Comparison of the GDx VCC Scanning Laser Polarimeter, HRT II Confocal Scanning Laser Ophthalmoscope, and Stratus OCT Optical Coherence Tomograph for the Detection of Glaucoma

Felipe A. Medeiros, MD; Linda M. Zangwill, PhD; Christopher Bowd, PhD; Robert N. Weinreb, MD

**Objective:** To compare the abilities of current commercially available versions of 3 optical imaging techniques: scanning laser polarimetry with variable corneal compensation (GDx VCC), confocal scanning laser ophthalmoscopy (HRT II [Heidelberg Retina Tomograph]), and optical coherence tomography (Stratus OCT) to discriminate between healthy eyes and eyes with glaucomatous visual field loss.

**Methods:** We included 107 patients with glaucomatous visual field loss and 76 healthy subjects of a similar age. All individuals underwent imaging with a GDx VCC, HRT II, and fast retinal nerve fiber layer scan with the Stratus OCT as well as visual field testing within a 6-month period. Receiver operating characteristic curves and sensitivities at fixed specificities (80% and 95%) were calculated for parameters reported as continuous variables. Diagnostic categorization (outside normal limits, borderline, or within normal limits) provided by each instrument after comparison with its respective normative database was also evaluated, and likelihood ratios were reported. Agreement on categorization between methods (weighted \( \kappa \)) was assessed.

**Results:** After the exclusion of subjects with unacceptable images, the final study sample included 141 eyes of 141 subjects (75 with glaucoma and 66 healthy control subjects). Mean±SD mean deviation of the visual field test result for patients with glaucoma was \(-4.87\pm3.9\) dB, and 70% of these patients had early glaucomatous visual field damage. No statistically significant difference was found between the areas under the receiver operating characteristic curves (AUCs) for the best parameters from the GDx VCC (nerve fiber indicator, AUC=0.91), Stratus OCT (retinal nerve fiber layer inferior thickness, AUC=0.92), and HRT II (linear discriminant function, AUC=0.86). Abnormal results for each of the instruments, after comparison with their normative databases, were associated with strong positive likelihood ratios. Chance-corrected agreement (weighted \( \kappa \)) among the 3 instruments ranged from moderate to substantial (0.50-0.72).

**Conclusions:** The AUCs and the sensitivities at high specificities were similar among the best parameters from each instrument. Abnormal results (as compared with each instrument’s normative database) were associated with high likelihood ratios and large effects on posttest probabilities of having glaucomatous visual field loss. Calculation of likelihood ratios may provide additional information to assist the clinician in diagnosing glaucoma with these instruments.

This observational cross-sectional study included 183 eyes of 183 patients (107 patients with glaucoma and 76 healthy control subjects). All patients were evaluated at the Hamilton Glaucoma Center, University of California, San Diego, from April 2002 to November 2003. These patients were included in a prospective longitudinal study designed to evaluate optic nerve structure and visual function in glaucoma (Diagnostic Innovations in Glaucoma Study). All patients who met the inclusion criteria were enrolled in the current study. 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 ophthalmologic examination including a review of the medical history, best-corrected visual acuity measurement, slitlamp biomicroscopy, intraocular pressure measurement using Goldmann application tonometry, gonioscopy, dilated funduscopy examination using a 78D lens, stereoscopic optic disc photography, and automated perimetry using the 24-2 Swedish Interactive Threshold Algorithm (Carl Zeiss Meditec, Inc). To be included, subjects had to have a best-corrected visual acuity of 20/40 or better in the affected eye, spherical refraction within ±5.0 diopters (D) and cylinder correction within ±3.0 D, and open angles on gonioscopy. Eyes with coexisting retinal disease, uveitis, or nonglaucomatous optic neuropathy were excluded from this investigation. One eye of each patient was randomly selected for inclusion in the study.

Normal control eyes had an intraocular pressure of 21 mm Hg or less (with no history of increased intraocular pressure) and a normal visual field test result. A normal visual field test was defined as a mean deviation and pattern standard deviation within 95% confidence limits, and a glaucoma hemifield test result within normal limits. Normal control eyes also had a healthy appearance of the optic disc and RNFL (no diffuse or focal rim thinning, cupping, optic disc hemorrhage, or RNFL defects), as evaluated by clinical examination. Eyes were classified as glaucomatous if they had repeated (2 consecutive) abnormal visual field test results, defined as a pattern standard deviation outside the 95% normal confidence limits or a glaucoma hemifield test result outside normal limits, regardless of the appearance of the optic disc.

All patients underwent ocular imaging with the GDx VCC, HRT II, and Stratus OCT. For each subject, all ocular imaging and visual field examinations were completed within 6 months.®

**INSTRUMENTATION AND DATA ANALYSIS**

**GDx VCC Scanning Laser Polarimeter**

All patients underwent imaging using a commercially available scanning laser polarimeter (GDx VCC; software version 5.0.1; Laser Diagnostic Technologies, Inc). The general principles of scanning laser polarimetry have been described in detail elsewhere. The GDx VCC is a modified scanning laser polarimetry system with variable corneal compensation. Images of the ocular fundus are formed by scanning the beam of a near infrared laser (780 nm) in a raster pattern. The scan raster covers an image field 40° horizontally and 20° vertically, including both the parapapillary and macular regions of the eye. With the GDx VCC, the method of variable corneal polarization compensation, as described by Zhou and Weinreb, has been automated and has replaced the original fixed corneal compensator. The variable corneal compensator in this system consists of 2 identical linear retarders in rotating mounts so that both the retardation and axis of the unit can be adjusted according to requirements. To measure eye-specific corneal polarization axis and magnitude, scanning laser polarimetry images of the macula are first acquired without compensation (the retardation is set to 0). The uniform radial birefringence of the Henle fiber layer in the macula is used as an “intraocular polarimeter,” and both the Henle fiber layer and corneal retardation can be determined from the macular retardation profile. Next, corneal birefringence-compensated scanning laser polarimetry images are obtained using the appropriate eye-specific corneal polarization axis and magnitude values by adjusting the variable corneal compensation retarders. The GDx VCC measures retardation in nanometers. To simplify communications, retardation values are converted into thickness values (micrometers) using a fixed conversion factor of 0.67 nm/μm.

In this study, a baseline image was automatically created from 3 images obtained for each subject. Assessment of GDx VCC image quality was performed by an experienced examiner masked to the subject’s identity and results from the other tests. The assessment was based on the appearance of the reflectance image, presence of residual anterior segment retardation, and presence of an atypical pattern of retardation. To be classified as good quality, an image required a focused and evenly illuminated reflectance image with a centered optic disc. To be acceptable, the baseline image also had to have a residual anterior segment retardation of 15 nm or less and an atypical scan score less than 23. The atypical scan score is a measure provided by the GDx VCC standard software, which indicates the presence of atypical patterns of retardation that can generate spurious RNFL thickness measurements. Twenty-four (13%) of 183 patients had unacceptable GDx VCC scans and were excluded from further analysis. The Venn diagram in Figure 1 shows the number of unacceptable images from each instrument.

The GDx VCC software calculates summary parameters based on quadrants that are defined as temporal (335° to 24°), superior (25° to 144°), nasal (145° to 214°), or inferior (215° to 334°). The GDx VCC parameters investigated in this study were superior ratio (superior quadrant thickness/temporal quadrant thickness), inferior ratio, superior/nasal ratio, superior maximum (mean of the 1500 thickest points in the superior quadrant), inferior maximum, superior average, inferior average, normalized superior area (area under the temporal-superior-nasal-inferior-temporal [TSNIT] curve in the superior quadrant), normalized inferior area, maximum modulation (thickness...
The commercially available optical coherence tomograph, the Stratus OCT, was used to assess parapapillary RNFL thickness measurements. Optical coherence tomography uses the principles of low-coherence interferometry and is analogous to ultrasound B-mode imaging, but it uses light instead of sound to acquire high-resolution images of ocular structures. A low-coherence near infrared (840-nm) light beam is directed onto a partially reflective mirror (beam splitter) that creates 2 light beams, a reference and a measurement beam. The measurement beam is directed onto the subject’s eye and is reflected from intraocular microstructures and tissues according to their distance, thickness, and reflectivity. The reference beam is reflected from the reference mirror at a known, variable position. Both beams travel back to the partially reflective mirror, recombine, and are transmitted to a photosensitive detector. The pattern of interference is used to provide information regarding distance and thickness of the retinal structures.
Bidimensional images are created by successive longitudinal scanning in transverse directions.

The fast RNFL algorithm was used to obtain RNFL thickness measurements with the Stratus OCT. Three images were acquired from each subject, with each image consisting of 256 A-scans along a 3.4-mm-diameter circular ring around the optic disc. A baseline image was automatically created using the Stratus OCT software. Quality assessment of Stratus OCT scans was determined by an experienced examiner masked to the subject's identity and the results of the other tests. Good-quality scans had to have focused images from the optic disc, an adequate signal-to-noise ratio, and the presence of a centered circular ring around the optic disc. Nineteen (10%) of 183 patients had unacceptable Stratus OCT scans and were excluded from further analysis (Figure 1).

Parapapillary RNFL thickness parameters automatically calculated by existing Stratus OCT software (version 3.1) and evaluated in this study were average thickness (360° measurement), temporal quadrant thickness (316° to 45°), superior quadrant thickness (46° to 135°), nasal quadrant thickness (136° to 225°), inferior quadrant thickness (226° to 315°), and thickness for each of 12 clock-hour positions with the 3-o’clock position as nasal, 6-o’clock position as inferior, 9-o’clock position as temporal, and 12-o’clock position as superior. Other parameters evaluated included superior maximum (Smax) (thickest point in the superior quadrant), inferior maximum (Imax) (thickest point in the inferior quadrant), and relative parameters such as Imax/Smax, Smax/Imax, Imax/Tavg, Smax/nasal average thickness (Smax/Navg), and difference between the thickest and thinnest points along the measurement circle (Max-Min). For each of these parameters, ROC curves were constructed and sensitivities at fixed specificities (≥80% and ≥95%) were reported.

For each parameter, the Stratus OCT software provides a classification (within normal limits, borderline, or outside normal limits) based on comparison with an internal normative database of 328 eyes. A parameter is classified as outside normal limits (outside normal limits) based on comparison with an internal normal database.

### Statistical Analysis

We used t tests to evaluate optic nerve head and RNFL measurement differences between glaucomatous and normal eyes. Results of statistical significance were also provided after Bonferroni correction based on the number of comparisons within each analysis (GDx VCC, HRT II, and Stratus OCT).

The ROC curves were used to describe the ability of each parameter from each instrument to differentiate glaucomatous from normal eyes. The ROC curve shows the trade-off between sensitivity and 1–specificity. An area under the ROC curve (AUC) of 1.0 represents perfect discrimination, whereas an AUC of 0.5 represents chance discrimination. The method of DeLong et al was used to compare AUCs. Sensitivities at fixed specificities were compared using the McNemar test.

Diagnostic categorization (outside normal limits, borderline, or within normal limits) provided by each instrument after comparison with its respective normative database was also evaluated, and LRs were reported. An LR is defined as the probability of a given test result in those with disease divided by the probability of that same test result in those without the disease. Once determined, an LR can be incorporated directly into the calculation of posttest probability of disease by using a formulation of the Bayes theorem. The LR for a given test result indicates how much that result will raise or lower the probability of disease. A value of 1 means that the test provides no additional information, and ratios higher or lower than 1 increase or decrease the likelihood of disease. A classification of the effect of LRs of different magnitudes on the posttest probability of disease has been suggested and was used in our study. According to this classification, LRs higher than 10 or lower than 0.1 would be associated with large effects on posttest probability, LRs from 5 to 10 or from 0.1 to 0.2 would be associated with moderate effects, LRs from 2 to 5 or from 0.2 to 0.5 would be associated with small effects, and LRs closer to 1 would be insignificant. The 95% CIs for LRs were calculated according to the method proposed by Simel et al.

Chance-corrected agreement on categorization between different instruments was assessed using a weighted k approach, with quadratic weighting assignment as proposed by Fleiss. This method allows for differences in the importance of disagreement, assuming that disagreement between adjacent categories (eg, between normal and borderline or borderline and outside normal limits) is not as important as that between distant categories (eg, between normal and outside normal limits). Strength of agreement was categorized according to the method proposed by Landis and Koch: less than 0 was poor, 0 to 0.20 was slight, 0.21 to 0.40 was fair, 0.41 to 0.60 was moderate, 0.61 to 0.80 was substantial, and 0.81 to 1.00 was almost perfect.

P < .05 was considered statistically significant. Statistical analyses were performed using SPSS version 10.0 (SPSS Inc, Chicago, Ill) and S-PLUS 2000 (Mathsoft Inc, Seattle, Wash) statistical software.

### RESULTS

After the exclusion of subjects with unacceptable images (Figure 1), the final study sample included 141 eyes of 141 subjects (75 patients with glaucoma and 66 healthy control subjects). There was no statistically significant difference between the mean ± SD ages of patients with glaucoma and healthy subjects (mean ± SD, 68 ± 10 years vs 65 ± 8 years, respectively; P = .06 using a t test). The mean ± SD mean deviation of the glaucomatous eyes on the visual field test nearest the imaging date was −4.87 ± 3.9 dB. According to the grading scale for severity of visual field defects developed by Hodapp et al, 53 patients (70%) were classified as having early visual field defects, 11 patients (15%) had moderate defects, and 11 patients (15%) had severe visual field defects.

#### Table 1 presents the mean values of GDx VCC parameters in glaucomatous and normal eyes. After Bonferroni correction (κ = .003; 15 comparisons), statistically significant differences were found for all parameters except symmetry. Table 1 also shows the ROC curve areas and sensitivities at fixed specificities. The 3 GDx VCC parameters with the largest AUCs were NFI (0.91), inferior normalized area (0.86), and TSNIT average (0.85). The AUC for NFI was significantly higher than those for the other parameters.

©2004 American Medical Association. All rights reserved.
Table 1. Values of GDx VCC Parameters With Areas Under the Receiver Operating Characteristic Curves and Sensitivities at Fixed Specificities *

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Glaucomatous (n = 75)</th>
<th>Normal (n = 66)</th>
<th>P Value</th>
<th>ROC (SE)</th>
<th>Sensitivity @95%</th>
<th>Sensitivity @80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFI</td>
<td>47.3 ± 21.2</td>
<td>19.8 ± 6.70</td>
<td>&lt;.001</td>
<td>0.91 (0.02)</td>
<td>61/97</td>
<td>87/80</td>
</tr>
<tr>
<td>Inferior normalized area</td>
<td>0.09 ± 0.03</td>
<td>0.13 ± 0.02</td>
<td>&lt;.001</td>
<td>0.86 (0.03)</td>
<td>49/95</td>
<td>72/80</td>
</tr>
<tr>
<td>TSNIT average</td>
<td>44.4 ± 7.21</td>
<td>53.0 ± 4.24</td>
<td>&lt;.001</td>
<td>0.85 (0.03)</td>
<td>59/95</td>
<td>72/80</td>
</tr>
<tr>
<td>Inferior average</td>
<td>48.4 ± 10.2</td>
<td>60.9 ± 7.42</td>
<td>&lt;.001</td>
<td>0.84 (0.03)</td>
<td>49/95</td>
<td>68/80</td>
</tr>
<tr>
<td>Superior average</td>
<td>52.3 ± 11.2</td>
<td>64.7 ± 6.58</td>
<td>&lt;.001</td>
<td>0.82 (0.04)</td>
<td>55/95</td>
<td>61/80</td>
</tr>
<tr>
<td>Superior normalized area</td>
<td>0.09 ± 0.03</td>
<td>0.13 ± 0.02</td>
<td>&lt;.001</td>
<td>0.82 (0.04)</td>
<td>53/95</td>
<td>68/80</td>
</tr>
<tr>
<td>TSNIT SD</td>
<td>15.3 ± 4.93</td>
<td>21.3 ± 4.08</td>
<td>&lt;.001</td>
<td>0.82 (0.03)</td>
<td>48/95</td>
<td>72/80</td>
</tr>
<tr>
<td>Superior maximum</td>
<td>61.1 ± 13.1</td>
<td>74.1 ± 7.71</td>
<td>&lt;.001</td>
<td>0.80 (0.04)</td>
<td>48/95</td>
<td>64/80</td>
</tr>
<tr>
<td>Superior ratio</td>
<td>2.23 ± 0.95</td>
<td>3.23 ± 0.98</td>
<td>&lt;.001</td>
<td>0.80 (0.04)</td>
<td>27/95</td>
<td>75/80</td>
</tr>
<tr>
<td>Inferior ratio</td>
<td>2.29 ± 0.95</td>
<td>3.38 ± 1.18</td>
<td>&lt;.001</td>
<td>0.79 (0.04)</td>
<td>36/95</td>
<td>61/80</td>
</tr>
<tr>
<td>Inferior maximum</td>
<td>62.8 ± 13.7</td>
<td>77.0 ± 11.0</td>
<td>&lt;.001</td>
<td>0.78 (0.04)</td>
<td>41/95</td>
<td>64/80</td>
</tr>
<tr>
<td>Maximum modulation</td>
<td>1.66 ± 0.94</td>
<td>2.56 ± 1.11</td>
<td>&lt;.001</td>
<td>0.76 (0.04)</td>
<td>23/95</td>
<td>60/80</td>
</tr>
<tr>
<td>Ellipse modulation</td>
<td>2.74 ± 1.23</td>
<td>4.00 ± 1.49</td>
<td>&lt;.001</td>
<td>0.75 (0.04)</td>
<td>25/95</td>
<td>59/80</td>
</tr>
<tr>
<td>Superior/nasal</td>
<td>2.07 ± 0.65</td>
<td>2.35 ± 0.62</td>
<td>&lt;.001</td>
<td>0.65 (0.05)</td>
<td>19/95</td>
<td>43/80</td>
</tr>
<tr>
<td>Symmetry</td>
<td>1.00 ± 0.24</td>
<td>0.97 ± 0.12</td>
<td>.38</td>
<td>0.50 (0.05)</td>
<td>9/95</td>
<td>25/80</td>
</tr>
</tbody>
</table>

Abbreviations: GDx VCC, GDx VCC scanning laser polarimeter with variable corneal compensation (Laser Diagnostic Technologies, Inc, San Diego, Calif); NFI, nerve fiber indicator; ROC, receiver operating characteristic; TSNIT, temporal-superior-nasal-inferior-temporal.

*Data are presented as mean ± SD unless otherwise indicated. P value is for comparison of mean values of GDx VCC parameters between glaucomatous and normal eyes using the t test.

Table 2. Likelihood Ratios and 95% Confidence Intervals for GDx VCC Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Within Normal Limits</th>
<th>Borderline</th>
<th>Outside Normal Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFI*</td>
<td>0.39 (0.29-0.53)</td>
<td>2.28 (0.86-6.08)</td>
<td>Infinity (NA)</td>
</tr>
<tr>
<td>TSNIT average</td>
<td>0.45 (0.31-0.64)</td>
<td>2.64 (0.85-8.21)</td>
<td>Infinity (NA)</td>
</tr>
<tr>
<td>Superior average</td>
<td>0.52 (0.38-0.73)</td>
<td>7.04 (4.28-11.6)</td>
<td>Infinity (NA)</td>
</tr>
<tr>
<td>Inferior average</td>
<td>0.56 (0.40-0.77)</td>
<td>2.93 (1.88-4.58)</td>
<td>Infinity (NA)</td>
</tr>
<tr>
<td>TSNIT SD</td>
<td>0.58 (0.42-0.80)</td>
<td>5.28 (3.51-7.95)</td>
<td>Infinity (NA)</td>
</tr>
<tr>
<td>Overall classification</td>
<td>0.24 (0.14-0.40)</td>
<td>1.60 (1.02-2.50)</td>
<td>Infinity (NA)</td>
</tr>
</tbody>
</table>

Abbreviation: NA, not applicable. Expansions of other abbreviations are provided in the footnote to Table 1.

*The NFI cut points as suggested by the manufacturer: 0 to 30, normal; 31 to 50, borderline; and 51 to 100, outside normal limits.

Table 3. Interval Likelihood Ratios and 95% Confidence Intervals for the Parameters With Highest Areas Under the Receiver Operating Characteristic Curves From the GDx VCC, Stratus OCT, and HRT II

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Interval LRs</th>
<th>Subjects in Each Category, No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDx VCC nerve fiber indicator</td>
<td>0 to 15: 0.04 (0.01-0.30)</td>
<td>22 (16)</td>
</tr>
<tr>
<td></td>
<td>&gt;15 to 25: 0.28 (0.15-0.50)</td>
<td>46 (32)</td>
</tr>
<tr>
<td></td>
<td>&gt;25 to 35: 2.09 (1.32-3.30)</td>
<td>27 (19)</td>
</tr>
<tr>
<td></td>
<td>&gt;35 to 50: 3.96 (2.05-7.65)</td>
<td>11 (8)</td>
</tr>
<tr>
<td></td>
<td>&gt;50: Infinity (NA)</td>
<td>35 (25)</td>
</tr>
<tr>
<td>HRT II Bathija function*</td>
<td>&lt;=1.0: 0.08 (0.02-0.34)</td>
<td>23 (16)</td>
</tr>
<tr>
<td></td>
<td>&gt;-1.0 to 0: 0.51 (0.30-0.85)</td>
<td>41 (29)</td>
</tr>
<tr>
<td></td>
<td>&gt;0 to 1.0: 0.93 (0.59-1.47)</td>
<td>37 (26)</td>
</tr>
<tr>
<td></td>
<td>&gt;1.0 to 2.0: 15.8 (9.90-25.30)</td>
<td>19 (14)</td>
</tr>
<tr>
<td></td>
<td>&gt;2.0: Infinity (NA)</td>
<td>21 (15)</td>
</tr>
<tr>
<td>Stratus OCT inferior thickness, µm</td>
<td>&lt;=70: Infinity (NA)</td>
<td>20 (14)</td>
</tr>
<tr>
<td></td>
<td>&gt;70 to 90: 21.1 (14.0-31.8)</td>
<td>25 (18)</td>
</tr>
<tr>
<td></td>
<td>&gt;90 to 110: 1.85 (1.20-2.86)</td>
<td>31 (22)</td>
</tr>
<tr>
<td></td>
<td>&gt;110 to 130: 0.24 (0.12-0.47)</td>
<td>38 (27)</td>
</tr>
<tr>
<td></td>
<td>&gt;130: 0.07 (0.02-0.28)</td>
<td>27 (19)</td>
</tr>
</tbody>
</table>

Abbreviations: GDx VCC, GDx VCC scanning laser polarimeter with variable corneal compensation (Laser Diagnostic Technologies, Inc, San Diego, Calif); HRT II, HRT II Heidelberg Retina Tomograph (Heidelberg Engineering, Dossenheim, Germany); LR, likelihood ratio; NA, not applicable; NFI, nerve fiber indicator; and Stratus OCT, Stratus OCT optical coherence tomograph (Carl Zeiss Meditec, Inc, Dublin, Calif).

*Bathija function refers to the linear discriminant function used by Bathija et al.*

Interval LRIs were also calculated for the parameter of NFI using arbitrarily selected cutoffs (Table 3); results of 0 to 15 and greater than 50 were associated with large effects on the posttest probability of disease, whereas the other test ranges were associated with small effects.

Table 2 presents LRIs with their 95% CIs for the GDx VCC parameters after comparison with the instrument’s normative database. For all parameters, outside normal limits results were associated with large effects on the posttest probability of disease. Borderline results were associated with small to moderate effects, whereas within-normal-limits results were associated with small effects on the posttest probability of disease. The LRIs of the overall GDx VCC classification were also evaluated. For this classification, an outside-normal-limits result was considered to be the presence of an NFI greater than 50 or any other parameter outside-normal-limits. The LR of an outside-normal-limits result in the GDx VCC overall classification was infinity. For a borderline result (NFI between 31 and 50 or any other parameter that was borderline), the LR was 1.60 (95% CI, 1.02-2.50). A within-normal-limits result (NFI ≤30 and all parameters within normal limits) had an LR of 0.24 (95% CI, 0.14-0.40).
II parameters by sector did not result in higher ROC curve areas and sensitivities at fixed specificities. The 3 HRT II parameters with the largest AUCs were the Bathija function (0.86), the Mikelberg function (0.83), and the vertical cup/disc ratio (0.83). There were no statistically significant differences in ROC curve areas for these parameters (P > .05 for all comparisons). The analysis of HRT II parameters by sector did not result in higher ROC curve areas, with the parameter of temporal inferior rim/disc area ratio having the largest ROC curve area (0.81).

### Table 5. Likelihood Ratios and 95% Confidence Intervals for HRT II Moorfields Regression Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Within Normal Limits</th>
<th>Borderline</th>
<th>Outside Normal Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>0.54 (0.39-0.75)</td>
<td>6.75 (4.5-10.2)</td>
<td>12.3 (7.25-20.9)</td>
</tr>
<tr>
<td>Temporal</td>
<td>0.82 (0.63-1.08)</td>
<td>4.10 (2.41-6.98)</td>
<td>Infinity (NA)</td>
</tr>
<tr>
<td>Temporal superior</td>
<td>0.59 (0.43-0.81)</td>
<td>Infinity (NA)</td>
<td>6.16 (3.82-10.5)</td>
</tr>
<tr>
<td>Temporal inferior</td>
<td>0.53 (0.39-0.75)</td>
<td>3.30 (1.98-5.51)</td>
<td>Infinity (NA)</td>
</tr>
<tr>
<td>Nasal</td>
<td>0.72 (0.54-0.96)</td>
<td>2.20 (1.18-4.11)</td>
<td>Infinity (NA)</td>
</tr>
<tr>
<td>Nasal superior</td>
<td>0.53 (0.38-0.74)</td>
<td>7.92 (4.95-12.7)</td>
<td>16.7 (10.6-26.4)</td>
</tr>
<tr>
<td>Nasal inferior</td>
<td>0.52 (0.37-0.73)</td>
<td>2.81 (1.71-4.63)</td>
<td>Infinity (NA)</td>
</tr>
<tr>
<td>Overall classification</td>
<td>0.35 (0.23-0.54)</td>
<td>0.88 (0.46-1.70)</td>
<td>19.4 (14.2-26.3)</td>
</tr>
</tbody>
</table>

Abbreviations: Expansions of abbreviations are provided in the footnote to Table 3. Data are presented as mean ± SD unless otherwise indicated. P value is for comparison of mean values of HRT II parameters between glaucomatous and normal eyes using the t test. *Bathija function refers to the linear discriminant function used by Bathija et al. ‡Mikelberg function refers to the linear discriminant function used by Mikelberg et al.

### HRT II RESULTS

Table 4 presents the mean values of HRT II parameters in glaucomatous and normal eyes. After Bonferroni correction (α = .003; 17 comparisons), statistically significant differences were found for all parameters except disc area and height variation contour. Table 4 also indicates ROC curve areas and sensitivities at fixed specificities. The 3 HRT II parameters with the largest AUCs were the Bathija function (0.86), the Mikelberg function (0.83), and the vertical cup/disc ratio (0.83). There were no statistically significant differences in ROC curve areas for these parameters (P > .05 for all comparisons). The analysis of HRT II parameters by sector did not result in higher ROC curve areas, with the parameter of temporal inferior rim/disc area ratio having the largest ROC curve area (0.81).

### Table 5 presents LRs with their 95% CIs for the HRT II Moorfields regression analysis. Global and sectorial results outside normal limits were generally associated with large effects on the posttest probability of disease. Borderline results were associated with small to moderate effects, whereas within normal limits results were associated with small effects. An outside normal limits result in the overall HRT II classification (ie, the Moorfields regression analysis classification) was associated with a large effect on the posttest probability of disease (LR=19.4), whereas borderline (LR=0.88) and within normal limits (LR=0.35) results were associated with small changes in the probability of disease.

We evaluated interval LRs for the HRT II parameter with the largest AUC, the Bathija function. Several cutoffs were arbitrarily created for this parameter, and the interval LRs are indicated in Table 3. Values for the Bathija function greater than 1.0 or smaller than –1.0 were associated with large effects on posttest probabilities of disease, whereas the other test results had small effects on the probability of disease.

### STRATUS OCT RESULTS

Table 6 presents the mean values of Stratus OCT parameters in glaucomatous and normal eyes. After Bonferroni correction (α = .002; 25 comparisons), statistically significant differences were found for all parameters except thickness at 8 o'clock, thickness at 9 o'clock, temporal thickness, Imax/Smax, Smax/Tavg, and Smax/Navg. Table 6 also indicates ROC curve areas and sensitivities at fixed specificities. The 3 Stratus OCT parameters with the largest AUCs were inferior thickness (0.92), average thickness (0.91), and Imax (0.91). There were no statistically significant differences in ROC areas and sensitivities at fixed specificities.
No statistically significant difference was found between AUCCs for the best parameters from the GDx VCC (NFI, AUC=0.91), Stratus OCT (inferior thickness, AUC=0.92), and HRT II (Bathija function, AUC=0.86) (P > .05 for all comparisons). Figure 2 shows the ROC curves for the best parameters from each instrument.

At specificities of at least 95%, no statistically significant differences were found among parameters with the highest sensitivities from each instrument: GDx VCC NFI (sensitivity, 61%), Stratus OCT average thickness (sensitivity, 71%), and HRT II Bathija function (sensitivity, 59%; P > .05 for all comparisons). At specificities of at least 80%, a statistically significant difference was found between the HRT II parameter with the highest sensitivity (Mikelberg function; sensitivity, 73%) and those from the GDx VCC (NFI; sensitivity, 87%; P = .05) and Stratus OCT (inferior thickness; sensitivity, 89%; P = .01).

No statistically significant difference was found between parameters with the highest sensitivities from the GDx VCC and Stratus OCT.

Agreement on diagnostic categorization between pairs of instruments was also evaluated. For this analysis, the overall classification of each instrument was used as described previously. The GDx VCC and Stratus OCT overall classifications agreed in 89% of cases, with a substantial chance-corrected agreement (κ = 0.72 [0.8]). The GDx VCC and HRT II overall classifications agreed in 81% of cases, with a moderate chance-corrected agreement (κ = 0.50 [0.8]). The Stratus OCT and HRT II overall classifications agreed in 81% of cases, with a moderate chance-corrected agreement (κ = 0.55 [0.08]).
COMMENT

To our knowledge, this is the first study to provide a comparison, using the same population, of the diagnostic accuracies of 3 instruments: the GDx VCC scanning laser polarimeter, HRT II confocal scanning laser ophthalmoscope, and Stratus OCT optical coherence tomograph. Each instrument represents the current commercially available version of a different technology for evaluation of the optic nerve head and RNFL in glaucoma.

Several measures of diagnostic accuracy were provided in our study, including ROC curve areas, sensitivities at fixed specificities, and LRs. For parameters reported as continuous variables, no statistically significant differences in ROC curves were found among the best parameters of the 3 instruments.

Previous studies have compared ROC curve areas derived from measures obtained using older versions of these technologies. Zangwill et al\(^1\) and Greaney et al\(^4\) reported that ROC curve areas were similar among the best parameters from the GDx Nerve Fiber Analyzer (Laser Diagnostic Technologies, Inc), OCT 2000 (Carl Zeiss Meditec, Inc), and HRT I (Heidelberg Engineering). In the study by Zangwill and colleagues, the best parameters from the OCT 2000 and HRT I had higher sensitivities than the best parameter from the GDx Nerve Fiber Analyzer. At 96% specificity, the best parameter from the GDx Nerve Fiber Analyzer (a linear discriminant function combining several parameters of the instrument) had a sensitivity of only 32%. The introduction of the GDx with variable corneal compensation, the GDx VCC, has reportedly resulted in improved diagnostic accuracy compared with scanning laser polarimetry using fixed corneal compensation.\(^{26}\) In our study, at a specificity of 97%, the sensitivity of the best parameter from the GDx VCC, NFI, was 61%. This confirms the improvement in diagnostic accuracy as described in other studies.\(^{26,44}\)
For optical coherence tomography, the ROC curve areas obtained in our study for the Stratus OCT were similar to those obtained with the previous versions of this technology. The AUCs for the earlier versions of the optical coherence tomographic reportedly ranged from 0.79 to 0.94 depending on the parameter and characteristics of the population evaluated. In studies evaluating the diagnostic ability of several optical coherence tomographic parameters, the RNFL thickness in the inferior region often had the best ability to discriminate healthy eyes from eyes with early to moderate glaucoma, with sensitivities between 67% and 79% for specificities of 90% or higher. In our study, the parameter of inferior thickness also had the highest AUC, with a sensitivity of 64% for a specificity set at 95%. Although results for the Stratus OCT in our study were at least as good as those reported for the previous version of the instrument, the Stratus OCT may still have advantages compared with its predecessor, including increased sampling, a reduced need for pupillary dilation, and easier image acquisition.

For the HRT II, the ROC curve areas for parameters reported as continuous variables were also similar to those reported for previous versions of the instrument. The incorporation of the Moorfields regression analysis has been demonstrated to improve the diagnostic accuracy of this instrument. In our study, ROC curve areas for parameters obtained from the HRT II Moorfields regression analysis were not provided because the small number of categories in these parameters may cause an underestimation of the ROC curve area. At high specificity (≥95%), the sensitivity of the Moorfields regression analysis overall classification (59%) was not statistically significantly different from the best parameters of the Stratus OCT and GDx VCC. Furthermore, the application of sophisticated methods of analysis of HRT data has recently been reported to improve the diagnostic ability of this instrument. Additional studies are necessary to compare these methods with measures provided by the Stratus OCT and GDx VCC.

Although sensitivity and specificity have commonly been reported as measures of diagnostic accuracy in medical studies, their clinical utility is limited. They reflect the probability that a particular test result is positive or negative given the presence (sensitivity) or absence (specificity) of disease. However, there is an inversion of clinical logic in this definition; knowledge of whether the patient had the disease would clearly obviate the need for a diagnostic test. Similarly, the AUC is important for comparing the diagnostic accuracy of different tests but has little intrinsic clinical meaning. The starting points of any diagnostic process are the patient developing a constellation of symptoms and signs and the clinician integrating this information to assign a pretest probability of disease. The results of diagnostic tests are then used to modify the pretest probability of disease, yielding a new posttest probability. The direction and magnitude of this change from pretest to posttest are determined by the test’s properties, particularly the LR. The LR represents the magnitude of change from a physician’s initial suspicion of disease (pretest probability) to the likelihood of disease after the test result (posttest probability).

In our study, LRs for outside-normal-limits results in all instruments were generally associated with large changes from pretest to posttest probability of glaucoma. However, LRs for within-normal-limits results were associated with small changes in probability. For the overall classifications from each instrument, the LRs of within-normal-limits results were 0.24, 0.28, and 0.35, respectively, for the GDx VCC, Stratus OCT, and HRT II. This indicates that normal results for each of these tests would induce only a small change in the pretest probability of disease; that is, they would be of limited value in excluding the presence of disease. For borderline results, the LRs were generally associated with small to moderate changes in the probability of disease. Depending on the pretest probability of disease and the clinical situation in which the test is used, even small changes in probability may be clinically relevant.

Previous studies using the HRT I have reported the results of Moorfields regression analysis in terms of sensitivity and specificity. To do so, the borderline test results had to be forced into the within-normal-limits or outside-normal-limits categories. This approach results in valuable loss of information and may cause distortions in interpretation of the test results when used in clinical practice. Combining borderline results with the within-normal-limits category reduces the sensitivity of the test and the importance of a within-normal-limits result, whereas combining borderline results with the outside-normal-limits category reduces the specificity of the test and the importance of an outside-normal-limits result. In contrast, LRs can be calculated for each diagnostic categorization, permitting the clinician to assess the diagnostic importance of each category. In this study, we showed that borderline results in the Moorfields regression analysis overall classification did not appreciably change the probability of disease but that an outside-normal-limits or within-normal-limits result produced a larger change in its probability.

The dichotomization of test results with continuous outcomes may also result in loss of information because results that are markedly abnormal are lumped with results that are only mildly abnormal, leading to distortion in their clinical interpretation. These distortions are especially exaggerated when the patient’s test result is close to the established cutoff. Interval LRs, however, assign a specific value to each level of abnormality, and this value can be used to calculate the posttest probability of disease for a given level of the test. Interval LRs calculated for parameters from each instrument are likely to provide more clinically relevant information than what is currently available in the printout. For instance, according to the manufacturer’s suggested cutoffs for the GDx VCC NFI, values from 0 to 30 should be considered normal. This range of values was associated with an LR of 0.38, inducing only a small change toward reduction in the probability of having glaucomatous visual field loss. Using the manufacturer’s suggested cutoffs, a test result with an NFI of 10 would be considered similar to one with an NFI of 27. Our results demonstrate the great difference between these 2 situations. Based on the in-
interval LRs calculated in our study, an NFI of 10 would almost exclude the presence of glaucomatous visual field loss, whereas a result of 27 would have a nearly insignificant effect on changing the pretest probability of disease. A similar analogy could be demonstrated for the Stratus OCT and HRT II results.

The usefulness of a diagnostic test is strongly influenced by the proportion of patients suspected of having the target disorder whose test results have very high (>10) or very low (<0.1) LRs, thus having a huge effect on the probability of disease. As indicated in Table 3, this proportion was 41% for the GDx VCC-NFI, 51% for Stratus OCT inferior thickness, and 45% for the HRT II Bathija function. For the HRT II Moorfields regression analysis, this proportion ranged from 1% to 33% depending on the specific parameter chosen. Selection of other cutoffs or parameters may produce different results, and studies with larger sample sizes are necessary to provide more robust estimations of LRs using smaller intervals of the range of possible test values. In our study, test results for some parameters had LRs of infinity, which indicates that a particular test result was not found in any of the healthy subjects (ie, the probability of the test result in healthy subjects was 0). When evaluating the clinical importance of such parameters, it is critical to evaluate the probability of the same test result in subjects with disease.

Agreement among the different instruments in our study varied from moderate to substantial. The chance-corrected agreement between the Stratus OCT and HRT II was 0.55, similar to the 0.58 index reported by Greaney et al when comparing earlier versions of these instruments. For the GDx VCC compared with the Stratus OCT and the GDx VCC compared with the HRT II, the agreements reported in our study were higher than those using earlier versions of these instruments. These findings are similar to those in recent reports showing an improvement in the correlation coefficients for associations between scanning laser polarimetry parameters and other measures of glaucomatous damage, such as RNFL semiquantitative photographic scores or OCT RNFL thickness measurements, when variable corneal compensation rather than fixed corneal compensation was used. Interestingly, the agreement of the GDx VCC with the Stratus OCT was higher than that between the GDx VCC and HRT II and between the Stratus OCT and HRT II, which most likely reflects the fact that both the GDx VCC and Stratus OCT measure RNFL properties, whereas the HRT II measures optic disc topography and provides only an indirect measure of the RNFL.

In conclusion, the AUCs and the sensitivities at high specificities were similar among the best parameters from each instrument. Abnormal results (as compared with each instrument’s normative database) were associated with high LRs and large effects on posttest probabilities of having glaucomatous visual field loss. The calculation of interval LRs may provide additional information to assist the clinician in diagnosing glaucoma.

Submitted for publication January 26, 2004; final revision received March 1, 2004; accepted March 4, 2004.

This study was supported in part by the Foundation for Eye Research, Rancho Santa Fe, Calif (Dr Medeiros), and by grant EY11008 from the National Institutes of Health, Bethesda, Md (Dr Zangwill).

Corresponding author: Robert N. Weinreb, MD, Hamilton Glaucoma Center, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093-0946.

REFERENCES


