Diagnostic Ability of the Heidelberg Retina Tomograph 3 for Glaucoma

ANTONIO FERRERAS, LUISS E. PABLO, ANA B. PAJARIN, JOSÉ M. LARROSA, VICENTE POLO, AND VICTORIA PUEYO

- PURPOSE: To compare the diagnostic ability of the Heidelberg Retina Tomograph 3 (HRT3) and the Heidelberg Retina Tomograph 2 (HRT2) for discriminating between healthy eyes and eyes with glaucomatous visual field loss.
- DESIGN: Retrospective cross-sectional study.
- METHODS: Participants were 93 healthy subjects and 90 patients with open-angle glaucoma. All participants underwent imaging of the optic nerve head with the HRT2. Afterward, HRT data also were analyzed using version 3 of the software without modifying the optic disk contour line. The receiver operating characteristic (ROC) curves between normal and glaucomatous subjects were plotted for the global stereometric parameters of both software versions. Moorfields regression analysis (MRA) and glaucoma probability score (GPS) diagnostic abilities also were compared.
- RESULTS: The parameters with the largest areas under the ROC curve were the Frederick S. Mikelberg (FSM) discriminant function for the HRT3 (0.948) and the vertical cup-to-disk ratio (0.914) for the HRT2. At a fixed specificity of 95%, the best sensitivity was 74.4% for the Reinhard O.W. Burk (RB) discriminant function of the HRT2 and 83.3% for the FSM discriminant function of the HRT3. The best sensitivity and specificity pairs for the HRT classifications were 85.5% and 76.3%, respectively, for overall MRA2, 84.4% and 83.8%, respectively, for overall MRA3, 93.3% and 58.0%, respectively, for the global color-coded GPS, and 84.4% and 74.1%, respectively, for the global GPS numerical value.
- CONCLUSIONS: At 95% fixed specificity, most HRT3 parameters exhibited at least the same sensitivity for glaucoma diagnosis as the analogous parameters for the HRT2. The diagnostic ability overall of MRA3 was similar to that of the previous version. GPS exhibited higher sensitivity and somewhat lower specificity than the MRA. (Am J Ophthalmol 2008;145:354–359. © 2008 by Elsevier Inc. All rights reserved.)

Accepted for publication Sep 26, 2007.
From the Department of Ophthalmology, Miguel Servet University Hospital, Zaragoza, Spain (A.F., L.E.P., J.M.L., V.Po., V.Pu.); and the Department of Family Medicine, Euroresidencia, Zaragoza, Spain (A.B.P.).
Inquiries to Antonio Ferreras, Department of Ophthalmology, Miguel Servet University Hospital, Isabel la Católica 1-3, 50-009 Zaragoza, Spain; e-mail: aferreras@msn.com

T he detection of early structural changes at the optic disk is key for the diagnosis and follow-up of glaucoma. For decades, stereophotographs and planimetry have been used together to evaluate optic disk morphologic features qualitatively. However, the highly subjective nature of this method and the need for experienced evaluators limit its general applicability. In the last few years, various instruments have been introduced to measure optic disk parameters quantitatively, one of which is the Heidelberg Retina Tomograph (HRT; Heidelberg Engineering, Heidelberg, Germany), which provides rapid, quantitative, and reproducible measurements of the optic nerve head.1–13 Recently, a new version of the software for the HRT was introduced. This new version, called Advanced Glaucoma Analysis 3.0 (HRT3), is an enhanced version of the previous HRT2 software. The HRT3 provides larger, ethnic-selectable normative databases and includes new data analysis tools such as the glaucoma probability score (GPS) classification.14–17

In general, up to a 4% difference can be expected in the data obtained with the HRT2 compared with that obtained with the HRT3. A horizontal scaling error was reported by the manufacturer, in which horizontal HRT2 stereometric measurements were enlarged by 4%; this was corrected in the HRT3.14 The new version 3.0 software for the HRT glaucoma module has enhanced alignment algorithms and software analysis that can be applied to past examinations.

The purpose of the present study was to compare both software versions to determine if the changes applied to the latest software version increase the diagnostic ability of the HRT. Recently, Shunmugan and Azuara-Blanco evaluated the quality of the reporting of diagnostic accuracy using the HRT.18 They concluded that the quality was suboptimal and suggested the Standards for Reporting of Diagnostic Accuracy (STARD) initiative for evaluating the strengths and weaknesses of diagnostic ability studies.19 Thus, the design of our study followed the 25 items of the STARD guidelines, which are aimed at increasing the quality of the reporting of diagnostic accuracy research.

METHODS

- SUBJECTS AND MEASUREMENT PROTOCOL: The study protocol was approved by the ethics committee of Miguel Servet University Hospital, and informed written consent was obtained from all participants.
Subjects with normal eyes were recruited from among patients referred for refraction who underwent routine examination without abnormal ocular findings, hospital staff, and relatives of patients in our hospital. Patients with glaucoma were recruited from an ongoing longitudinal follow-up study at the Miguel Server University Hospital.

In total, 93 consecutive healthy subjects and 90 consecutive White patients with open-angle glaucoma who underwent imaging of the optic disk with the HRT2 from September 1, 2005 through April 30, 2007 were included in the statistical analysis. One eye from each subject was chosen randomly for the study, unless only one eye met the inclusion criteria.

Inclusion criteria were best-corrected visual acuity of 20/40 or better, refractive error within ± 5.00 diopters equivalent sphere and ± 2.00 diopters of astigmatism, and open anterior chamber angle. Exclusion criteria were previous intraocular surgery, lens opacity (nuclear color or opalescence, cortical or posterior subcapsular lens opacity < 1) according to the Lens Opacities Classification System III system, diabetes or other diseases affecting the visual field, history of ocular or neurologic disease, or current use of a medication that could affect visual field sensitivity.

All subjects underwent a full ophthalmologic examination: clinical history, visual acuity, biomicroscopy of the anterior segment using a slit-lamp, gonioscopy, Goldmann applanation tonometry, central corneal ultrasonic pachymetry (model DGH 500; DGH Technology, Exton, Pennsylvania, USA), and ophthalmoscopy of the posterior segment. All subjects underwent at least two reliable standard automated perimetry (SAP) tests. SAPs were performed using a Humphrey Field analyzer, model 750 (Zeiss Humphrey Systems, Dublin, California, USA), with the Swedish interactive threshold algorithm standard 24-2 algorithm. Reliability criteria were fixation losses and false-positive or false-negative rates of less than 20%. The second reliable perimetry test obtained was used in this study to minimize the impact of the learning effect. Abnormal SAP results were considered to be a reproducible glaucomatous visual field loss in the absence of other abnormalities to explain the defect. Visual field loss was defined as the presence of a cluster of three points with a P value of less than 0.05 or a cluster of two points with a P value of less than 0.01 on the pattern deviation plot, a pattern standard deviation significantly greater than the 5% level, a glaucoma hemifield test result outside normal limits, or a combination thereof. The subjects completed the perimetry tests before any clinical examination or structural test. Each perimetry test was performed on different days to avoid a fatigue effect.

Topographic analysis of the optic nerve head was performed using a confocal scanning laser ophthalmoscope, the HRT2, which uses a diode laser of 670-nm wavelength. The HRT provides topographic measurements of the optic nerve head derived from 16 to 64 optical sections to a depth of 4 mm, depending on the longitudinal field of view. The spherical equivalent refractive error of each eye was adjusted in the dioptric ring of the HRT. After keratometric readings were entered into the software program (to correct for magnification errors), topographic images were acquired through pupils dilated with 1% tropicamide (Alcon Laboratories, Inc, Fort Worth, Texas, USA). All scans had to have an interscan standard deviation of less than 30 μm. The margin of the optic disks was traced manually by the same glaucoma specialist, who was masked to patient identity and clinical history, defining the inner edge of the Elschnig ring with at least a four-point contour line. Scans were analyzed using first the HRT2 software and, afterward, the Advanced Glaucoma Analysis 3.0 software. The same optic disk margin was used for both analyses. All the ocular examinations, perimetry tests, and the topographic analyses were performed within two months.

- **CLASSIFICATION INTO GROUPS**: Healthy eyes had an intraocular pressure (IOP) of less than 21 mm Hg, no history of increased IOP, and normal SAP results, regardless of the appearance of the optic disk. The glaucoma group comprised subjects with primary open-angle glaucoma, pseudoexfoliative glaucoma, and pigmented glaucoma. Glaucomatous eyes had an IOP of more than 21 mm Hg (on at least three readings on different days) and altered SAP, regardless of the appearance of the optic disk.

- **STATISTICAL ANALYSIS**: All statistical analyses were calculated using SPSS software version 15.0 (SPSS, Inc, Chicago, Illinois, USA) and MedCalc version 9.2.1.0 (MedCalc Software, Mariakerke, Ghent, Belgium) statistical software.

The receiver operating characteristic (ROC) curves were plotted between normal and glaucomatous eyes for all global stereometric parameters of both HRT2 and HRT3. The ROC curve has a trade-off between sensitivity and 1-specificity (false-positive rate). An area under the ROC curve (AUC) of 1.0 represents perfect discrimination, whereas an AUC of 0.5 represents chance discrimination. ROC curves were used to assess the usefulness of each parameter for differentiating glaucomatous eyes from healthy eyes.

Sensitivities at 95% (5% false-positive rate) fixed specificity were calculated for all global stereometric parameters. The diagnostic ability of the color-coded Moorfields regression analysis (MRA) results, color-coded GPS classification, and GPS numerical values also were calculated. The cut-off points were defined by the MedCalc software as the points with the best sensitivity and specificity balance. Positive and negative likelihood ratios (LRs) were calculated for these classifications. A positive LR is the ratio between the probability of a positive test result given the presence of the disease and the probability of a positive test result given the absence of the disease (sensitivity/[1 − specificity]) or true positive rate/false positive rate. The LR for a given test result indicates how much that result increases or decreases the posttest probability of disease. A negative LR is the ratio between the probability of a negative test result given the presence of the disease and...
the probability of a negative test result given the absence of the disease (false-negative rate/true-negative rate or \(1 - \text{sensitivity}/\text{specificity}\)). An LR value of 1 indicates that the test provides no additional information, but ratios higher or lower than 1 indicate an increased or decreased posttest probability of disease, respectively.

### RESULTS

The clinical characteristics of both groups included in the study are shown in Table 1. The normal group consisted of 93 eyes of 93 subjects. The mean age was 56.4 ± 9.8 years. The glaucoma group included 90 eyes.
of 90 patients: 71 with primary open-angle glaucoma, 15 with pseudoexfoliative glaucoma, and four with pigmentary glaucoma. The mean age (± standard deviation) was 60.4 ± 9.1 years. Neither age (P = .103) nor central corneal thickness (P = .386) differed significantly between the groups, as determined using the Student t test.

The HRT2 parameters with the greatest AUCs (Table 2) were vertical cup-to-disk ratio (0.914), cup-to-disk area ratio (0.906), and rim-to-disk area ratio (0.906). At a fixed specificity of 95%, the best sensitivity value was 74.4% for the RB57 discriminant function and 73.3% for the vertical cup-to-disk ratio of the HRT2. The HRT3 parameters with the largest AUCs (Table 3) were the FSM3 discriminant function (0.948), the cup-to-disk area ratio (0.941), the rim-to-disk area ratio (0.941), and the vertical cup-to-disk ratio (0.940). At a fixed specificity of 95%, the best sensitivity value was 83.3% for FSM and 82.2% for the vertical cup-to-disk ratio of the HRT3.

At 95% fixed specificity, rim volume, horizontal cup-to-disk ratio, vertical cup-to-disk ratio, contour line modulation (CLM) temporal superior, and FSM discriminant function of the HRT3 yielded higher sensitivity than the analogous parameters for the HRT2. CLM temporal inferior had better sensitivity at 95% specificity for the HRT2.

The remaining HRT parameters had similar diagnostic abilities for both software versions.

Sensitivity and specificity pairs were 85.5% and 76.3%, respectively, for overall MRA2, 84.4% and 83.8%, respectively, for overall MRA3, and 77.7% and 87.0%, respectively, for global MRA3. Temporal superior, temporal inferior, and nasal MRA3 classifications had higher sensitivity values (75% vs 55%) with similar specificities (88% and 95%) than the equivalent MRA2 classifications. Temporal, nasal superior, and nasal inferior MRA classifications exhibited similar sensitivity and specificity balance for both HRT software versions.

Sensitivity and specificity values were 74.4% and 84.9%, respectively, for overall color-coded GPS, 93.3% and 58.0%, respectively, for global color-coded GPS, and 84.4% and 74.1%, respectively, for global GPS probability value (at a cut-off point of > 47), respectively. GPS sectoral classifications yielded higher sensitivity values than MRA2 and MRA3, but lower specificity values. The cut-off point with the best sensitivity and specificity balance for the color-coded classifications was borderline, except for the HRT3 MRA final and GPS final.

Temporal superior MRA2, temporal inferior MRA2, and temporal superior MRA3 classifications had the high-

### Table 3. Areas Under the Receiver Operating Characteristic Curve for Each Global Topographic Parameter Obtained with the Heidelberg Retina Tomograph 3

<table>
<thead>
<tr>
<th>HRT3 Global Topographic Parameters</th>
<th>AUC</th>
<th>SD</th>
<th>95% CI</th>
<th>P value</th>
<th>Sensitivity at 95% Fixed Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk area</td>
<td>0.649</td>
<td>0.069</td>
<td>0.523 to 0.795</td>
<td>.031</td>
<td>18.8%</td>
</tr>
<tr>
<td>Cup area</td>
<td>0.916</td>
<td>0.036</td>
<td>0.840 to 0.982</td>
<td>&lt;.001</td>
<td>88.9%</td>
</tr>
<tr>
<td>Rim area</td>
<td>0.889</td>
<td>0.046</td>
<td>0.774 to 0.956</td>
<td>&lt;.001</td>
<td>48.3%</td>
</tr>
<tr>
<td>Cup-to-disk area ratio</td>
<td>0.941</td>
<td>0.032</td>
<td>0.866 to 0.992</td>
<td>&lt;.001</td>
<td>65.5%</td>
</tr>
<tr>
<td>Rim-to-disk area ratio</td>
<td>0.941</td>
<td>0.032</td>
<td>0.866 to 0.992</td>
<td>&lt;.001</td>
<td>65.5%</td>
</tr>
<tr>
<td>Cup volume</td>
<td>0.845</td>
<td>0.050</td>
<td>0.738 to 0.933</td>
<td>&lt;.001</td>
<td>38.8%</td>
</tr>
<tr>
<td>Rim volume</td>
<td>0.904</td>
<td>0.043</td>
<td>0.795 to 0.966</td>
<td>&lt;.001</td>
<td>57.7%</td>
</tr>
<tr>
<td>Mean cup depth</td>
<td>0.738</td>
<td>0.063</td>
<td>0.605 to 0.853</td>
<td>.002</td>
<td>28.8%</td>
</tr>
<tr>
<td>Maximum cup depth</td>
<td>0.558</td>
<td>0.074</td>
<td>0.410 to 0.700</td>
<td>.456</td>
<td>3.3%</td>
</tr>
<tr>
<td>Height variation contour</td>
<td>0.656</td>
<td>0.071</td>
<td>0.501 to 0.780</td>
<td>.057</td>
<td>11.1%</td>
</tr>
<tr>
<td>Cup shape measure</td>
<td>0.896</td>
<td>0.040</td>
<td>0.810 to 0.966</td>
<td>&lt;.001</td>
<td>58.8%</td>
</tr>
<tr>
<td>Mean RNFL thickness</td>
<td>0.851</td>
<td>0.051</td>
<td>0.731 to 0.932</td>
<td>&lt;.001</td>
<td>37.7%</td>
</tr>
<tr>
<td>RNFL cross-sectional area</td>
<td>0.834</td>
<td>0.057</td>
<td>0.697 to 0.919</td>
<td>&lt;.001</td>
<td>28.8%</td>
</tr>
<tr>
<td>Horizontal cup-to-disk ratio</td>
<td>0.842</td>
<td>0.051</td>
<td>0.731 to 0.931</td>
<td>&lt;.001</td>
<td>35.5%</td>
</tr>
<tr>
<td>Vertical cup-to-disk ratio</td>
<td>0.940</td>
<td>0.030</td>
<td>0.883 to 1.000</td>
<td>&lt;.001</td>
<td>82.2%</td>
</tr>
<tr>
<td>Maximum contour elevation</td>
<td>0.813</td>
<td>0.056</td>
<td>0.707 to 0.925</td>
<td>&lt;.001</td>
<td>45.5%</td>
</tr>
<tr>
<td>Maximum contour depression</td>
<td>0.662</td>
<td>0.068</td>
<td>0.540 to 0.807</td>
<td>.019</td>
<td>31.1%</td>
</tr>
<tr>
<td>CLM temporal superior</td>
<td>0.820</td>
<td>0.057</td>
<td>0.684 to 0.907</td>
<td>&lt;.001</td>
<td>36.6%</td>
</tr>
<tr>
<td>CLM temporal inferior</td>
<td>0.734</td>
<td>0.064</td>
<td>0.606 to 0.858</td>
<td>.002</td>
<td>45.5%</td>
</tr>
<tr>
<td>Average variability (SD)</td>
<td>0.566</td>
<td>0.074</td>
<td>0.411 to 0.699</td>
<td>.456</td>
<td>13.3%</td>
</tr>
<tr>
<td>Reference height</td>
<td>0.516</td>
<td>0.075</td>
<td>0.353 to 0.649</td>
<td>.989</td>
<td>3.3%</td>
</tr>
<tr>
<td>FSM discriminant function value</td>
<td>0.948</td>
<td>0.036</td>
<td>0.854 to 0.996</td>
<td>&lt;.001</td>
<td>83.3%</td>
</tr>
<tr>
<td>RB discriminant function value</td>
<td>0.927</td>
<td>0.036</td>
<td>0.838 to 0.981</td>
<td>&lt;.001</td>
<td>71.1%</td>
</tr>
<tr>
<td>Linear cup-to-disk ratio</td>
<td>0.928</td>
<td>0.035</td>
<td>0.832 to 0.978</td>
<td>&lt;.001</td>
<td>66.6%</td>
</tr>
</tbody>
</table>

AUC = area under the receiver operating characteristic curve; CI = confidence interval; CLM = contour line modulation; FSM = Frederick S. Mikelberg; RB = Reinhard O.W. Burk; HRT = Heidelberg Retina Tomograph; RNFL = retinal nerve fiber layer; SD = standard deviation.
est positive LRs (23.7, 16.1, and 20.5, respectively), whereas temporal inferior, nasal superior, and nasal inferior color-coded GPS classifications had the lowest negative LRs (0.10 for all).

**DISCUSSION**

**ASSESSMENT OF OPTIC DISK MORPHOLOGIC FEATURES AND the identification of optic disk changes are key for the diagnosis and follow-up of glaucoma. Many studies report the ability of the HRT to detect glaucomatous defects at the optic nerve head and its advantages for obtaining numerical values that can be analyzed and compared easily.**

In the present study, disk areas were similar between the normal and glaucoma groups. This is a critical factor when comparing groups because several HRT parameters, such as rim area, cup area, rim volume, etc., are directly related to disk size. MRA and GPS are also disk size dependent and tend to classify small and large optic disks with lower sensitivity and greater specificity than medium disks. Clinicians must take disk size into account when interpreting HRT outcomes.

Most global stereometric HRT3 parameters showed similar sensitivity figures at 95% specificity than the equivalent HRT2 parameters. The best global stereometric HRT parameters to discriminate between healthy and glaucomatous eyes were the same for both software versions: FSM and RB discriminant functions, cup-to-disk area ratio, rim-to-disk area ratio, vertical cup-to-disk ratio, and cup shape measure; these results are consistent with previous studies using the HRT2. Nevertheless, the diagnostic performance of rim volume, horizontal cup-to-disk ratio, vertical cup-to-disk ratio, CLM temporal superior, and FSM discriminant function is improved in the new HRT version. This is mainly because of enhanced alignment algorithms for the HRT3, which can correct the horizontal scaling error detected in the HRT2.

MRA is a linear regression that compares a subject’s rim area with the predicted rim area for a given disk area and age. Its results are based on confidence limits of the regression analysis derived from a normative database. Both software versions include the MRA, but the HRT3 includes a new ethnic-selectable database that comprises a greater range of disk area sizes. In general, MRA3 had better sensitivity than MRA2 with similar specificity. The increased diagnostic ability of MRA3 with respect to MRA2 may be because of the wider disk area range provided by the new HRT3 database. Burgansky-Eliash and associates compared the overall MRA2 and MRA3 diagnostic abilities by calculating ROC curves and did not find differences between classifications. Their results must be interpreted with caution, however, because analyses and comparisons of the AUCs are not suitable for variables with such low test result outcomes (within normal limits, borderline, and outside normal limits). Results that are clearly abnormal are included in the same cut-off point as results that are only mildly abnormal, leading to a loss of information. This effect is greater when the subject’s test result is closer to the cut-off point. In our study, ROC curves were plotted only for the stereometric parameters, because the MRA is not a continuous variable. The small number of categories in this case may lead to an underestimation of the ROC curve area.

The GPS provides a disease probability value based on the 3-dimensional shape of the optic disk and peripapillary retinal nerve fiber layer. The GPS is an approach to optic disk analysis that eliminates operator-dependent factors, which are considered to be important sources of variability. This classification represents the likelihood of glaucoma and not the level of damage; thus, higher GPS values do not necessarily indicate more advanced disease. At a borderline cut-off point, sensitivity was higher than 90% for the color-coded GPS classifications and specificity was approximately 60%. GPS probability values exhibited a better sensitivity and specificity balance than color-coded GPS classifications. The sensitivity was slightly lower, but the specificity increased up to 85% for the numerical GPS classification. The diagnostic performances of GPS and MRA3 were similar, confirming the results of previous studies.

Temporal superior MRA2, temporal inferior MRA2, and temporal superior MRA3 had the highest positive LRs (more than 15), indicating that abnormal results would be associated with important posttest effects for these variables. The color-coded GPS variables, however, yielded the lowest negative LRs and had the highest ability to exclude the presence of glaucoma.

The quality of the data obtained by the imaging devices is influenced by media opacity, retinal pigment epithelium status, instrument variability, and positioning and centering of the images. Also, the severity of visual field loss has an important influence on imaging instrument sensitivity. More severe disease is associated with increased sensitivity. These factors must be taken into account in clinical practice. Another limitation is that by ignoring optic disk appearance as a classification criterion, we might have included preperimetric glaucoma subjects in the normal group, leading to an underestimation of the HRT diagnostic accuracy. Diagnostic ability of imaging technologies, including the HRT, should be tested in population-based settings.

For most parameters and classifications, the diagnostic ability of the HRT3 was at least as good as that of the HRT2. Further studies are needed to demonstrate the promising clinical applicability of the ethnic-selectable normative database of the HRT3.
REFERENCES


Antonio Ferreras, MD, PhD, received his medical degree from the Zaragoza University, Spain followed by eye surgery training and an ophthalmology residency at the Miguel Servet University Hospital, Zaragoza, Spain. Dr Ferreras was awarded his doctorate by the University of Zaragoza in 2003. His research interests include all aspects of glaucoma diagnosis and treatment, with a particular focus on early glaucoma diagnosis techniques.
Luís E. Pablo, MD, is a Professor of Ophthalmology and Optics at Zaragoza University in Spain. He completed his residency in ophthalmology in 1994. His areas of research interest are clinical and experimental glaucoma. Currently, Dr Pablo is Vice-Secretary-General of the Spanish Glaucoma Society and practices at the Glaucoma Unit in the Department of Ophthalmology at Miguel Servet University Hospital, Zaragoza, Spain.