Exploring the Heidelberg Retinal Tomograph 3 Diagnostic Accuracy across Disc Sizes and Glaucoma Stages

A Multicenter Study

Francesco Oddone, MD,[1] Marco Centofanti, MD, PhD,[1,2] Luca Rossetti, MD,[3] Michele Iester, MD, PhD,[4] Paolo Fogagnolo, MD,[1] Elisabetta Capris, MD,[4] Gianluca Manni, MD[1,2]

Purpose: To investigate and compare the diagnostic accuracy of the Heidelberg Retinal Tomograph 3 (HRT3) diagnostic algorithms and establish whether they are affected by optic disc size and glaucoma severity.

Design: Multicenter cross-sectional evaluation of diagnostic tests.

Participants: Two hundred forty-two eyes from 139 normal subjects and 103 glaucomatous patients classified by the presence of a repeatable visual field (VF) defect.

Testing: Eyes were imaged by the HRT3. The diagnostic accuracies of Moorfields regression analysis (MRA) and the glaucoma probability score (GPS) was explored by sensitivity and specificity and area under the receiver operating characteristics curves (AUC). The analysis was performed globally and by optic disc size quartiles and by 3 VF severity groups.

Main Outcome Measures: Sensitivity, specificity, and AUC.

Results: The GPS showed a sensitivity (80% vs. 77%) similar to and a specificity (57% vs. 67%) lower than that of MRA Result. It showed a higher specificity in small discs than MRA Result (77% vs. 68%) but a low to very low specificity in medium to very large discs (medium, 61%; large, 50%; very large, 26%). Moorfields regression analysis Global showed the highest sensitivity and specificity (68% and 78%) in very large discs. R. Burke linear discriminant function (RB-LDF) and cup shape measure (CSM) showed the best and least variable AUC across optic nerve head sizes and glaucoma stages. The sensitivity of both MRA and the GPS decreased at the earlier glaucoma stages. The MRA and GPS agreement was moderate throughout the entire population and in small discs and early stage, whereas it was weaker among the other disc size and glaucoma stage subgroups.

Conclusions: HRT3 diagnostic algorithms' accuracy is moderate. The GPS is less specific and more influenced by disc size than MRA. Cup shape measure and the RB-LDF offer the best and less variable performances across different disc sizes and glaucoma stages. *Ophthalmology 2008;115:1358–1365 © 2008 by the American Academy of Ophthalmology.*

Primary open-angle glaucoma (POAG) is a chronic disease characterized by progressive loss of retinal ganglion cells (RGCs) that leads to structural damage, as shown by progressive regional or diffuse thinning of the retinal nerve fiber layer (RNFL) and of the neuroretinal rim within the optic nerve head (ONH), followed by functional loss, as shown by progressive visual field (VF) defects.

The temporal sequence of glaucomatous structural/functional damage suggests that looking for structural changes at the ONH/RNFL level should theoretically allow an earlier diagnosis than seeking for functional defects. However, in the earlier stages of the disease the broad overlap between normal and glaucomatous ONH characteristics precludes a

Originally received: July 7, 2007.
Final revision: December 17, 2007.
Accepted: January 4, 2008.
Available online: March 5, 2008. Manuscript no. 2007-905.

[1] G. B. Bietti Eye Foundation for the Study and Research in Ophthalmology–IRCCS, Rome, Italy.

[2] Dipartimento di Biopatologia e Diagnostica per Immagini, University of Tor Vergata, Rome, Italy.

[3] Clinica Oculistica, Università di Milano, Ospedale San Paolo, Milan, Italy.

[4] Clinica Oculistica, University of Genoa, Genoa, Italy.

Presented as a poster at: Association for Research in Vision and Ophthalmology Annual Meeting, May 5–10, 2007, Fort Lauderdale, Florida.

No conflicting relationship exists for any author.

Reprint requests to Francesco Oddone, Via Livenza 3, 00198, Rome, Italy. E-mail: francesco.oddone@tiscali.it.

ISSN 0161-6420/08/$–see front matter
doi:10.1016/j.ophtha.2008.01.007

clear classification when the ONH is examined subjectively, thus delaying the diagnosis and appropriate treatment.

Quantitative investigations of the ONH anatomy and identification of ONH changes in the earlier stages of the glaucomatous process are the task that imaging devices are often called to perform in clinical practice.

Among imaging devices, the Heidelberg Retinal Tomograph (HRT; Heidelberg Engineering GmbH, Dossenheim, Germany) is a leading device for a 3-dimensional quantitative study and classification of the ONH shape.[1–6] Despite its good overall diagnostic performances, the accuracy of the discrimination between normality and glaucoma has been shown to be significantly influenced by disease severity.[7–9] In other terms, it has been shown that its diagnostic performances are significantly reduced in early stages of the disease when the overlap between the normal and glaucomatous ONH anatomy is broader. Besides disease severity, ONH size represents another factor able to influence HRT diagnostic accuracy; specifically, it has been shown that larger optic discs are associated with lower specificities and smaller discs with lower sensitivities.[8–11]

Recently, a new version of the HRT (HRT3) was released equipped with a larger normative database and a new operator-independent classification algorithm. The purpose of this study was to investigate and compare the diagnostic accuracies of the HRT3 diagnostic algorithms and to establish whether they are affected by optic disc size and glaucoma severity.

Materials and Methods

This multicenter, observational, cross-sectional study included a series of consecutive normal and POAG subjects from the population attending the glaucoma clinics of 3 Italian academic hospitals (University of Rome Tor Vergata, University of Milan San Paolo, and University of Genoa). Normal control subjects were subjects attending the outpatient clinics, spouses and friends of the recruited patients, or volunteers from the hospital staff. The study, approved by the institutional ethical committees, was in agreement with the tenets of the Declaration of Helsinki, and all recruited subjects were asked to sign an informed consent form after the nature of the procedure was fully explained.

Each subject underwent a comprehensive ophthalmologic evaluation including history, autorefractometry, keratometry (Javal keratometer), best-corrected visual acuity (BCVA), slit-lamp biomicroscopy, intraocular pressure (IOP) measurement by Goldmann applanation tonometry, gonioscopy, and indirect ophthalmoscopy with a 78-diopter (D) lens.

In addition to the clinical examination, all subjects performed a VF test by automated standard achromatic perimetry (SAP) using the Humphrey Field Analyzer program 24-2 Swedish Interactive Threshold Algorithm–Standard (Carl Zeiss Meditec, Inc., Dublin, CA). Subjects experienced with SAP but with the last VF test performed more than 3 months before study enrollment and subjects without previous experience with SAP were asked to undergo a second VF test within 1 week.

Both normal subjects and glaucoma patients had to have BCVA of 20/40 or better, spherical refraction within ±5 D and astigmatism within ±3 D, and an open angle by gonioscopy. The optic disc appearance was not part of the inclusion criteria in either group. Common exclusion criteria were history of neuro-ophthalmologic or retinal diseases, uveitis, history of ocular surgery or laser treatments, history of ocular trauma, rheumatologic systemic diseases, and diabetes.

Specific inclusion criteria for control subjects were IOP < 22 mmHg in both eyes, a glaucoma hemifield test result within normal limits, and a mean deviation (MD) and pattern standard deviation (PSD) within 95% confidence limits confirmed in 2 reliable consecutive SAP tests. Exclusion criteria for control subjects were family history of glaucoma, any active or past ocular pathology, history of any IOP measurement > 21 mmHg, and history of long-term use of topical or systemic steroids.

Inclusion criteria for glaucoma patients were documented history of IOP > 24 mmHg in the hospital note, 2 consecutive reliable VFs with the glaucoma hemifield test result outside normal limits, MD and PSD outside 95% confidence limits, and a cluster of at least 3 points with $P<0.05$ in the pattern deviation plot, one of each with $P<0.01$ affecting the same hemifield; the cluster did not have to be contiguous with the blind spot and did not have to cross the horizontal midline.

Heidelberg Retina Tomograph 3

All the participants were imaged using the HRT3 (software version 3.0), which is a confocal scanning laser ophthalmoscope that uses a diode laser ($\lambda = 670$ nm) to scan the retinal surface at multiple consecutive parallel focal planes. The pixel with the highest reflectivity on the z-axis across the focal planes for each x, y location is used to identify the retinal surface and to construct a topographic image of the ONH. Relative topographic heights are then calculated from a reference ring placed on the retinal surface at the periphery of the scanned area. Average corneal curvature was recorded for automatic magnification error correction, and the appropriate ethnicity database was selected before scanning. If the patient's astigmatism exceeded ±0.75 D, a supplemental cylinder lens in front of the objective lens was placed and oriented according to the axis obtained by autorefractometry and keratometry.

A mean topographic image was automatically obtained by the HRT3 software from 3 consecutive scans centered on the ONH and was used for analysis. Only high-quality images with acquisition sensitivity $> 90\%$ and a standard deviation (SD) < 40 were considered acceptable and used for the study purposes.

After scanning, a contour line was manually placed around the ONH edge by 3 experienced investigators masked to the subject's diagnosis (one for each participating center: MI, FO, PF) according to a common standard operating procedure. Briefly, the investigators took into consideration both the mean reflectance and the mean intensity images to identify the very edge of the optic disc better, corresponding to the inner edge of the Elshnig's ring, where 4 or 5 points were placed to define the contour line. Special care was taken to avoid any peripapillary atrophy within the contour line.

Once the contour line was drawn, the HRT3 image analysis algorithm automatically places a standard reference plane 50 μm below the retinal surface between 350° and 356°, which was used to split the topographic heights of each x, y location included within the contour line: relative heights above and below the reference plane were then arbitrarily considered as belonging to the neuroretinal rim or to the cup, respectively, and used to calculate ONH stereometric parameters.

Besides stereometric parameters, the HRT3 provides 2 different automatic classification algorithms of the ONH morphology.

Moorfields Regression Analysis

Moorfields regression analysis requires the placement of the contour line and compares the regression between the obtained disc area and the logarithmic transformation of the rim area against a

Table 1. Characteristics of the Glaucoma and Control Groups

	Controls	Glaucoma	*P* Value
n	137	99	
Gender (male/female)	60/77	45/54	0.8
Age	60.9±13	62.7±11	0.07
Eye (right/left)	67/70	53/46	0.48
SD of HRT3 scans (μm)	17.47±6.7	24.61±12.1	<0.001

HRT3 = Heidelberg Retinal Tomograph 3; SD = standard deviation. Continuous variables are expressed as mean ± standard deviation.

normative database made of 733 eyes from white subjects and 215 eyes from black subjects. Optic nerve heads are then classified as abnormal if the parameter is below the 99.9% confidence interval (CI) or as borderline if between the 95% and 99.9% CI. A similar classification output is also given for each separate sector. Moorfields regression analysis provides results for the global rim area (MRA Global) as well as a final classification (MRA Result). A normal MRA Result requires MRA of all sectors as well as the global rim area to be within normal limits. A borderline MRA Result occurs when at least 1 of the sectors or the global rim area is borderline, and an outside normal limits MRA Result occurs when at least 1 sector or the global rim area is outside normal limits.

Glaucoma Probability Score

Glaucoma probability score classification involves the use of a geometric model to approximate the shape of the ONH topography with a 3-dimensional surface described by 5 parameters derived from the ONH and peripapillary retinal morphology (cup size, cup depth, rim steepness, and horizontal and vertical RNFL curvature). With a standard nonlinear least-squares fitting technique, these parameters are adapted to the individual topography globally as well as in the 6 sectors of the ONH. The obtained parameters are then put into a vector machine-learning classifier that estimates the probability of finding similar data in the glaucoma group of the training data. The probability score is expressed as on ordinal index (range, 0–1), and sectors with scores > 0.28 or > 0.64 are classified as borderline or outside normal limits, respectively. The global outcome of the GPS classification is determined by the sector with the highest probability score.

Statistical Analysis

Continuous data were described as mean (± SD) and categorical data as frequency analysis. Differences among control and glaucoma groups were assessed by Mann–Whitney *U* test for continuous parameters, and chi-square test for categorical parameters.

The outcome classifications from MRA and GPS classification algorithms were treated as categorical data. Both the GPS and MRA have a borderline classification placed between the within normal limits and outside normal limits classifications, which corresponds to a value lying outside the 95% but inside the 99.9% lower CI. The borderline classification has been excluded or included either as within normal limits or outside normal limits to dichotomize the test.

Sensitivity, specificity, positive predictive value, and negative predictive value were then calculated for global MRA and GPS classification. Positive predictive value, which expresses the probability that a test positive has the disease, was calculated as (true positives/[true positives + false positives]). Negative predictive value, which expresses the probability that a test negative is healthy, was calculated as (true negatives/[true negatives + false negatives]). The area under the receiver operating characteristic curves (AUC) was used to quantify the discrimination capabilities of each continuous parameter between healthy and glaucomatous eyes.

With the purpose of exploring the hypothesis that disc size may influence the diagnostic accuracy of the HRT3, the analysis was repeated dividing the study population by disc size in 4 quartiles (small, medium, large, very large).

Moreover, the analysis was repeated dividing the study population into 3 subgroups according to the stage of the VF defect (MD > -6 decibels [dB], MD < -6 dB > -12 dB, MD < -12 dB) so as to assess the influence of glaucoma severity on the diagnostic performances of the HRT3. The variability of the diagnostic performances of continuous parameters across disc size or glaucoma stage groups has been quantified by calculating the SD of the average AUC across disc size or glaucoma stage groups.

The agreement between MRA and GPS classifications was analyzed by Cohen's κ coefficient globally and for each disc size and glaucoma stage group.

Results

A total of 268 subjects were screened to be enrolled in this study. Twenty-one candidates were not enrolled for unreliable VF results (9 control candidates and 12 glaucoma candidates), and 5 control candidates were not enrolled for abnormal VF results, so the number of enrolled subjects was 242 (103 glaucomatous and 139 control). In 6 eyes (2.5%), 4 glaucomatous (1.28-, 1.35-, 1.46-, and 3.05-mm^2 disc areas) and 2 control (1.36- and 3.13-mm^2 disc areas), the GPS was unable to provide a classification, and those eyes were not included in the analysis, which was then performed on 99 glaucoma patients and 137 controls. No differences were found in demographic characteristics among the glaucoma and control groups (demographic details given in Table 1).

Descriptives of glaucoma stages and disc size subgroups are given in Table 2. Descriptive statistics of global parameters and HRT discriminant functions are given in Table 3 (available at http://aaojournal.org). Most of the parameters statistically differed among the 2 groups except for disc area, mean cup depth, maximum cup depth, and height variation contour.

Table 2. Descriptives (Mean ± Standard Deviation) of Disease Severity and Disc Size Groups

	n	MD	PSD	Disc Area (mm^2)
Stage				
1	42	−3.74±1.29	4.67±1.72	
2	29	−8.35±1.83	7.50±2.41	
3	28	−18.07±4.93	10.4±2.88	
Controls	137	−0.72±1.26	1.73±0.45	
Disc size				
Small	59			1.48±0.19
Medium	59			1.86±0.01
Large	59			2.19±0.07
Very large	59			2.75±0.34

MD = visual field mean deviation; PSD = visual field pattern standard deviation.

Table 4. Sensitivity and Specificity (Sn/Sp), Positive Predictive Value (PPV), and Negative Predictive Value (NPV) of Moorfields Regression Analysis (MRA) and Glaucoma Probability Score (GPS) Classifications in All Discs and in Each Disc Size Group

	Without Borderline		Borderline as Normal		Borderline as Glaucoma	
Parameter/Disc Size	*Sn/Sp (%)*	*PPV/NPV*	*Sn/Sp (%)*	*PPV/NPV*	*Sn/Sp (%)*	*PPV/NPV*
MRA Result/All	77/67	0.64/0.79	68/72	0.64/0.76	80/55	0.56/0.79
MRA Global/All	50/87	0.74/0.70	42/89	0.74/0.68	58/72	0.60/0.70
GPS Global/All	80/57	0.61/0.77	66/69	0.61/0.74	84/40	0.50/0.77
MRA Result/S	76/68	0.59/0.82	70/69	0.59/0.78	78/64	0.58/0.82
MRA Global/S	50/87	0.71/0.73	43/89	0.71/0.71	57/53	0.59/0.50
GPS Global/S	60/77	0.63/0.75	55/81	0.63/0.74	64/67	0.54/0.75
MRA Result/M	67/81	0.74/0.75	56/85	0.74/0.73	72/62	0.58/0.75
MRA Global/M	33/93	0.78/0.66	28/94	0.78/0.64	44/79	0.61/0.66
GPS Global/M	75/61	0.63/0.74	60/74	0.63/0.71	80/41	0.50/0.74
MRA Result/L	75/73	0.65/0.81	63/77	0.65/0.75	79/63	0.59/0.81
MRA Global/L	48/87	0.71/0.71	42/89	0.71/0.69	54/77	0.62/0.71
GPS Global/L	88/50	0.54/0.86	58/66	0.54/0.70	92/34	0.49/0.86
MRA Result/VL	88/42	0.61/0.77	81/56	0.61/0.78	89/31	0.52/0.77
MRA Global/VL	68/78	0.75/0.72	56/84	0.75/0.69	74/56	0.59/0.72
GPS Global/VL	96/26	0.63/0.83	89/55	0.63/0.85	96/16	0.50/0.83

L = large; M = medium; S = small; VL = very large.

Diagnostic Accuracy of the Moorfields Regression Analysis and Glaucoma Probability Score Classifications

Sensitivity and specificity of MRA overall Result were 77% and 67%, respectively (positive predictive value = 0.64, negative predictive value = 0.79). The GPS global classification showed a sensitivity of 80% and specificity of 57% (positive predictive value = 0.61, negative predictive value = 0.77). For full details, see Table 4.

Diagnostic Accuracy of Stereometric Parameters

The AUC with 95% CI of stereometric parameters and discriminant functions are given in Table 5 (available at http://aaojournal.org). Cup shape measure (CSM) and the R. Burke linear discriminant function (RB-LDF) offered the best discriminatory performances, with AUCs of 0.75 (95% CI, 0.69–0.81) and 0.76 (95% CI, 0.70–0.82), respectively.

Diagnostic Accuracy of Glaucoma Probability Score Parameters

Descriptives of global GPS parameters are given in Table 3. Glaucoma probability, horizontal RNFL curvature, and cup size statistically differed among glaucoma and control eyes, whereas cup depth, vertical RNFL curvature, and rim steepness were found to be similar in the 2 groups.

The best discriminatory ability among GPS global parameters was observed for glaucoma probability, with an AUC of 0.73 (95% CI, 0.66–0.79). Full details are given in Table 5.

Influence of Disc Size

Glaucoma probability score global classification was more influenced by ONH size than MRA, with lower sensitivities in smaller discs and lower specificities in larger discs. Both algorithms showed very low specificities in very large discs (Table 4).

The diagnostic capabilities of all continuous parameters were found to depend on the size of the ONH, as shown in Table 5.

R. Burke LDF and CSM showed the best and less variable diagnostic performance across different ONH sizes. The largest AUCs were given by CSM in small discs, CSM and RB-LDF in medium discs, CSM and F. S. Mikelberg LDF in large discs, and vertical cup-to-disc (C/D) ratio in very large discs, as described in Table 5. The parameter with the less variable AUC across disc size groups was found to be CSM. The relationship between disc area and GPS was found not to be linear, for either glaucoma or control eyes, with disc areas larger than 2.5 mm^2 associated with scores higher than 0.4 and discs smaller than 1.7 mm^2 associated with lower scores (Fig 1).

Influence of Glaucoma Stage

Moorfields regression analysis Result, MRA Global, and GPS classifications showed a trend of decreasing sensitivity at the earlier stages of the disease (Table 6). Areas under the receiver operating characteristics curves for continuous parameters at different stages of glaucoma are given in Table 7 (available at http://aaojournal.org). All parameters showed decreased diagnostic performances at stage 1, with the largest AUCs given by RB-LDF (0.66) and CSM (0.66).

Among the parameters with the largest AUCs, the more stable diagnostic performances across the glaucoma stages were observed for RB-LDF (Table 7). The observed relation between GPS and glaucoma severity, as expressed by MD values, was found not to be linear, as shown in Figure 2.

Borderline Classification

When considering the borderline classification as within normal limits, specificities increased and sensitivities decreased for both the GPS and MRA in all subgroups considered in the study. When the borderline classification was considered as outside normal limits, an increased sensitivity at the cost of decreased specificity was observed in all subgroups for both MRA and the GPS.

Agreement between Moorfields Regression Analysis and Glaucoma Probability Score

The agreement between MRA Result and GPS global was moderate overall ($\kappa = 0.42$) and fair to moderate across all disc sizes

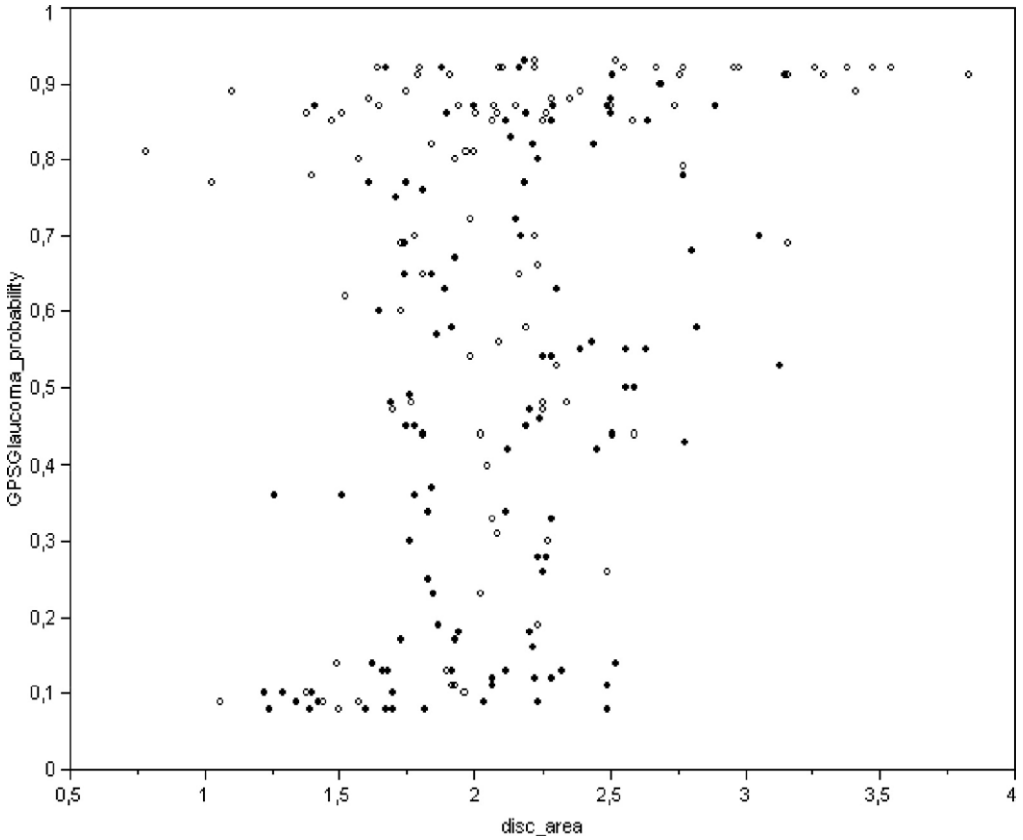

Figure 1. Overlay scatterplot of optic disc area versus the Glaucoma Probability Score (GPS) in glaucoma (○) and control (●) eyes.

(small = 0.52, medium = 0.38, large = 0.29, very large = 0.39) and glaucoma stages (stage 1 = 0.42, stage 3 = 0.38, controls = 0.30). In the stage 3 group, it was not possible to calculate the κ statistic because no borderline classification was given by GPS to match the MRA borderline class.

Discussion

One of the applications of imaging devices involves helping the clinician performing a correct glaucoma diagnosis in the early stages of the disease and when it is clinically more difficult to ascertain glaucomatous ONH changes (i.e., in the presence of larger or smaller ONHs). In this study, we investigated the diagnostic performances of the HRT3 classification tools and their agreement and how these performances may vary across different ONH sizes and glaucoma stages. The best continuous parameters to discriminate normal eyes from glaucomatous eyes in the entire sample population were found to be CSM and the RB-LDF that uses height variation contour, RNFL thickness, CSM, and rim area as input parameters (AUC, 0.75 and 0.76). We found a similar diagnostic performance for the operator-independent GPS, with an AUC of 0.73. These figures are somewhat lower than the ones previously found by Burgansky-Eliash et al[12]; this might be explainable by the multicenter nature of our study, which might have added some noise to

Table 6. Sensitivity and Specificity (Sn/Sp), Positive Predictive Value (PPV), and Negative Predictive Value (NPV) of Moorfields Regression Analysis (MRA) and Glaucoma Probability Score (GPS) Classifications in Different Glaucoma Stages

	Without Borderline		Borderline as Normal		Borderline as Glaucoma	
Parameter/Stage	*Sn/Sp (%)*	*PPV/NPV*	*Sn/Sp (%)*	*PPV/NPV*	*Sn/Sp (%)*	*PPV/NPV*
MRA Result/stage 1	66/67	0.36/0.87	53/72	0.36/0.84	73/55	0.32/0.87
MRA Global/stage 1	28/87	0.40/0.79	25/89	0.40/0.80	35/72	0.27/0.79
GPS Global/stage 1	60/57	0.30/0.82	46/69	0.30/0.82	69/40	0.25/0.82
MRA Result/stage 2	79/67	0.37/0.93	79/72	0.37/0.94	79/55	0.27/0.93
MRA Global/stage 2	61/87	0.48/0.92	50/89	0.48/0.90	68/72	0.33/0.92
GPS Global/stage 2	93/57	0.37/0.96	89/69	0.37/0.97	93/40	0.24/0.96
MRA Result/stage 3	89/67	0.39/0.96	77/72	0.39/0.93	90/55	0.31/0.96
MRA Global/stage 3	72/87	0.55/0.93	58/89	0.55/0.90	77/72	0.39/0.93
GPS Global/stage 3	92/57	0.34/0.96	71/69	0.34/0.91	94/40	0.26/0.96

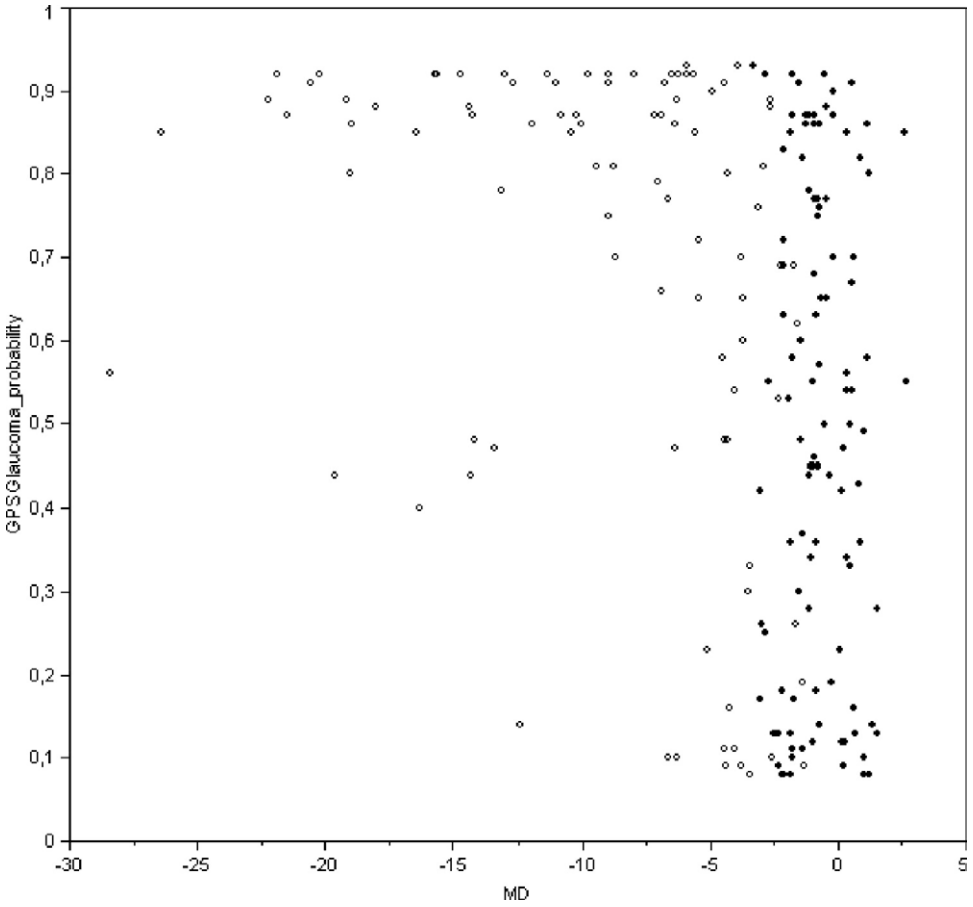

Figure 2. Overlay scatterplot of mean deviation (MD) versus the Glaucoma Probability Score (GPS) of glaucoma (○) and control (●) eyes.

the results, but with the potential advantage of being more similar to the noise of the imaging devices when used in standard clinical practice by different operators and in different environments.

Both MRA Result and GPS classifications were shown to be more sensitive than specific, and interestingly, GPS, despite a good sensitivity (80%), failed to classify correctly normal eyes in our sample population, showing a specificity of just 57%. This finding disagrees with previous reports in which the GPS and MRA showed similar sensitivities but rather higher specificities, ranging from 82% to 94%.[8,9,13] This difference might be explained looking at the gold standard used in these studies to define normal eyes. In these studies, an outcome variable of the test under investigation (i.e., appearance of the optic disc) was used as a criterion to include eyes in the normal group, and this might have represented a source of verification bias[14] that could have led to the finding of specificities higher than those found in our study, in which all subjects were merely classified by functional criteria. The choice of a reference standard significantly influences the outcome of a diagnostic study, and in the glaucoma field, a perfect reference standard for the disease is so far absent.[14,15] Nevertheless, the use of an outcome variable of a test (like the optic disc appearance for a morphological test) to restrict the entry of subjects into either the normal group or the disease group may anticipate the outcome of the study, making such studies difficult to compare with.[16]

The diagnostic accuracy of most of the HRT3 classification tools investigated in our study was shown to be significantly influenced by the severity of the disease, with weaker discriminatory performances in the earlier stage of glaucoma, corresponding to MDs > -6 dB. This finding agrees with previous reports in which a similar relationship between diagnostic accuracy and glaucoma severity has been found.[7–9] In our study, the GPS did slightly worse than MRA in detecting the early stage of the disease (sensitivity, 60% vs. 66%), whereas it did remarkably better at an intermediate stage (sensitivity, 93% vs. 79%) and similarly at an advanced stage (sensitivity, 92% vs. 89%). Similarly, Harizman et al found a decreased sensitivity for both the GPS and MRA Result in the early stage of glaucoma, but compared with our study, GPS was found to perform slightly better than MRA.[17] On the contrary, Coops et al did not find any relationship between the HRT3 diagnostic accuracy and extent of VF damage; this might be explained considering that the inclusion of glaucomatous subjects was restricted to subjects with a VF MD not worse than -10 dB, thus with a narrower spectrum of disease severity compared with our study.[18]

When an optic disc is under subjective evaluation, it is likely that its size influences the probability to be classified

as glaucomatous or normal. A glaucomatous change is hardly detected in a small and crowded optic disc. A large disc will get the attention of the observer, because of the presence of a large cup with a well visible lamina. Similarly, in this study we found that the classification of the optic disc morphology by HRT3 is influenced by the size of the disc under evaluation. We found that the GPS classification is more influenced by the size of the optic discs than MRA. Specifically, the GPS misclassified as glaucomatous 74% of the healthy very large discs and about 50% of healthy medium and large discs, whereas MRA Result showed a more constant, although not optimal, performance from small to large discs, correctly classifying from 68% to 81% of the healthy discs. In very large discs, MRA Result also showed a fall of specificity (42%). The MRA Global classification, which generally showed lower performances than both MRA Result and GPS, showed a better performance in very large discs, proving a specificity of 68% and sensitivity of 78%; this might be considered the classification of choice in this type of optic discs. Both MRA Result and the GPS performed fairly well in classifying glaucomatous eyes across disc sizes from medium to very large, with a worse performance of GPS only in small glaucomatous discs, which were correctly classified in only 60% of cases.

Both MRA Result and the GPS easily detected glaucoma in very large optic discs and performed worse in smaller discs, with a GPS sensitivity abrupt fall in the smallest discs (58%). Moreover, the GPS showed low specificities in medium, large, and very large discs, whereas MRA Result showed more constant behavior from small to large discs, with a fall only in very large discs. The GPS, when considered as a continous outcome, was found to be significantly related to disc size measurements, with disc areas larger than 2.5 mm^2 associated with scores higher than 0.4 and discs smaller than 1.7 mm^2 associated with lower scores, showing a nonlinear relationship between these variables (Fig 1).

In this study, among the parameters CSM, RB-LDF, and vertical C/D ratio showed the best discriminatory performances across different optic disc sizes, and CSM was found to be the parameter with the less variable performance across the different sizes. These results are in agreement with previous work that showed, although with different methodologies, that disc size may significantly influence the diagnostic performances of the HRT classification tools.[7–9,19]

Moreover, the weak agreement between GPS and MRA classifications across different optic disc sizes and glaucoma stages found in our study is likely to reflect the different diagnostic performances of the 2 classification algorithms described above and to support the concept that the GPS and MRA may not be used alternatively when helping the diagnosis of glaucoma.

According to our result, the diagnostic performances of HRT3 are moderate with the new operator-independent diagnostic tool that lacks in specificity compared with the traditional classification algorithms (MRA) and parameters (CSM, RB-LDF). In the real world, assessment methods for screening and early identification of a disorder rarely have perfect sensitivity and specificity. There is no general agreement about what the acceptable levels of sensitivity and specificity for an assessment test are. Acceptable levels vary depending upon the intent of the test, setting of testing (e.g., general population or a specific subgroup at risk for the condition), prevalence of the condition in the group being tested, alternate methods of assessment, and costs and benefits of testing. In the diagnostic process of detecting glaucoma, care should be taken when integrating the HRT3 classification tools; this is because, according to our results, a large normal disc might be more likely misclassified as glaucomatous by the GPS than a small disc by MRA Result classification, and in very large discs, the MRA Global classification might be preferable for its higher sensitivity and specificity. Also, particular care is required when using HRT3 diagnostic tools to help to diagnose early glaucoma, which might be more likely misclassified as normal than more advanced disease stages.

Both the GPS and MRA have a borderline category placed between the within normal limits and the outside normal limits classifications that corresponds to a value lying outside the 95% but inside the 99.9% lower CI. Clinically, the borderline category may be considered as either within or outside normal limits, and because the eyes classified as borderline might be either diseased or healthy, it will lead to either an increased specificity/decreased sensitivity or a decreased specificity/increased sensitivity, respectively. Whether one of the two choices is best may be a matter of debate; even if considering the consequences of a positive diagnosis of glaucoma, the clinician might be more interested in limiting the number of false positives (increased specificity) and thus manage the borderline classification as within normal limits.

Finally, considering that more than a few studies in the literature showed that none of the currently available imaging techniques shows superiority to subjective assessment of the ONH in detecting glaucoma,[20,21] the cost-to-benefit ratio of high-tech testing for diagnosing glaucoma should be critically considered, and the HRT classification, when available, should be used only as an integration of the entire clinical picture.

References

1. Miglior S, Guareschi M, Albe' E, et al. Detection of glaucomatous visual field changes using the Moorfields regression analysis of the Heidelberg retina tomograph. Am J Ophthalmol 2003;136:26–33.
2. Miglior S, Guareschi M, Zanchi S, et al. Diagnostic value for glaucoma of a new HRT analysis correcting for disc size. Acta Ophthalmol Scand Suppl 2002;236:44.
3. Reus NJ, de Graaf M, Lemij HG. Accuracy of GDx VCC, HRT I, and clinical assessment of stereoscopic optic nerve head photographs for diagnosing glaucoma. Br J Ophthalmol 2007;91:313–8.
4. Medeiros FA, Zangwill LM, Bowd C, Weinreb RN. 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. Arch Ophthalmol 2004;122:827–37.
5. Andreou PA, Wickremasinghe SS, Asaria RH, et al. A com-

parison of HRT II and GDx imaging for glaucoma detection in a primary care eye clinic setting. Eye 2007;21:1050–5.
6. Iester M, Mikelberg FS, Swindale NV, Drance SM. ROC analysis of Heidelberg retina tomograph optic disc shape measures in glaucoma. Can J Ophthalmol 1997;32:382–8.
7. Ferreras A, Pajarin AB, Polo V, et al. Diagnostic ability of Heidelberg retina tomograph 3 classifications. Glaucoma probability score versus Moorfields regression analysis. Ophthalmology 2007;114:1981–7.
8. Medeiros FA, Zangwill LM, Bowd C, et al. Influence of disease severity and optic disc size on the diagnostic performance of imaging instruments in glaucoma. Invest Ophthalmol Vis Sci 2006;47:1008–15.
9. Zangwill LM, Jain S, Racette L, et al. The effect of disc size and severity of disease on the diagnostic accuracy of the Heidelberg retina tomograph glaucoma probability score. Invest Ophthalmol Vis Sci 2007;48:2653–60.
10. Mardin CY, Horn FK. Influence of optic disc size on the sensitivity of the Heidelberg retina tomograph. Graefes Arch Clin Exp Ophthalmol 1998;236:641–5.
11. Iester M, Mikelberg FS, Drance SM. The effect of optic disc size on diagnostic precision with the Heidelberg retina tomograph. Ophthalmology 1997;104:545–8.
12. Burgansky-Eliash Z, Wollstein G, Bilonick RA, et al. Glaucoma detection with the Heidelberg retina tomograph 3. Ophthalmology 2007;114:466–71.
13. Pueyo V, Polo V, Larrosa JM, et al. Diagnostic ability of the Heidelberg retina tomograph, optical coherence tomograph, and scanning laser polarimeter in open-angle glaucoma. J Glaucoma 2007;16:173–7.
14. Zhou XH. Correcting for verification bias in studies of a diagnostic test's accuracy. Stat Methods Med Res 1998;7:337–53.
15. Medeiros FA, Ng D, Zangwill LM, et al. The effects of study design and spectrum bias on the evaluation of diagnostic accuracy of confocal scanning laser ophthalmoscopy in glaucoma. Invest Ophthalmol Vis Sci 2007;48:214–22.
16. Garway-Heath DF, Hitchings RA. Sources of bias in studies of optic disc and retinal nerve fibre layer morphology. Br J Ophthalmol 1998;82:986.
17. Harizman N, Zelefsky JR, Ilitchev E, et al. Detection of glaucoma using operator-dependent versus operator-independent classification in the Heidelberg Retinal Tomograph-III. Br J Ophthalmol 2006;90:1390–2.
18. Coops A, Henson DB, Kwartz AJ, Artes PH. Automated analysis of Heidelberg retina tomograph optic disc images by glaucoma probability score. Invest Ophthalmol Vis Sci 2006;47:5348–55.
19. Ford BA, Artes PH, McCormick TA, et al. Comparison of data analysis tools for detection of glaucoma with the Heidelberg retina tomograph. Ophthalmology 2003;110:1145–50.
20. Deleon-Ortega JE, Arthur SN, McGwin G Jr, et al. Discrimination between glaucomatous and nonglaucomatous eyes using quantitative imaging devices and subjective optic nerve head assessment. Invest Ophthalmol Vis Sci 2006;47:3374–80.
21. Kesen MR, Spaeth GL, Henderer JD, et al. The Heidelberg retina tomograph vs clinical impression in the diagnosis of glaucoma. Am J Ophthalmol 2002;133:613–6.

Table 3. Mean and Standard Deviation (SD) Values of Heidelberg Retinal Tomograph 3 Global Stereometric and Glaucoma Probability Score (GPS) Parameters in the Control and Glaucoma Groups

	Controls		Glaucoma		
	Mean	*SD*	*Mean*	*SD*	***P* Value**
Global stereometric parameters					
Disc area	2.04	0.44	2.13	0.59	0.3736
Cup area	0.65	0.4	0.98	0.6	<0.001*
Rim area	1.39	0.32	1.14	0.44	<0.001*
Cup-to-disc area ratio	0.3	0.16	0.44	0.19	<0.001*
Rim-to-disc area ratio	0.7	0.16	0.56	0.19	<0.001*
Cup volume	0.17	0.14	0.33	0.41	0.0073*
Rim volume	0.35	0.15	0.26	0.17	<0.001*
Mean cup depth	0.24	0.1	0.3	0.17	0.0636
Maximum cup depth	0.65	0.23	0.67	0.28	0.9676
Height variation contour	0.39	0.1	0.39	0.17	0.2585
Cup shape measure	−0.17	0.07	−0.09	0.08	<0.001*
Mean RNFL thickness	0.23	0.09	0.17	0.08	<0.001*
RNFL cross sectional area	1.15	0.42	0.88	0.44	<0.001*
Horizontal cup-to-disc ratio	0.56	0.22	0.64	0.23	0.0033*
Vertical cup-to-disc ratio	0.45	0.21	0.63	0.24	<0.001*
FSM	0.87	1.81	−0.9	2.57	<0.001*
RB	1.13	1.02	0.02	1.24	<0.001*
Global GPS parameters					
Glaucoma probability	0.43	0.3	0.67	0.29	<0.001*
Cup depth	0.62	0.19	0.62	0.24	0.5
Horizontal RNFL curvature	−0.02	0.05	−0.07	0.06	<0.001*
Vertical RNFL curvature	−0.11	0.05	−0.12	0.05	0.199
Rim steepness	−0.25	0.53	−0.32	0.52	0.34
Cup size	0.43	0.19	0.57	0.29	0.001*

FSM = F. S. Mikelberg linear discriminant function; RB = R. Burke linear discriminant function; RNFL = retinal nerve fiber layer.
*$P<0.05$.

Table 5. Area under the Receiver Operating Characteristic Curves (AUC) and 95% Confidence Interval (CI) for Global Stereometric and Glaucoma Probability Score (GPS) Parameters

	All Discs		**Small**		**Medium**		**Large**		**Very Large**		
	AUC	*95% CI*	*AUC*	*95% CI*	*AUC*	*95% CI*	*AUC*	*95% CI*	*AUC*	*95% CI*	**AUC SD**
Stereometric parameters											
RB	**0.76**	0.70–0.82	0.79	0.66–0.91	**0.73**	0.60–0.87	0.70	0.56–0.84	0.83	0.72–0.93	0.059
Cup shape measure	0.75	0.69–0.81	**0.81**	0.70–0.93	**0.73**	0.60–0.86	**0.75**	0.62–0.87	0.76	0.64–0.89	**0.034**
Vertical cup-to-disc ratio	0.73	0.67–0.80	0.77	0.64–0.90	0.72	0.59–0.86	0.69	0.55–0.83	**0.85**	0.74–0.95	0.070
FSM	0.73	0.66–0.79	0.78	0.67–0.90	0.65	0.50–0.79	**0.75**	0.61–0.88	0.76	0.63–0.89	0.058
Mean RNFL thickness	0.71	0.65–0.78	0.76	0.64–0.89	0.70	0.55–0.84	0.68	0.54–0.83	0.70	0.56–0.84	0.035
Cup-to-disc area ratio	0.70	0.64–0.77	0.74	0.61–0.88	0.67	0.53–0.82	0.69	0.55–0.83	0.78	0.66–0.90	0.050
Linear cup-to-disc ratio	0.70	0.63–0.77	0.74	0.60–0.88	0.68	0.54–0.82	0.68	0.54–0.82	0.78	0.65–0.90	0.049
RNFL cross-sectional area	0.70	0.63–0.77	0.78	0.66–0.90	0.67	0.52–0.82	0.69	0.55–0.84	0.69	0.55–0.83	0.049
Rim volume	0.69	0.62–0.76	0.76	0.64–0.88	0.67	0.53–0.82	0.67	0.53–0.82	0.70	0.56–0.84	0.042
Rim area	0.69	0.62–0.76	0.77	0.64–0.90	0.62	0.47–0.77	0.70	0.57–0.84	0.75	0.61–0.88	0.067
Cup area	0.66	0.59–0.73	0.71	0.57–0.85	0.70	0.56–0.84	0.68	0.54–0.82	0.76	0.64–0.89	0.034
Horizontal cup-to-disc ratio	0.61	0.54–0.69	0.66	0.51–0.82	0.59	0.44–0.74	0.60	0.45–0.75	0.61	0.46–0.76	0.031
Cup volume	0.60	0.53–0.68	0.64	0.49–0.78	0.57	0.42–0.72	0.60	0.45–0.75	0.70	0.56–0.85	0.056
Mean cup depth	0.57	0.50–0.64	0.59	0.44–0.74	0.51	0.36–0.66	0.56	0.41–0.71	0.66	0.51–0.81	0.063
Height variation contour	0.54	0.47–0.62	0.64	0.49–0.80	0.60	0.44–0.75	0.51	0.36–0.67	0.43	0.28–0.59	0.094
Maximum cup depth	0.50	0.42–0.57	0.52	0.36–0.67	0.62	0.48–0.77	0.51	0.35–0.66	0.62	0.47–0.77	0.061
GPS parameters											
Glaucoma probability	0.73	0.66–0.79	0.72	0.58–0.86	0.71	0.58–0.85	0.70	0.57–0.84	0.84	0.73–0.95	0.066
Horizontal RNFL curvature	0.72	0.66–0.79	0.67	0.51–0.82	0.69	0.55–0.83	0.70	0.58–0.85	0.81	0.70–0.92	0.063
Cup size	0.63	0.56–0.70	0.66	0.51–0.81	0.63	0.49–0.78	0.57	0.41–0.73	0.74	0.61–0.88	0.071
Vertical RNFL curvature	0.54	0.47–0.62	0.57	0.42–0.73	0.51	0.36–0.66	0.59	0.44–0.74	0.53	0.38–0.68	0.037
Rim steepness	0.54	0.46–0.61	0.57	0.41–0.72	0.42	0.27–0.57	0.58	0.43–0.73	0.57	0.42–0.72	0.077
Cup depth	0.47	0.40–0.55	0.54	0.38–0.69	0.63	0.49–0.78	0.54	0.39–0.69	0.57	0.42–0.73	0.042

FSM = F. S. Mikelberg linear discriminant function; RB = R. Burke linear discriminant function; RNFL = retinal nerve fiber layer; SD = standard deviation of AUC across the 4 disc size groups.
In **bold** are the largest AUCs per disc size and the lowest SD among the best parameters.

Table 7. Area under the Receiver Operating Characteristic Curves (AUC) and 95% Confidence Intervals (CIs) of Stereometric and Glaucoma Probability Score (GPS) Parameters at Different Stages of Glaucoma Severity

	Stage 1		**Stage 2**		**Stage 3**		
	AUC	*95%CI*	*AUC*	*95%CI*	*AUC*	*95%CI*	**AUC SD**
Stereometric parameters							
RB	**0.67**	0.57–0.67	0.78	0.68–0.88	0.86	0.79–0.93	**0.095**
Cup shape measure	0.66	0.57–0.75	0.76	0.66–0.87	**0.87**	0.80–0.94	0.105
Vertical cup-to-disc ratio	0.62	0.52–0.72	0.77	0.65–0.89	**0.87**	0.79–0.95	0.126
FSM	0.61	0.51–0.72	0.8	0.70–0.90	0.83	0.75–0.91	0.119
Mean RNFL thickness	0.63	0.53–0.73	0.73	0.63–0.84	0.8	0.71–0.88	0.085
Cup-to-disc area ratio	0.59	0.49–0.69	0.75	0.63–0.86	0.83	0.75–0.92	0.122
Linear cup-to-disc ratio	0.58	0.48–0.68	0.74	0.63–0.86	0.83	0.75–0.92	0.127
RNFL cross-sectional area	0.63	0.53–0.73	0.73	0.62–0.84	0.76	0.65–0.86	0.068
Rim volume	0.62	0.51–0.72	0.72	0.61–0.83	0.77	0.68–0.86	0.076
Rim area	0.6	0.49–0.70	0.75	0.64–0.86	0.76	0.66–0.86	0.090
Cup area	0.57	0.47–0.66	0.69	0.57–0.81	0.8	0.71–0.88	0.115
Horizontal cup-to-disc ratio	0.53	0.43–0.63	0.66	0.54–0.79	0.67	0.57–0.78	0.078
Cup volume	0.5	0.40–0.60	0.67	0.54–0.80	0.69	0.59–0.79	0.104
Mean cup depth	0.47	0.38–0.57	0.68	0.56–0.81	0.62	0.51–0.73	0.108
Height variation contour	0.6	0.49–0.71	0.5	0.37–0.64	0.53	0.40–0.66	0.051
Maximum cup depth	0.41	0.32–0.51	0.62	0.49–0.74	0.52	0.40–0.64	0.105
GPS parameters							
Glaucoma probability	0.6	0.50–0.70	**0.83**	0.75–0.91	0.79	0.70–0.88	0.123
Horizontal RNFL curvature	**0.62**	0.52–0.62	0.82	0.72–0.93	0.76	0.66–0.86	0.103
Cup size	0.55	0.45–0.64	0.68	0.56–0.80	0.7	0.59–0.81	0.081
Vertical RNFL curvature	0.55	0.45–0.65	0.57	0.44–0.69	0.53	0.43–0.63	0.020
Rim steepness	0.5	0.40–0.59	0.58	0.45–0.70	0.55	0.43–0.66	0.040
Cup depth	0.39	0.30–0.48	0.63	0.50–0.75	0.45	0.32–0.57	0.125

FSM = F. S. Mikelberg linear discriminant function; RB = R. Burke linear discriminant function; RNFL = retinal nerve fiber layer; SD = standard deviation of AUC across the 3 glaucoma stage groups.
In **bold** are the largest AUCs per disc size and the lowest SD among the best parameters.