enhanced Corneal Compensation

Felipe A. Medeiros, Christopher Bowd, Linda M. Zangwill, Chirag Patel, and Robert N. Weinreb

PURPOSE. To evaluate and compare the diagnostic accuracies for glaucoma detection of scanning laser polarimetry (SLP) with enhanced corneal compensation (GDx ECC) and variable corneal compensation (GDx VCC; both by Carl Zeiss Meditec, Dublin, CA), according to different levels of disease severity and presence of atypical retardation patterns.

METHODS. The study included 102 eyes of 68 patients with glaucoma and 94 eyes of 55 normal subjects. All patients underwent SLP imaging with ECC and VCC methods on the same day. Severity of disease was based on the AGIS (Advanced Glaucoma Intervention Study) visual field score. An ROC regression model was fitted to evaluate the influence of disease severity and atypical retardation patterns (typical scan score [TSS]) on the diagnostic performance of the SLP parameters for both methods.

RESULTS. GDx ECC performed significantly better than GDx VCC in glaucoma detection in patients with more severe atypical retardation patterns. For average disease severity and arbitrarily chosen TSS values of 20, 50, 70, and 100, the ROC curve areas for GDx ECC were 0.910, 0.935, 0.948, and 0.964. Corresponding values for GDx VCC were 0.684, 0.850, 0.920, and 0.975. For lower values of TSS and lower AGIS scores, GDx ECC performed significantly better than GDx VCC.

CONCLUSIONS. GDx ECC performed significantly better than VCC for diagnosing glaucoma in patients with more severe atypical patterns of retardation and at earlier stages of disease.

Scanning laser polarimetry (SLP) is an imaging technology designed to provide objective assessment of the thickness of the retinal nerve fiber layer (RNFL), with potential use for diagnosis and follow-up of patients with glaucoma. It is based on the principle that polarized light passing through the birefringent RNFL undergoes a measurable phase shift, known as retardation, which is linearly related to histologically measured RNFL tissue thickness. The introduction of variable corneal compensation (VCC) in the GDx VCC (Carl Zeiss Meditec, Inc., Dublin, CA) has resulted in improved diagnostic accuracy compared with earlier versions of the SLP technology, which used fixed corneal compensation. However, shortly after the introduction of the GDx VCC, several scans showing atypical retardation patterns (ARPs) were observed in some patients. These scans show irregular patches of elevated retardation values that do not match the expected retardation based on the RNFL anatomy. ARPs seem to result from poor signal-to-noise ratio as a consequence of light scattering in the eye. To compensate for a decrease in signal, the instrument automatically increases the gain to augment the polarization signal, which paradoxically increases the noise from deeper structures such as the sclera. Using subjective evaluation, Bagga et al. found ARPs in 25% of normal and in 51% of glaucomatous eyes, and their presence seems to affect the diagnostic performance of the GDx VCC adversely.

A new software-based method of corneal compensation called Enhanced Corneal Compensation (ECC) has been developed recently, to improve the signal-to-noise ratio while still achieving customized corneal compensation. This method can be incorporated into the GDx VCC instrument and does not require changes in the hardware. The ECC was developed to improve neutralization of ARPs and to increase the dynamic range of the measurements in the low signal range.

To justify the introduction of a new methodology, it is important to demonstrate its superior performance for disease detection under different circumstances. Although some recent studies have suggested that scans obtained with the ECC method provide a better assessment of the RNFL morphology than do those obtained with the VCC method, no study has yet been performed comparing the diagnostic accuracy of these two methods, while adjusting for potentially confounding factors, such as disease severity and amount of ARP in the scans. In the present study, we used statistical models that allowed simultaneous evaluation of the effect of these covariates on the comparisons between the two methods. Diagnostic accuracies of the VCC and ECC methods were then compared for different levels of disease severity and presence of ARPs.

METHODS

This was an observational case-control study. Patients enrolled in the study were included in a prospective longitudinal study designed to evaluate optic nerve structure and visual function in glaucoma (DiGstudy, Diagnostic Innovations in Glaucoma Study) conducted at the Hamilton Glaucoma Center, University of California, San Diego. Patients in the DiGstudy were longitudinally evaluated according to a preestablished protocol that includes regular follow-up visits in which they undergo clinical examination and several other imaging and functional tests. All the data was entered in a computer database, which currently contains information on 1883 subjects, including normal subjects, patients with glaucoma, and those with suspected glaucoma. For the current study, all subjects who had GDx ECC and VCC examinations acquired on the same day were considered for inclusion. From the 226 participants with GDx VCC and ECC tests, 88 were excluded, as they did not fulfill inclusion criteria (including definition of glaucoma and control groups) as specified later in the article. Informed consent was obtained from all participants. The University of California San Diego Human Subjects...
Committee approved all protocols, and the methods described adhered to the tenets of the Declaration of Helsinki. Each subject underwent a comprehensive ophthalmic examination including review of medical history, best corrected visual acuity, slit-lamp biomicroscopy, intraocular pressure (IOP) measurement using Goldmann applanation tonometry, gonioscopy, dilated funduscopic examination using a 78-D lens, stereoscopic optic disc photography, and automated perimetry using 24-2 Swedish Interactive Threshold Algorithm (SITA; Carl Zeiss Meditec Inc.). To be included, subjects had to have best corrected visual acuity of 20/40 or better, spherical refraction within ±5.00 D and cylinder correction within ±3.00 D, and open angles on gonioscopy. Eyes with coexisting retinal disease, uveitis, or nonglaucomatous optic neuropathy were also excluded from this investigation.

Eyes were classified as glaucomatous if they had repeatable (at least two consecutive) abnormal visual field test results, defined as a PSD outside the 95% normal confidence limits and/or a Glaucoma Hemifield Test result outside normal limits, regardless of the appearance of the optic disc. Eyes were also classified as glaucomatous if they had documented evidence of progressive glaucomatous change in the appearance of the optic disc as assessed by simultaneous stereoscopic optic disc photographs (TRC-SS; Topcon Instrument Corp. of America, Paramus, NJ), regardless of visual field test results. The evidence of progressive glaucomatous damage had to be present before the imaging test date. The use of this composite reference standard for glaucoma diagnosis allowed us to evaluate the accuracy of diagnostic tests in a broad spectrum of patients with the disease, because we could include both patients with visual field loss, as well as patients with normal visual fields but confirmed progressive glaucomatous optic nerve damage.

For evaluation of progressive optic disc damage, stereoscopic sets of slides were examined using a stereoscopic viewer (Asahi, Pentax, Tokyo, Japan). The photographs were evaluated by two experienced graders, and each was masked to the subject’s identity and to the other test results. For inclusion, photographs had to be of adequate quality or better. To identify a subgroup of patients with progressive glaucomatous optic disc change for this study, our research database was reviewed for all patients who had been imaged using both GDx VCC and ECC and who had been observed for at least 1 year before the imaging test date. For each patient, the most recent stereophotograph was compared with the oldest available one, to maximize the chance of detecting progressive optic disc change. Each observer was masked to the temporal sequence of the photographs. The definition of change was based on focal or diffuse thinning of the neuroretinal rim, increased excavation, or enlargement of RNFL defects. Changes in rim color, presence of disc hemorrhage or progressive parapapillary atrophy were not sufficient for characterization of progression. Discrepancies between the two graders were resolved either by consensus or by adjudication of a third experienced grader.

Patients with glaucoma were required to have visual field loss and/or progressive GON in at least one eye, regardless of intraocular pressure. When both eyes of the same patient had glaucoma as defined, both eyes were included in the study, provided that they satisfied other inclusion criteria. When only one eye had a diagnosis of glaucoma, that eye was included.

For the population with progressive optic disc change, visual field results were not used as a criterion for inclusion in the study. However, results of the visual field test closest to the imaging date were analyzed as part of the study results for all patients. The AGIS (Advanced Glaucoma Intervention Study) score was used to evaluate the severity of visual field loss. This score has been described in detail elsewhere. It is based on the extent of depression at different locations of the visual field test in the total deviation plot and can range from 0 (no field loss) to 20 (end stage). A computer program was developed in commercial software (MatLab ver. 7.0; The MathWorks, Inc., Natick, MA) to allow automatic calculation of the AGIS score from data exported from the Humphrey perimeter (Carl Zeiss Meditec, Inc.).

Normal subjects were recruited from the general population through advertisement, as well as from the staff and employees of the University of California San Diego. They were selected so that their age range was similar to that of subjects with glaucoma (40-90 years). Normal control eyes had normal findings in clinical examination, intraocular pressure of 21 mm Hg or less with no history of increased IOP, and a normal visual field result. A normal visual field was defined as a mean deviation (MD) and pattern SD (PSD) within 95% confidence limits, and a Glaucoma Hemifield Test (GHT) result within normal limits. Normal control eyes were required to have a normal appearance of the optic nerve (no focal rim thinning, glaucomatous excavation or RNFL defects). When both eyes of normal subjects satisfied inclusion criteria, both were included in the analyses.

All patients had had visual field and imaging tests within 6 months.

**Instrumentation**

**SLP with VCC.** Patients were imaged with a commercially available scanning laser polarimeter with variable corneal compensation (GDx VCC, software ver. 5.5.1; Carl Zeiss Meditec, Dublin, CA). The general principles of SLP and the algorithm used for VCC have been described in detail elsewhere. Because corneal polarization axis and magnitude affect SLP measurements and are not similar in all eyes, the GDx VCC employs a variable corneal polarization compensator that allows eye-specific compensation of anterior chamber birefringence. After determining the axis and magnitude of corneal polarization in each eye by macular scanning, three appropriately compensated retinal polarization images per eye were automatically obtained and combined to form each mean image used for analysis. Only well focused, evenly illuminated and centered scans with SD ≤ 7 µm, determined by GDx VCC software, were included (cutoff suggested by oral communication, June 2005, Qienyuan Zhou, PhD, Carl Zeiss Meditec, Inc.). From an original sample of 215 eyes of 138 patients, 19% (eyes of 15 patients were excluded due to poor-quality images with the GDx VCC.

GDx VCC parameters provided in the standard printout of the instrument and investigated in this article were superior average, inferior average, TSNT (circumpapillary RNFL thickness measured under the automatically defined 3.2-mm-diameter calculation circle: T, temporal sector; S, superior sector; N, nasal sector; I, inferior sector) average, TSNT standard deviation, and the Nerve Fiber Indicator (NFI). The NFI is calculated by using a support vector machine algorithm based on several RNFL measures and assigns a number from 0 to 100 to each eye. According to the manufacturer, the higher the NFI, the greater the likelihood that the patient has glaucoma.

To quantify the presence of ARPs on GDx VCC scans, we used the software-provided parameter Typical Scan Score (TSS). The TSS is a continuous variable ranging from 0 to 100 and is the result of a support vector machine analysis of SLP data labeled for training based on the subjective appearance of each scan (typical versus atypical). TSS is based on the slope, SD, and average magnitude of RNFL thickness measurements from the edge of the optic disc extending outward to 20°. Low TSSs indicate atypical scans and high TSSs indicate typical ones.

**SLP with ECC.** The ECC algorithm was implemented by the manufacturer on a commercially available scanning laser polarimeter. For each patient, GDx ECC scans were obtained on the same day and on the same machine as were the GDx VCC scans. The principles and rationale of the ECC method have been described in detail elsewhere. In brief, a known birefringence bias is introduced into the measurement beam path to shift the measurement of total retardation into a more sensitive region of the curve of detection of polarization of the instrument. The bias retarder is formed by the combination of the VCC and cornea. However, instead of completely canceling corneal birefringence, the retarder is adjusted so that the combination has retardance close to 55 nm and slow axis of polarization close to vertical. After image acquisition, the birefringent bias is removed mathematically, point-by-point, to yield the RNFL retardation values. As with GDx VCC scans, only well focused, evenly illuminated and centered GDx ECC scans with SD ≤ 7 µm were included. Three images per eye were automatically obtained and combined to form each mean
image used for analysis. From an original sample of 215 eyes from 138 patients, 15 (7%) eyes from 12 patients were excluded due to poor-quality images with GDx ECC. To evaluate GDx ECC diagnostic accuracy, we used the same parameters provided by the GDx VCC: NFI, TSNIT average, superior average, inferior average and TSNIT SD. Although the NFI has been originally developed for VCC scans, preliminary analyses showed that it still performs well with ECC when compared with other parameters. Therefore, we also analyzed the results of the NFI for GDx ECC in the present study (see the Discussion section).

Data Analysis

Descriptive statistics included mean and standard deviation for normally distributed variables and median, first quartile, and third quartile values for nonnormally distributed variables. Student’s t-tests or Mann-Whitney tests were used to evaluate demographic and clinical differences between subjects with glaucoma and normal subjects.

Receiver operating characteristic (ROC) curves were used to describe the ability of the SPL to discriminate patients with glaucoma from healthy subjects. The ROC curve shows the tradeoff between sensitivity and 1 − specificity. An ROC curve area of 1.0 represents perfect discrimination, whereas an area of 0.5 represents chance discrimination. Empiric ROC curves were reported for each parameter of GDx VCC and ECC methods. Although these “pooled” ROC curves represent the average performance of each parameter in the population, they are limited by the lack of adjustment for potentially confounding factors, such as disease severity. For example, it is conceivable that the performance of the test for detection of patients with more advanced disease will be better than in patients with early disease and, therefore, the ROC curve area for test parameters will be different in these two situations. Also, ROC curves for glaucoma detection with GDx VCC and ECC could vary with the severity of ARPs. Therefore, it is important to characterize the relationship between the performance of the diagnostic test and these covariates and to evaluate how this relationship affects the comparison between these two tests. To accomplish this, we used an ROC regression methodology.

ROC Regression Model

In the present study, we used an ROC regression modeling technique to evaluate the influence of atypical scan patterns and severity of disease on the diagnostic accuracy of GDx VCC and ECC in glaucoma. This modeling approach was initially described by Medeiros et al.15 for evaluation of the influence of covariates on the performance of diagnostic tests in glaucoma. This methodology allows the evaluation of the influence of covariates on the diagnostic performance of the test, so that ROC curves for specific values of the covariate of interest can be obtained. Also, it allows adjustment for the possible confounding effects of other covariates. Details of the modeling procedure have been described previously.15,16 In brief, the $\text{ROC}_{X,X_0}(q)$ is the probability that a diseased individual with disease-specific covariates $X_0$ (that is, covariates specific to diseased subjects such as disease severity, for example) and common covariates $X$ (covariates common to both diseased and healthy subjects) has test results $X_1$ that are greater than or equal to the $q$th quantile of the distribution of test results from nondiseased individuals. That is, when the specificity of the test is $1 - q$, the sensitivity is $\text{ROC}_{X,X_0}(q)$. The general ROC regression model can be written as:

$$\text{ROC}_{X,X_0}(q) = \Phi(\alpha_1 + \alpha_2 + \beta_1 X + \beta_2 X_0)$$

where the coefficients $\alpha_1$ and $\alpha_2$ are the intercept and slope of the ROC curve, respectively, and $\Phi$ is the normal cumulative distribution function. If the coefficient for a specific variable $X$ ($\beta_1$) is greater than zero, then the discrimination between diseased and nondiseased subjects increases with increasing values of this covariate. Similarly, if the coefficient for the disease-specific covariate $X_0$ ($\beta_2$) is greater than zero, then diseased subjects with larger values of this covariate are more distinct from nondiseased subjects than are diseased subjects with smaller values of $X_0$.

In the present study, an ROC model was fitted to assess the influence of the common covariate TSS on the diagnostic performance of the GDx ECC and VCC parameters. The model was adjusted for the disease-specific covariate severity and the common covariate age. The following ROC regression model was fitted for each parameter evaluated:

$$\text{ROC}_{X,X_0}(q) = \Phi(\alpha_1 + \alpha_2 \Phi^{-1}(q) + \beta_1 \text{ECC} + \beta_2 \text{ECC} \times \Phi^{-1}(q) + \beta_3 \text{TSS} + \beta_4 \text{TSS} \times \text{ECC} + \beta_5 \text{severity} + \beta_6 \text{age})$$

where $\text{ECC}$ is a binary variable indicating the type of test (GDx VCC was used as the reference test), and $\text{TSS}$ is a continuous variable quantifying the presence of atypical patterns of retardation. An interaction term between the variable $\text{ECC}$ and $\Phi^{-1}(q)$ was included to verify whether the performance of the ECC differed by varying amounts depending on the false-positive rate $q$ (or specificity $1 - q$)—that is, whether the shapes of ROC curves for VCC and ECC tests were different. The interaction term between ECC and TSS was included to assess whether the influence of ARPs was similar or different between GDx VCC and ECC tests. Finally, to adjust for severity of disease and age, the model included a variable severity indicating severity of glaucomatous damage as measured by the AGIS score, and a variable age indicating patient’s age. No variable selection method was used. Because the models were developed for hypothesis testing, there was little concern for parsimony. The full prespecified model fit, including all variables, results in more accurate probabilities for tests of variables of interest.17

Parameters were estimated using probit regression. To obtain confidence intervals for regression parameters, a bootstrap resampling procedure was used ($n = 500$ resamples).18 As measurements from both eyes of the same subject are likely to be correlated, the use of standard statistical methods for parameter estimation can lead to underestimation of standard errors and to confidence intervals that are too narrow.19 Therefore, to account for the fact that both eyes of some subjects were used for analyses, the cluster of data for the study subject was considered as the unit of resampling when calculating standard errors. This procedure has been used to adjust for the presence of multiple correlated measurements from the same unit.18,20

Statistical analyses were performed using commercial software (STATA ver. 9.0; StataCorp, College Station, TX; and SPSS ver. 13.0; SPSS Inc., Chicago, IL). A level of 0.05 was set as the cutoff. The study had a power of 0.85 to detect a difference of 0.05 between ROC curve areas for the parameter NFI, assuming $\alpha = 0.05$.

Results

The study included 102 eyes of 68 patients with glaucoma and 94 eyes of 55 normal subjects. The mean ± SD age of patients with glaucoma and normal subjects was 68 ± 11 and 59 ± 11 years, respectively ($P < 0.001$). Median (first quartile, third quartile) MD and PSD of the visual field closest to the imaging date in glaucomatous eyes were −4.05 dB (−7.56 dB, −1.60 dB) and 3.60 dB (2.19 dB, 9.00 dB). Corresponding values for normal eyes were −0.05 dB (−0.90 dB, 0.32 dB) and 1.45 dB (1.29 dB, 1.64 dB). MD and PSD of glaucomatous eyes were significantly different from those from normal eyes ($P < 0.001$ for both comparisons). From the 102 eyes of patients with glaucoma, 63 (62%) had evidence of progressive optic disc changes based on longitudinal assessment of stereophotographs, and 74 (73%) had repeatable glaucomatous visual field defects. Thirty-five (34%) eyes had both evidence of progressive optic disc change and abnormal visual fields. Twenty-eight of the 102 (27%) eyes had evidence of progressive optic disc change, but no visual field loss and were classified as preperimetric glaucoma. A total of 74 (73%) of 102 eyes had visual field defects and were classified as perimetric glaucoma. The
average AGIS score for all 102 glaucomatous eyes was 4.23 (median: 2.50; first quartile: 0; third quartile: 6.25). Figure 1 shows the distribution of AGIS scores in the glaucomatous eyes included in the study.

Table 1 shows the areas under the empiric ROC curves for each parameter from GDx ECC and VCC. These ROC curve areas represent the overall performance of the parameters in the population sample and are not adjusted for severity of disease and TSS. The parameter NFI had the best performance for discriminating glaucomatous from healthy eyes for both the GDx ECC and the VCC. For this parameter, there was no statistically significant difference between the two methods ($P = 0.254$). Figure 2 shows empiric ROC curves for the parameter NFI for both methods. For the parameters TSNIT average, superior average, and inferior average, GDx ECC performed significantly better than VCC. No significant difference was found for the parameter TSNIT SD.

Tables 2 and 3 show the areas under the empiric ROC curves for each parameter from GDx ECC and VCC for discriminating between patients with preperimetric and perimetric glaucoma with both technologies. These ROC curves were also not adjusted for TSS. The differences between ECC and VCC performances tended to remain similar; however, as expected, the ROC curve areas were lower for detection of patients with preperimetric glaucoma than for patients with perimetric glaucoma with both technologies.

To evaluate the influence of atypical scans and severity of disease on the diagnostic performance of GDx ECC and VCC, we used an ROC regression model for the parameter with best overall performance in both methods (i.e., the NFI). Table 4 shows the estimates of the coefficients of the ROC regression model. The results of the model indicated a significant influence of atypical scan patterns on the diagnostic performance of the GDx VCC, as indicated by the statistically significant value attributed to the coefficient representing atypical patterns ($P = 0.001$). Another significant influence was found for the parameter TSNIT SD.

The severity of disease, as measured by the AGIS score had a significant influence on the diagnostic performance of both GDx ECC and VCC, as indicated by the statistically significant value attributed to the coefficient representing severity ($P = 0.001$). As expected, both tests performed better in patients with more severe disease. The severity of disease, as measured by the AGIS score had a significant influence on the diagnostic performance of both GDx ECC and VCC, as indicated by the statistically significant value attributed to the coefficient representing severity ($P = 0.001$). As expected, both tests performed better in patients with more severe disease.

Figure 4 illustrates the relationship among ROC curve areas, disease severity, and TSS for the parameter NFI with GDx ECC (Fig. 4A) and VCC (Fig. 4B). Figure 4C shows a contour plot of the differences in ROC curve areas for GDx ECC and VCC for arbitrarily chosen levels of TSS, adjusted for mean disease severity and age. ROC curve areas and probabilities for the comparison between ROC curves are shown in Table 5. At TSSs of 20 and 50, GDx ECC had a significantly better performance than did GDx VCC. At TSSs of 70 and 100, the differences were not statistically significant.

The severity of disease, as measured by the AGIS score had a significant influence on the diagnostic performance of both GDx ECC and VCC, as indicated by the statistically significant value attributed to the coefficient representing severity ($P = 0.001$). As expected, both tests performed better in patients with more severe disease. Figure 4 illustrates the relationship among ROC curve areas, disease severity, and TSS for the parameter NFI with GDx ECC (Fig. 4A) and VCC (Fig. 4B). Figure 4C shows a contour plot of the differences in ROC curve areas between the two methods, according to different levels of the AGIS score and TSSs. The differences between the

![Figure 1](http://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/933444/)

**Figure 1.** Distribution of AGIS score values among the 102 glaucomatous eyes included in the study.

![Figure 2](http://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/933444/)

**Figure 2.** Empiric ROC curves for the parameter NFI for the GDx ECC and GDx VCC.

### Table 1. Areas under the Empirical ROC Curves for GDx ECC and GDx VCC Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GDx ECC</th>
<th>GDx VCC</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFI</td>
<td>0.944 (0.916-0.971)</td>
<td>0.920 (0.881-0.958)</td>
<td>0.254</td>
</tr>
<tr>
<td>TSNIT average</td>
<td>0.930 (0.896-0.964)</td>
<td>0.751 (0.681-0.821)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Superior average</td>
<td>0.884 (0.835-0.932)</td>
<td>0.821 (0.759-0.882)</td>
<td>0.011</td>
</tr>
<tr>
<td>Inferior average</td>
<td>0.918 (0.881-0.955)</td>
<td>0.792 (0.727-0.856)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>TSNIT SD</td>
<td>0.816 (0.757-0.875)</td>
<td>0.881 (0.836-0.927)</td>
<td>0.080</td>
</tr>
</tbody>
</table>

Data are areas under the ROC curves (95% CI). These ROC curve areas represent the overall performance of the tests and are not adjusted for severity of disease or presence of atypical patterns of retardation.
performances of GDx ECC and VCC were marked for lower values of TSS, especially at early disease states.

**DISCUSSION**

Using ROC regression analysis, we demonstrated that SLP measurements obtained with the ECC method performed significantly better for detection of glaucoma than measurements acquired with the VCC method. Differences in performance between the two methods were more marked for detection of damage in patients showing atypical patterns of retardation, when imaging with the GDx VCC had poor diagnostic accuracy. The results of our study support the use of ECC as a better method for evaluation of the RNFL with SLP.

The parameter NFI had the best performance for detection of glaucoma for the GDx VCC, in agreement with previous studies in the literature. For the GDx ECC, the NFI also had the best overall performance. When only empiric ROC curves were analyzed, no significant difference was found for the diagnostic performance of this parameter between the VCC (ROC curve area = 0.920) and ECC (ROC curve area = 0.944) methods. Although the empiric ROC curves represent the overall performance of a test in the population sample, they are limited by the lack of adjustment for potentially confounding factors. In fact, by incorporating disease severity and TSS as covariates on ROC regression models, we were able to demonstrate a significant effect of these two variables on the diagnostic performance of VCC and ECC. The GDx VCC performed poorly in subjects with atypical patterns of retardation, whereas the performance of the GDx ECC was largely unaffected. This resulted in a marked difference between the performance of the two methods in subjects with low TSS, as can be seen in Table 5 and Figure 4. The ROC curve areas for a TSS of 20 were 0.910 for the GDx ECC but only 0.684 for the VCC, considering an average disease severity (AGIS score of 4). The lack of significant effect of TSS on the diagnostic performance of the GDx ECC indicates that scans obtained by this method do not seem to be affected by atypical patterns of retardation. Figure 4C also shows that the influence of atypical patterns on the difference in performance between the two methods was more marked in subjects with earlier stages of disease. In fact, when the same regression model is used, the estimated ROC curves for TSS of 20 and AGIS score of 0 would be 0.576 for VCC and 0.856 for ECC. For the same level of TSS, but AGIS score of 15, for example, the corresponding values would be 0.897 and 0.982. This is not surprising, as in subjects with advanced disease the degree of RNFL damage would generally enable its detection by the VCC even in the presence of ARPs. It is important to emphasize that these differences in performance between the two methods would have been missed if only the overall ROC curves had been analyzed.

The NFI is calculated using a sophisticated machine learning classifier method that takes into account several parameters of the RNFL and is intended to provide the best measure of the current RNFL status obtained with SLP. It is important to emphasize that the NFI was originally derived for use with the GDx VCC. Although subjects with very atypical scans were excluded from the original dataset used to develop the NFI algorithm (Q Zhou, PhD, Carl-Zeiss Meditec, Inc., written communication, September 2005), it is likely that scans with ARPs could still have influenced the parameters (or their relative importance) used to compose the support vector machine. Therefore, it is possible that a machine learning classifier developed specifically for the GDx ECC would perform even better than the current NFI. Further studies are necessary to investigate this question. It should also be noted that the parameter TSS was originally developed for estimating the presence of ARPs on scans obtained using the VCC method and, in our study, the TSS variable represents this parameter on VCC scans. Although the TSS has also been used to estimate the presence of ARPs on ECC scans, it is possible that atypical patterns on these scans would not be detected by this parameter.

It is interesting to note that the difference in performance between the NFI and the other RNFL parameters was more marked for VCC than ECC. Absolute thickness parameters such as average thickness performed much better with the ECC than with the VCC method, and this could be the result of an increased susceptibility of absolute thickness parameters to atypical patterns of retardation. A recent study by Reus et al. found overall ROC curve areas of 0.87 and 0.96 for the parameter average thickness for the VCC and ECC methods, respectively. In our study, corresponding empiric overall ROC curve areas for this parameter were 0.751 and 0.930. In interpreting differences between our study and the one by Reus et al., it is important to note that while in our study we acquired three images for both GDx VCC and ECC to obtain final mean images used for analysis, they used only one image from each scan.

**Table 2. Areas under the Empirical ROC Curves for GDx ECC and GDx VCC Parameters for Discriminating Patients with Preperimetric Glaucoma from Healthy Subjects**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GDx ECC</th>
<th>GDx VCC</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFI</td>
<td>0.892 (0.829–0.954)</td>
<td>0.893 (0.833–0.953)</td>
<td>0.944</td>
</tr>
<tr>
<td>TSNIT average</td>
<td>0.877 (0.800–0.954)</td>
<td>0.674 (0.537–0.812)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Superior average</td>
<td>0.817 (0.714–0.920)</td>
<td>0.725 (0.597–0.852)</td>
<td>0.013</td>
</tr>
<tr>
<td>Inferior average</td>
<td>0.861 (0.781–0.941)</td>
<td>0.724 (0.593–0.855)</td>
<td>0.017</td>
</tr>
<tr>
<td>TSNIT SD</td>
<td>0.747 (0.649–0.845)</td>
<td>0.796 (0.704–0.888)</td>
<td>0.428</td>
</tr>
</tbody>
</table>

Data are the area under the ROC curve (95% CI).

**Table 3. Areas under the Empirical ROC Curves for GDx ECC and GDx VCC Parameters for Discriminating Patients with Perimetric Glaucoma from Healthy Subjects**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GDx ECC</th>
<th>GDx VCC</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFI</td>
<td>0.964 (0.941–0.986)</td>
<td>0.930 (0.887–0.972)</td>
<td>0.167</td>
</tr>
<tr>
<td>TSNIT average</td>
<td>0.950 (0.919–0.981)</td>
<td>0.780 (0.705–0.855)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Superior average</td>
<td>0.909 (0.860–0.957)</td>
<td>0.857 (0.792–0.922)</td>
<td>0.081</td>
</tr>
<tr>
<td>Inferior average</td>
<td>0.940 (0.906–0.974)</td>
<td>0.817 (0.749–0.884)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>TSNIT SD</td>
<td>0.842 (0.780–0.903)</td>
<td>0.914 (0.872–0.955)</td>
<td>0.052</td>
</tr>
</tbody>
</table>

Data are areas under the ROC curves (95% CI).
type for inclusion in the analyses. However, the larger values found in the study by Reus et al. are most likely due to the worse disease severity of the patients with glaucoma included in their study (average MD = −12 dB) compared with ours (average MD = −6 dB). In fact, when an ROC regression model was used in our population to estimate the accuracy of the parameter average thickness for MD of −12 dB, ROC curve areas of 0.868 and 0.972 for VCC and ECC, respectively, were obtained. These values are almost identical with those found by Reus et al.

A significant effect of disease severity was demonstrated on the diagnostic accuracy of both VCC and ECC. The inclusion of patients with glaucoma without visual field loss along with patients with early, moderate, and advanced visual field defects allowed us to evaluate the performance of these tests in a broad spectrum of patients with the disease. The diagnostic performance estimates for detection of glaucoma in subjects with AGIS score equal to 0 would represent the performance expected for detection of disease in subjects with no visual field loss. These estimates are relevant in the clinical situation of using these tests for detection of disease in those with suspected glaucoma who do not show visual field abnormalities. For the GDx ECC, the ROC curve area for the NFI parameter (estimated from the ROC regression model) was 0.920, assuming an AGIS score of 0 and a TSS of 82 (average TSS in the study population). Estimated values of sensitivity at 80% and 95% specificities, extracted from the ROC curves, were 87% and 65%, respectively. When empiric ROC curves were analyzed (Table 2), the area under the ROC curve for GDx ECC NFI (0.892) was similar to that estimated by the regression model. These results show the potential of this technology for detection of patients with preperimetric glaucoma.

It should be emphasized that the diagnosis of glaucoma in patients without visual field loss in our study was performed based on documented evidence of progressive optic disc damage on stereophotographs and, therefore, was more robust than a diagnosis of glaucomatous optic neuropathy based on an optic disc examination at a single time point, as used in other studies. This approach was originally proposed by Medeiros et al. for evaluation of diagnostic tests in glaucoma. In the absence of visual field loss, a diagnosis of certainty of glaucoma usually requires demonstrating a history of progressive glaucomatous changes to the optic nerve. It should be noted, however, that this inclusion criterion is still limited in its ability to identify glaucoma. It is possible that patients with glaucomatous damage to the optic nerve but no visual field loss were not included in the study simply because the disease did not progress in them during the follow-up period covered by the stereophotographs. This factor would limit the inclusion of patients with early disease who did not show progressive disc damage. However, this is a current limitation of all diagnostic studies in glaucoma, as no perfect reference standard exists for the disease. Also, normal subjects did not undergo longitudinal evaluation of optic disc photographs. An ideal control group would consist of patients with suspect optic nerve appearance observed over a very long period without showing any evidence of progressive glaucomatous optic disc changes, providing reasonable confidence that they had only a normal variation of optic disc features, but no GON. These patients would also need to be left untreated during the entire follow-up time to avoid any confounding effects of treatment in the prevention of development of progressive optic disc changes. Unfortu-

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient Estimate</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>α₁</td>
<td>2.160</td>
<td>(1.730–2.589)</td>
</tr>
<tr>
<td>Φ⁻¹ (q₁)</td>
<td>α₂</td>
<td>0.913</td>
<td>(0.659–1.167)</td>
</tr>
<tr>
<td>ECC</td>
<td>β₁</td>
<td>0.185</td>
<td>(−0.245–0.614)</td>
</tr>
<tr>
<td>ECC × Φ⁻¹ (q₁)</td>
<td>β₂</td>
<td>0.050</td>
<td>(−0.220–0.320)</td>
</tr>
<tr>
<td>TSS</td>
<td>β₄</td>
<td>0.025</td>
<td>(0.011–0.039)</td>
</tr>
<tr>
<td>TSS × ECC</td>
<td>β₅</td>
<td>−0.017</td>
<td>(−0.029–0.006)</td>
</tr>
<tr>
<td>Severity</td>
<td>β₆</td>
<td>0.097</td>
<td>(0.018–0.177)</td>
</tr>
<tr>
<td>Age</td>
<td>β₇</td>
<td>0.001</td>
<td>(−0.026–0.027)</td>
</tr>
</tbody>
</table>

VCC was used as the reference test. The variables severity, age, and TSS were centered on their mean values. q, false-positive fraction (1 - specificity); Φ⁻¹, inverse normal cumulative distribution function.

<table>
<thead>
<tr>
<th>TSS</th>
<th>ROC Curve Area</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.910</td>
<td>0.684</td>
</tr>
<tr>
<td>50</td>
<td>0.935</td>
<td>0.850</td>
</tr>
<tr>
<td>70</td>
<td>0.949</td>
<td>0.920</td>
</tr>
<tr>
<td>100</td>
<td>0.964</td>
<td>0.975</td>
</tr>
</tbody>
</table>

TABLE 4. Results of the ROC Regression Model Incorporating TSS, Type of Scan, Age and Disease Severity as Covariates

FIGURE 3. ROC curves for the parameter NFI for GDx ECC and GDx VCC for arbitrary values of TSS, according to the regression model.
nately, such a control group was not available for the present study. It should be noted, however, that the inclusion criteria for the control group used in the present study are unlikely to influence the comparison between the ECC and VCC methods for GDx.

Although ROC curves are a useful and important index for evaluation and comparison of the performance of diagnostic tests under certain circumstances, they have limited intrinsic clinical meaning. Other indexes, such as likelihood ratios, may have more straightforward clinical interpretation and application. Medeiros et al.6 have recently demonstrated the usefulness of likelihood ratios for interpretation of results of imaging tests in glaucoma. Further studies are necessary in which likelihood ratios for glaucoma diagnosis are evaluated with the GDx ECC.

In conclusion, our results demonstrate the superiority of the ECC over the VCC method for SLP analysis of the RNFL in glaucoma. The difference in performance between these two methods was most evident for diagnosing RNFL loss in patients showing atypical patterns of retardation and at early stages of the disease.

References

15. Medeiros FA, Sample PA, Zangwill LM, et al. A statistical approach to the evaluation of covariate effects on the receiver operating

**Figure 4.** Relationship between areas under the ROC curves for the parameter NFI and TSS, according to the level of visual field damage measured by the AGIS score. Areas under the ROC curves are shown for (A) GDx ECC and (B) GDx VCC. (C) A contour plot of the differences in the areas under the ROC curves for the parameter NFI for GDx ECC and GDx VCC according to the levels of AGIS score and TSS. Differences between the ROC curve areas (GDx ECC – GDx VCC) are represented by the color scale. Differences were marked for lower TSSs, especially at early disease states.


