# Influence of Disc Size on Optic Nerve Head versus Retinal Nerve Fiber Layer Assessment for Diagnosing Glaucoma

*Francesco Oddone, MD, PhD,[1] Marco Centofanti, MD, PhD,[1,2] Lucia Tanga, MD,[1] Mariacristina Parravano, MD,[1] Manuele Michelessi, MD,[2] Mauro Schiavone, MD,[3] Carlo Maria Villani, MD,[3] Paolo Fogagnolo, MD,[1] Gianluca Manni, MD[2]*

***Purpose:*** To explore and compare the influence of optic disc size on the diagnostic accuracy of retinal nerve fiber layer (RNFL) thickness and optic nerve head (ONH) quantitative assessment.

***Design:*** Observational, cross-sectional evaluation of diagnostic tests.

***Participants:*** We included 120 eyes from 50 normal subjects and 70 glaucomatous patients classified by the presence of a repeatable visual field defect for the analysis.

***Testing:*** The RNFL thickness was measured by scanning laser polarimetry with variable corneal compensator (GDx-VCC, Carl-Zeiss Meditec, Dublin, CA) and spectral-domain optical coherence tomography (Cirrus HD-OCT, Carl Zeiss Meditec, Inc). We obtained ONH imaging by means of confocal scanning laser ophthalmoscopy (HRT3; Heidelberg Engineering, GmbH, Dossenheim, Germany).

***Main Outcome Measures:*** Sensitivity and specificity for normative classifications, sensitivity at fixed specificity and area under the receiver operating characteristics curve (AUC) for continuous parameters. A logistic marginal regression model and coefficients of variation (CoV) have been used to test and quantify the influence of optic disc size on the diagnostic accuracy of the 3 technologies under investigation.

***Results:*** Among continuous parameters average RNFL thickness for Cirrus HD-OCT, nerve fiber indicator for GDx-VCC and cup shape measure for the HRT3 showed the best diagnostic accuracy with an AUC of 0.97, 0.94, and 0.94, respectively. Among normative classifications, the highest sensitivity and specificity were found for OCT average RNFL thickness (75.8% and 94.7%), for GDx superior thickness (77.1% and 97.5%), for HRT3 Moorfields regression analysis result (89.4% and 73.7%) and for HRT3 GPS global (92.3% and 76.5%). The diagnostic performance of HRT3 parameters seemed to be significantly influenced by optic disc size, although the same was not true for Cirrus HD-OCT and GDx VCC. The most steady performers for each imaging device across disc size groups were Cirrus HD-OCT average thickness (CoV, 1.6%), GDx-VCC inferior thickness (CoV, 2.5%), and HRT3 GPS temporal and nasal (CoV, 21.4%).

***Conclusions:*** The diagnostic accuracy of quantitative RNFL assessment as performed by Cirrus HD-OCT and GDx-VCC is high and virtually unaffected or only minimally affected by the size of the optic disc and may provide more consistent diagnostic outcomes across small and large discs than ONH assessment as performed by HRT3.

***Financial Disclosure(s):*** The authors have no proprietary or commercial interest in any of the materials discussed in this article. *Ophthalmology 2011;118:1340–1347 © 2011 by the American Academy of Ophthalmology.*

Primary open-angle glaucoma is a chronic disease characterized by progressive loss of retinal ganglion cells 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.[1] However, the broad overlap between normal and glaucomatous ONH morphologic characteristics may preclude a clear classification when the optic disc is examined subjectively in clinical practice, thus potentially delaying diagnosis and appropriate treatment.

With the emergence of optical imaging instruments, assessment of the ONH and RNFL has been made more objective, providing quantitative information. Confocal scanning laser ophthalmoscopy (CSLO), scanning laser polarimetry (SLP), and spectral domain optical coherence tomography (OCT) are technologies that make use of the different properties of light and different characteristics of retinal tissue to obtain their measurements.[2–8] The commer-

cially available CSLO, the Heidelberg Retinal Tomograph (HRT; Heidelberg Engineering GmbH, Dossenheim, Germany) is the leading device for a 3-dimensional quantitative study and classification of the ONH shape.[9–14] Despite its good overall diagnostic performance, the accuracy of the discrimination between normality and glaucoma has been shown to be significantly influenced by ONH shape and size; specifically, it has been shown that larger optic discs are associated with lower specificities and smaller discs with lower sensitivities,[15–20] thus limiting the role of quantitative ONH assessment in the clinical diagnostic process.

Although it has been reported that RNFL quantitative assessment by either the commercially available SLP with variable corneal and lens compensator (GDx-VCC, Carl Zeiss Meditec, Dublin, CA), and spectral domain OCT may perform similarly or better than topographic ONH assessment for discriminating glaucomatous from normal eyes,[21] little is known about the influence of disc size on the diagnostic accuracy of quantitative RNFL assessment.

A previously published study by Medeiros et al[10] investigated and directly compared the influence of disc size and disease severity on the diagnostic accuracy of ONH and RNFL structural assessment in glaucoma as performed by HRT-II, GDx-VCC, and time-domain OCT. The authors found, using a logistic marginal regression model, that the diagnostic performance of the technologies under investigation was significantly influenced by these covariates with potential impact on their clinical use.

Recently, new advances in CSLO and OCT have been implemented in the commercially available glaucoma imaging devices. The latest version of CSLO, the HRT3, has been released with an improved normative database and diagnostic algorithms, and the spectral domain OCT technology allows faster 3-dimensional imaging of retina and RNFL with potential improvements in the diagnosis of glaucoma. Despite several reports that have been recently published[5] about the diagnostic accuracy of these technologies, no direct comparative data are yet available about whether and how their diagnostic performances are influenced by the size of optic disc and severity of the disease.

This study explores and compares the influence of ONH size on the diagnostic accuracy of RNFL and ONH quantitative assessment performed by the spectral domain Cirrus HD-OCT (Carl Zeiss Meditec), GDx-VCC, and HRT3.

## Material and Methods

This observational, cross-sectional evaluation of diagnostic tests was approved by the institutional ethical committees, was in agreement with the tenets of the Declaration of Helsinki; all recruited subjects were asked to sign an informed consent form after the nature of the procedure was fully explained. A series of consecutive healthy subjects and primary open-angle glaucoma patients from the population attending the clinics of the G.B. Bietti Eye Foundation, Rome, Italy, and the University of Rome Tor Vergata were enrolled. Normal control subjects were either subjects attending the outpatient clinics, spouses and friends of the recruited patients, or volunteers from the hospital staff.

Each subject underwent a comprehensive ophthalmologic evaluation including history, autorefractometry, keratometry (Javal

keratometer), best-corrected visual acuity, slit-lamp biomicroscopy, intraocular pressure (IOP) measurement by Goldmann applanation tonometry, gonioscopy, and indirect ophthalmoscopy with a 78-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 (Carl Zeiss Meditec). Subjects experienced with SAP but with the last VF test performed >3 months before the study enrollment and subjects without previous experience with SAP were asked to perform a second VF test within 1 week.

Both normal subjects and glaucoma patients had to have a best-corrected visual acuity of ≥20/40, a spherical refraction within ±5 diopters, astigmatism within ±3 diopters, 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, a 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 "within normal limits" (WNL), and a mean deviation (MD) and a pattern standard deviation 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 a 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 VF with the glaucoma Hemifield test "outside normal limits" (ONL), MD and pattern standard deviation outside 95% confidence limits, and a cluster of ≥3 points with P<5% in the pattern deviation plot, one of each with P<1% affecting the same Hemifield; the cluster had not to be contiguous with the blind spot and had not to cross the horizontal midline.[22]

## Heidelberg Retina Tomograph 3

All the participants were imaged using the HRT3 (software version 3.0). The HRT3 is a CSLO that uses a diode laser (l = 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 diopters, 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 <30 were considered acceptable and used for the study purposes.

After scanning, a contour line was manually placed around the ONH edge by 1 experienced investigator masked to the subjects diagnosis. Briefly, the investigator took into consideration both the mean reflectance and the mean intensity images to better identify the very edge of the optic disc, corresponding to the inner edge of the Elschnig'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 is drawn, the HRT3 image analysis algorithm automatically places a standard reference plane 50 $\mu$m below the retinal surface between 350° and 356°, which is used to split the topographic heights of each *x, y* location included within the contour line: Relative heights above and below the reference plane are then arbitrarily considered as belonging to the neuroretinal rim or to the cup respectively and used to calculate ONH stereometric parameters. Beside stereometric parameters, the HRT3 provides 2 different automatic classification algorithms of the ONH morphology.

**Moorfields Regression Analysis.** Moorfields regression analysis (MRA) requires the placement of the contour line and compares the regression between the obtained disc area (DA) and the logarithmic transformation of the rim area (RA) against a normative database made of 733 eyes from white subjects and 215 eyes from black subjects.[23] The ONHs are then classified as ONL if the parameter is below the 99.9% confidence interval, as borderline if between the 95% and 99.9% confidence interval, and as WNL if within 95% confidence interval. This classification is given for the whole optic disc, considering the global RA and for each separate sector.

**Glaucoma Probability Score.** The GPS 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 ONH and peripapillary retinal morphology[24] (cup size, cup depth, rim steepness, and horizontal and vertical RNFL curvature). These parameters are input in a vector machine learning classifier that estimates the probability of finding similar data in the glaucoma group of the training data and gives a probability score as an ordinal index between 0 and 1. Sectors with scores >0.28 or >0.64 are classified as borderline or ONL, respectively. The global outcome of the GPS classification is determined by the sector with the highest probability score.

## GDx-VCC

The RNFL was imaged by GDx-VCC, an imaging technology that measures the retardation of reflected light caused by parallel birefringent microtubules of retinal axons. The retardation of reflected light has been proven to be linearly related to RNFL thickness.[4] The GDx-VCC software calculates summary parameters based on eye quadrants, defined as superior (25°–144°), temporal (355°–24°), inferior (215°–354°), and nasal (145°–214°). The GDx-VCC parameters considered for this study were temporal-superior-nasal-inferior-temporal average, superior average, inferior average, temporal-superior-nasal-inferior-temporal standard deviation, and nerve fiber indicator (NFI). The NFI is calculated with a support vector machine algorithm[25] that is based on several RNFL measurements, and a score from 0 to 100 is assigned to each eye: The higher the NFI score, the greater the likelihood that patients have a RNFL defect compatible with a glaucomatous optic neuropathy.

For each parameter, the GDx-VCC provides a normative classification which includes 5 categories: 5th to 100th percentiles, 2nd to 5th percentiles, 1st to 2nd percentiles, 0.5th to 1st percentiles, and below 0.5th percentiles. The 5th to 100th percentiles classification was considered WNL, whereas all the others were considered ONL in our analysis. Only high-quality images, defined by a well-focused and uniformly illuminated reflectance image, with a centered optic disc and a quality score >8, and without an atypical retardation pattern, were included.

## Cirrus HD-OCT

Cirrus HD-OCT (software version 3.0) utilizes spectral domain OCT technology to acquire reproducible data with better resolution (5 $\mu$m compared with approximately 10 $\mu$m axial resolution in tissue) about 70 times faster (27 000 vs 400 A-scans/s) than time-domain OCT technology.[26–28]

The optic disk cube 200 × 200 protocol was used for acquisition and analysis. This protocol generates a cube of data through a 6 mm$^2$ grid by acquiring a series of 200 horizontal scan lines each composed of 200 A-scans. For analysis, Cirrus algorithms identify the center of the optic disc and automatically place a calculation circle of 3.46 mm diameter around it. The anterior and posterior margins of the RNFL are delineated, and after extracting from the data cube 256 A-scan samples along the path of the calculation circle, the system calculates the RNFL thickness at each point on the circle.

All images were acquired by a single, well-trained investigator during the same visit. Only scans with a signal strength of ≥6, without RNFL discontinuity or misalignments, eye movements, or blinking artefacts were included in the analysis. Cirrus HD-OCT technology provides average and sectorial RNFL thickness classified according to a internal normative data base.

For Cirrus HD-OCTs, normative classification for each parameter consisted of 4 categories: The 95th to 100th percentiles were hypernormal (white color on thickness map); 5th to 95th percentiles were normal (green); 1st to 5th percentiles were borderline (yellow); and <1st percentile was abnormal (red). White and green areas were regarded as normal and yellow and red as abnormal in our analysis. All OCT data were aligned according to the orientation of the right eye. In this way, clock-hour 9 of the circumpapillary scan represented the temporal side of the optic disc for both eyes. The RNFL thickness and normative classification were used for analysis.

## Statistical Analysis

Continuous data were described by mean values ± standard deviation and categorical data by frequency analysis. Differences among control and glaucoma groups were assessed by the Mann–Whitney U test for continuous parameters, and the chi-square test for categorical parameters.

The target condition under investigation was "definite glaucoma" and the reference standard for classification in this study was SAP (for VF defect definition, see Methods). The index test was Cirrus HD-OCT and the comparators were GDx-VCC and HRT3.

The area under the receiver operating characteristic curve was calculated to quantify the diagnostic accuracy of continuous parameter and sensitivities were compared between parameters choosing cutoff points corresponding to fixed 90% specificity.

The diagnostic accuracy of the normative classifications of the imaging devices under study was estimated by calculating sensitivity and specificity. The borderline classifications placed between the WNL and the ONL classifications has been considered as ONL to dichotomize the test outcomes and maximize sensitivity.

The agreement between global classification algorithms was analyzed by Cohen's kappa coefficients (kappa 0–0.29 = week agreement; 0.3–0.59 = moderate agreement; 0.6–0.89 = good agreement; and 0.9–1 = optimal agreement).

The influence of disc size on the diagnostic accuracy of the imaging devices under evaluation was assessed in 2 different ways. First, the analysis was repeated dividing the study population in 2 disc size groups using as cutoff the median value of DA as measured by HRT3 and calculating the coefficient of variation (CoV) for sensitivity estimates (for fixed specificities of 90%) across disc size subgroups (CoV = [standard deviation of sensitivities across disc size groups/mean of sensitivities across disc size groups]*100).

Moreover, a logistic marginal regression modeling approach, as proposed by Leisenring et al,[29] was used to fit sensitivities of the

## Table 1. Demographics of Healthy Control Subjects and Patients with Glaucoma

| Variable | Controls | Glaucoma | P Value |
|---|---|---|---|
| N | 50 | 70 | — |
| Age (y) | 64.3±6.0 | 66.2±8.6 | 0.12 |
| Female/male (no.) | 21/29 | 28/42 | 0.23 |
| Spherical equivalent (D) | −1.12±1.76 | −1.0±1.52 | 0.65 |
| VF MD (dB) | −0.4±1.1 | −8.4±6.8 | <0.001 |
| VF PSD | 1.6±0.3 | 7.2±4.5 | <0.001 |
| Visual field index | 99.27±0.97 | 78.5±20.1 | <0.001 |
| Disc size (mm$^2$) | 1.96±0.4 | 2.05±0.5 | 0.09 |

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

tests as a function of the optic disc size. This model was previously used and described by Medeiros et al.[10,30] In brief, it allows the evaluation of sensitivities of different tests taking into account the potentially confounding effect of different variables. The diagnostic tests were included in the logistic model after being dichotomized, so that their sensitivities can be compared at the same level of specificity.

For each subject, the result of the diagnostic test provided a binary variable for sensitivities that was 1 or 0 (above or below the cutoff point matching the 95% specificity). This binary variable was used as the dependent variable in the regression model, whereas, optic disc size, and test type were entered as independent variables. To avoid, the possible confounding effect of disease severity, the stage of the disease as expressed by the VF index[31] (VFI) was also input into the model as well as the interactions terms between disc size or disease severity with test type.

## Results

A total of 136 subjects were screened to be enrolled in this study. Six candidates were not enrolled for unreliable VF results (3 glaucoma candidates and 3 control candidates), so the number of enrolled subjects was 130 (76 glaucomatous and 54 controls). In 6 eyes, GDx-VCC provided an atypical pattern (4 controls and 2 glaucoma eyes) and in 4 eyes the HRT3 GPS analysis was unable to provide a classification (2 glaucoma and 2 control eyes); those eyes were not included in the analysis, which was then performed on 120 eyes (70 glaucoma and 50 controls) with a mean DA of

2.01±0.5 mm$^2$ (median, 2.02). No differences were found in demographic characteristics among the glaucoma and control groups except for VFIs (Table 1).

## Continuous Parameters

The RNFL thickness assessment as performed by Cirrus HD-OCT provided AUCs ranging from 0.77 to 0.94, with the parameters average RNFL thickness and inferior and superior RNFL thickness providing the best results. Average RNFL thickness provided a sensitivity of 86% at a specificity of 90% corresponding to a cutoff of 79 $\mu$m. Complete details are given in Table 2.

The RNFL thickness assessment as performed by GDx-VCC provided AUCs ranging from 0.91 to 0.97, with the parameter NFI providing the best result. The sensitivity of NFI was estimated to be 90% for a fixed specificity of 90%, corresponding to a cutoff value of 28. Complete details are given in Table 3.

The ONH assessment as performed by HRT3 provided AUCs ranging from 0.62 to 0.94 with the parameter cup shape measure (CSM), FSM linear discriminant function and cup/DA ratio providing the best results. The sensitivity of CSM was estimated to be 86% for a fixed specificity of 90% corresponding to a cutoff value of −0.15. Complete details are given in Table 4 (available online at http://aaojournal.org).

## Classification Algorithms

Average thickness classification provided the highest sensitivity (75.8%) at the highest specificity (94.7%) among the Cirrus HD-OCT classifications. Superior RNFL thickness classification offered the highest sensitivity (77.1%) and specificity (97.5%) among the GDx-VCC parameters.

Among the HRT3 classification algorithms, MRA-based classifications have been found to be associated with higher specificities and lower sensitivities, whereas GPS-based classifications were found associated with higher sensitivities and somewhat lower specificities (full details are given in Table 5; available online at http://aaojournal.org).

The agreement between global classification algorithms provided by different devices was found to be moderate ranging from 0.42 to 0.57 between GDx-VCC and HRT3, from 0.49 to 0.57 between Cirrus HD-OCT and HRT3, and from 0.60 to 0.61 between GDx-VCC and Cirrus HD-OCT (Table 6).

## Table 2. Area Under the Receiver Operating Characteristics Curve of Retinal Nerve Fiber Layer Thickness Measured by Cirrus HD-OCT in Controls and in Glaucoma

| | All Discs | | | Small Discs | | | Large Discs | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | AUC | Sensitivity at Specificity >90% | Cutoff ($\mu$m) | AUC | Sensitivity at Specificity >90% | Cutoff ($\mu$m) | AUC | Sensitivity at Specificity >90% | Cutoff ($\mu$m) | Sensitivity CoV (%) |
| Cirrus HD-OCT | | | | | | | | | | |
| Average thickness | 0.94 | 0.86 | 79 | 0.95 | 0.87 | 79 | 0.90 | 0.85 | 78 | 1.6 |
| Inf thickness | 0.91 | 0.82 | 96 | 0.90 | 0.77 | 99 | 0.88 | 0.73 | 92 | 3.8 |
| Sup thickness | 0.86 | 0.79 | 91 | 0.89 | 0.77 | 91 | 0.84 | 0.80 | 91 | 2.7 |
| Tmp thickness | 0.77 | 0.59 | 53 | 0.80 | 0.63 | 54 | 0.70 | 0.60 | 54 | 3.4 |
| Nas thickness | 0.77 | 0.38 | 56 | 0.81 | 0.40 | 57 | 0.73 | 0.43 | 57 | 5.1 |

Data are reported for the whole sample population and from the small and large optic discs subgroups.
AUC = area under the receiver operating characteristics curve; CoV = coefficient of variation of sensitivity across small and large disc groups; HD-OCT = high definition optical coherence tomography; Inf = inferior quadrant; Nas = nasal quadrant; Sup = superior quadrant; Tmp = temporal quadrant.

Table 3. Area Under the Receiver Operating Characteristics Curve and Sensitivity at Fixed Specificity of Retinal Nerve Fiber Layer Thickness Measured by Gdx VCC in Controls and in Glaucoma

| | All Discs | | | Small Discs | | | Large Discs | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| GDx VCC | AUC | *Sensitivity at Specificity >90%* | Cutoff | AUC | *Sensitivity at Specificity >90%* | Cutoff | AUC | *Sensitivity at Specificity >90%* | Cutoff | **Sensitivity CoV (%)** |
| NFI | 0.97 | 0.90 | 28 | 0.96 | 0.87 | 28 | 0.98 | 0.98 | 25 | 8.5 |
| TSNIT($\mu$m) | 0.94 | 0.84 | 48.7 | 0.93 | 0.8 | 48.7 | 0.93 | 0.93 | 50.1 | 10.8 |
| SUP ($\mu$m) | 0.94 | 0.83 | 55 | 0.92 | 0.77 | 54.2 | 0.97 | 0.95 | 58.9 | 15.4 |
| INF ($\mu$m) | 0.93 | 0.83 | 54.8 | 0.91 | 0.83 | 54.8 | 0.93 | 0.86 | 56.1 | 2.5 |
| SD ($\mu$m) | 0.91 | 0.79 | 15.1 | 0.88 | 0.7 | 14.6 | 0.94 | 0.89 | 15.8 | 16.6 |

Data are reported for the whole sample population and from the small and large optic discs subgroups.
AUC = area under the receiver operating characteristics curve; CoV = coefficient of variation of sensitivity across small and large disc groups; INF = inferior quadrant; NFI = nerve fiber indicator; SD = standard deviation; SUP = superior quadrant; TSNIT = temporal-superior-nasal-inferior thickness; VCC = variable corneal compensator.

## Influence of Optic Disc Size on RNFL and ONH Assessment Diagnostic Accuracy

The variability of diagnostic sensitivity across disc size groups of Cirrus HD-OCT parameters was found to be minimal, ranging from 1.6% to 5.1%, with the parameter average thickness representing the most steady performer among small and large discs (sensitivity of 87% and 85% in small and large discs respectively; CoV, 1.6%). Full details are given in Table 2.

The variability of diagnostic sensitivity across disc size groups of GDx-VCC parameters was found to range between 2.5% and 15.4%, with the parameters inferior thickness (INF) and NFI representing the most steady performers among small and large discs (INF: sensitivity 83% and 86% in small and large discs respectively, CoV 2.5%; NFI: sensitivity 87% and 98% in small and large discs respectively, CoV 8.5%). Full details are given in Table 3.

The variability of diagnostic sensitivity across disc size groups of HRT3 stereometric and GPS parameters was found to be higher, ranging between 7.6% and 95%. When considering only parameters with overall sensitivities >70%, the variability ranged from 21.4% to 43.9% with the most steady performer among small and large discs represented by CSM among stereometric parameters and GPS temporal and nasal among GPS parameters (CSM: sensitivity 69% and 98% in small and large discs respectively, CoV 24%; GPS temporal and nasal: sensitivity 68% and 92% in small and large discs respectively, CoV 21.4%).

When considering the variability of the diagnostic performances of classification algorithms across small and large discs, specificities where found to be minimally variable for either Cirrus HD-OCT, GDx-VCC, and HRT3, whereas sensitivities where

found to be minimally variable across disc sizes for Cirrus HD-OCT and GDx-VCC classifications and more variable for HRT3 parameters (Table 5). For inclusion in the logistic regression model, the sensitivity of the parameters average thickness for the Cirrus HD-OCT, NFI for the GDx-VCC, and CSM for HRT3 has been dichotomized using the cut-off point matching 95% specificity.

The logistic model had the following framework:

$$\text{Logit}(S) = \beta_0 + \beta_1 \text{GDx} + \beta_2 \text{HRT} + \beta_3 \text{VFI} + \beta_4 \text{VFI} \times \text{HRT} + \beta_5 \text{VFI} \times \text{GDx} + \beta_6 \text{Disc} + \beta_7 \text{Disc} \times \text{GDx} + \text{GDx} + \beta_8 \text{Disc} \times \text{HRT}$$

where $S$ is the sensitivity of the test, and GDx and HRT variables are for the test type; VFI and Disc represent disease severity and optic DA, respectively, and the variables VFI × HRT and VFI × GDx represent the interaction terms between optic disc, disease severity, and the type of test. Cirrus HD-OCT was used as reference test. Table 7 shows the estimated coefficients of the regression model.

The results of the logistic regression model suggest that Cirrus HD-OCT sensitivity is not significantly influenced by optic disc size as indicated by coefficient $\beta_6$ ($P = 0.946$). A similar effect was also observed for the GDx-VCC sensitivity as indicated by the lack of statistical significance of the coefficient $\beta_7$ ($P = 0.068$), whereas the statistically significant coefficient $\beta_8$ (1.297; $P = 0.015$) indicates that larger optic discs are associated with increased sensitivity of HRT3. Disease severity was found to influence the sensitivity of Cirrus HD-OCT with increased sensitivity associated with increasing severity of VF loss as indicated by the statistical significance of coefficient $\beta_3$ ($-0.058$; $P = 0.043$). This

Table 6. Agreement as Expressed by Cohen's Kappa Coefficients (standard errors) between Classification Algorithms of GDx-VCC, Cirrus HD-OCT, and HRT3

| | GDx TSNIT | GDx NFI | HRT3 GPS Global | HRT3 MRA Result | HRT3 MRA Global | HD-OCT Average |
|---|---|---|---|---|---|---|
| **GDx** TSNIT | 1 (0) | 0.73 (0.06) | 0.42 (0.08) | 0.45 (0.07) | 0.53 (0.08) | 0.60 (0.08) |
| **GDx** NFI | 0.73 (0.06) | 1 (0) | 0.57 (0.08) | 0.48 (0.08) | 0.50 (0.08) | 0.6 (0.08) |
| **HRT3** GPS global | 0.42 (0.08) | 0.57 (0.08) | 1 (0) | 0.62 (0.08) | 0.54 (0.07) | 0.49 (0.09) |
| **HRT3** MRA Result | 0.45 (0.07) | 0.48 (0.08) | 0.62 (0.08) | 1 (0) | 0.59 (0.07) | 0.57 (0.08) |
| **HRT3** MRA Global | 0.53 (0.08) | 0.50 (0.08) | 0.54 (0.07) | 0.59 (0.07) | 1 (0) | 0.53 (0.08) |
| **HD-OCT** Average | 0.60 (0.08) | 0.61 (0.08) | 0.49 (0.09) | 0.57 (0.08) | 0.53 (0.08) | 1 (0) |

GPS = glaucoma probability score; HD-OCT = high definition optical coherence tomography; HRT3 = Heidelberg Retinal Tomograph 3; MRA = Moorfields Regression Analysis; NFI = nerve fiber indicator; TSNIT = temporal-superior-nasal, inferior thickness; VCC = variable corneal compensator.

Table 7. Results of the Logistic Regression Model of Sensitivity Incorporating Disease Severity, Optic Disc Area, and Test Type as Covariates

| Variable | Coefficient | Estimate | 95% CI | P |
|---|---|---|---|---|
| Constant | $\beta_0$ | 2.588 | 0.289–5.107 | – |
| Type of test | | | | |
|   GDx VCC NFI | $\beta_1$ | 0.542 | −0.018–1.196 | 0.965 |
|   HRT3 CSM | $\beta_2$ | −0.011 | −0.542–0.548 | 0.516 |
| VFI | $\beta_3$ | −0.058 | −0.089–0.038 | 0.043 |
| VFI * HRT3 CSM | $\beta_4$ | <0.0001 | −0.030–0.029 | 0.997 |
| VFI * GDx VCC NFI | $\beta_5$ | −0.006 | −0.045–0.023 | 0.707 |
| Optic disc area | $\beta_6$ | 0.025 | −0.726–0.791 | 0.946 |
| Optic disc area * GDx VCC NFI | $\beta_7$ | −1.007 | −2.155–0.075 | 0.068 |
| Optic disc area * HRT3 CSM | $\beta_8$ | 1.297 | 0.234–2.475 | 0.015 |

Cirrus-HD OCT was used as the reference test.
CI = confidence interval; CSM = cup shape measure; HD-OCT = high definition optical coherence tomography; HRT3 = Heidelberg Retinal Tomograph 3; NFI = nerve fiber indicator; VCC = variable corneal compensator; VFI = visual field index.

influence was similar among the 3 instruments as indicated by the lack of significance of coefficients $\beta_4$ and $\beta_5$.

## Discussion

One of the applications of imaging devices involves helping the clinician in performing a correct glaucoma diagnosis when it is clinically more difficult to ascertain glaucomatous ONH changes (i.e., in the presence of larger or smaller ONHs).

From previously published studies,[10–16] it is known that the size of optic disc may significantly influence the diagnostic accuracy of quantitative ONH assessment as performed by CSLO (HRT) with larger discs associated with lower specificities and smaller discs with lower sensitivities. Less or no information is available concerning the influence of optic disc size on the diagnostic accuracy of quantitative RNFL assessment as performed by either spectral-domain OCT or SLP.

The present study investigated and compared the influence of optic disc size on the accuracy in discriminating between glaucomatous and healthy eyes of RNFL versus ONH quantitative assessment as performed by spectral-domain OCT, SLP, and CSLO.

The overall most sensitive RNFL parameters were Cirrus HD-OCT average thickness (sensitivity 86%) and GDx-VCC NFI (sensitivity 90%). The OCT superior and inferior RNFL thickness provided good sensitivities (79% and 82% respectively), whereas nasal and temporal OCT RNFL thicknesses were found to be poor performers. The GDx-VCC temporal-superior-nasal-inferior-temporal and sectorial parameters were found to have very good sensitivities all above 80% for fixed specificities of 90% indicating very good discriminating capabilities between healthy and glau-

comatous eyes. These figures are somewhat higher than those reported by Brusini et al,[32] where sensitivities of GDx parameters were reported to be lower than those found in the present study for similar fixed specificities. The reason of this difference is likely to be related to the severity of the disease in the 2 sample populations that, as expressed by the VF MD, was significantly milder in the Brusini population (MD = −3.7±3.0 dB) than in the population of the present study (MD = −8.4±6.8 dB). The results of the present study are nonetheless in agreement with the results or Reus et al,[33] where the diagnostic accuracy of GDx VCC was assessed in a population of glaucoma patients with similar disease severity characteristics (MD = −8.45 dB).

Despite the information provided by continuous parameters being useful in clinical settings, dichotomous information provided by the classification algorithms implemented in the software of the imaging devices under investigation may have a more direct impact on the clinical diagnostic process and in decision making by directly influencing the posttest probability of having the disease. Classifications associated with RNFL parameters as measured by both Cirrus HD-OCT and GDx-VCC were found associated with very high specificities, ranging between 94.7% and 100%, indicating a very low prevalence of false positives in our sample population. These specificities were associated with somewhat more variable sensitivities across different classifications with Superior thickness for GDx-VCC and RNFL average thickness for Cirrus HD-OCT associated with the highest sensitivities (77.1% and 75.8%, respectively).

The ONH quantitative analysis was performed in the present study by HRT3, which returns several continuous stereometric parameters among whose CSM, FSM discriminant function, and ratio parameters where found to be associated with the highest sensitivities. The CSM overall was found to be the best performer, with a sensitivity of 86%, followed by FSM and cup/DA ratio (sensitivity 82%). The HRT3 classification algorithms are represented by the MRA and by the GPS. The MRA provides results for the global RA (MRA global) for RA sectors as well as a final classification (MRA result) that requires MRA from all ONH sectors as well as the global MRA to be WNL for a final WNL classification. In the present study, MRA global was found to be very highly specific (100%) and moderately sensitive (71.2%), whereas the MRA result was found to be less specific (73.7%) but more sensitive (89.4%), indicating that the 2 outcomes may not be used interchangeably. It has to be highlighted that, in our sample population, normative classifications of all imaging devices under investigation provided quite higher specificities compared with previous published data, despite the arbitrary choice of considering the borderline outcome as ONL, thus privileging sensitivity rather than specificity.

The marginal logistic regression model approach used to test the influence of disc size and disease severity on the sensitivity of the 3 imaging devices returned interesting results. According to our data, in fact, the sensitivity of GDx VCC and Cirrus HD-OCT parameters seemed not to be significantly influenced by the size of the optic disc under investigation, which was found, instead, to significantly influence the diagnostic performance of HRT with larger

discs associated with higher sensitivities, confirming the results of Medeiros et al.[10]

With the aim of conveying in a clinically useful way this influence, the variability of sensitivity of each parameter was also expressed as the CoV for sensitivity estimates across disc size subgroups.

This approach confirmed the results of the logistic regression analysis, showing that sensitivity of GDx-VCC and Cirrus HD-OCT is only minimally affected by disc size with sensitivities varying across different disc size groups <15% for GDx-VCC parameters and <5% for OCT parameters. The diagnostic accuracy of quantitative ONH assessment was found significantly more variable with the sensitivity of the best overall performing parameters (CSM, FSM, cup/DA ratio) varying from 24% to 30.2% across small and large discs. The least variable diagnostic performance was that of the least overall sensitive HRT3 parameters maximum cup depth (overall sensitivity 12%) and height variation contour (overall sensitivity 21%), whose clinical utility is very limited or remains to be ascertained.

Similar results were confirmed for the influence of disc size on the classification algorithms. In fact, specificities where found to be minimally variable for either Cirrus HD-OCT, GDx-VCC, and HRT3, whereas sensitivities where found to be minimally variable across disc sizes for Cirrus HD-OCT and GDx-VCC and significantly more variable for HRT3 classifications algorithms. These results are in agreement with the results of Coops et al,[34] where optic DA was found to have a significant effect on the HRT3 classification, with an estimated 21% increase in the odds of a GPS positive classification and 15% of a positive MRA classification for each 0.1 mm$^2$ increase in area.

Whether the more consistent diagnostic accuracy of RNFL parameters across discs of different size, compared with ONH parameters, purely rely on different performances of the measuring devices being investigated or whether it might be related to a more consistent distribution of RNFL anatomic characteristics across discs of different size remain to be ascertained.

In clinical diagnostic studies, several factors may influence the accuracy and generalizability of the results. Among these factors, a key influence is exerted by the reference standard used to define the disease that may be responsible for a incorporation bias and ultimately to a inflated sensitivity of the test. To avoid this problem in the present study, the appearance of the optic disc was not used as part of the reference standard. The spectrum of the disease in the sample population is another factor that may significantly influence the performances of the diagnostic test. Although the entire spectrum of the glaucomatous disease was represented in our sample population, most of the patients showed an early VF defect that may have resulted in a lesser influence of disease severity on the diagnostic performance of the tests under investigation. Another limitation of the present study may be represented by the case-control design with a controlled high prevalence of glaucoma that does not allow generalization of the results to settings with a lower prevalence of glaucoma, such as population screening settings.

In conclusion, the results of the present study indicate that the diagnostic accuracy of quantitative RNFL assessment as performed by Cirrus HD-OCT or GDx-VCC is high and virtually unaffected or only minimally affected by the size of the optic disc and may provide more consistent diagnostic outcomes across small and large discs than ONH assessment as performed by HRT3.

Finally, it should be highlighted that these results regard the performances of RNFL and ONH imaging in discriminating between healthy and glaucomatous eyes and do not apply to the diagnostic performance of RNFL versus ONH assessment in detecting morphologic progression of the disease, an area where further studies are needed.

# References

1. Weinreb RN, Friedman DS, Fechtner RD, et al. Risk assessment in the management of patients with ocular hypertension. Am J Ophthalmol 2004;138:458–67.
2. Bathija R, Zangwill L, Berry CC, et al. Detection of early glaucomatous structural damage with confocal scanning laser tomography. J Glaucoma 1998;7:121–7.
3. Weinreb RN, Lusky M, Bartsch DU, Morsman D. Effect of repetitive imaging on topographic measurements of the optic nerve head. Arch Ophthalmol 1993;111:636–8.
4. Weinreb RN, Dreher AW, Coleman A, et al. Histopathologic validation of Fourier-ellipsometry measurements of retinal nerve fiber layer thickness. Arch Ophthalmol 1990;108:557–60.
5. Weinreb RN. Evaluating the retinal nerve fiber layer in glaucoma with scanning laser polarimetry. Arch Ophthalmol 1999;117:1403–6.
6. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. Science 1991;254:1178–81.
7. Hoh ST, Greenfield DS, Mistlberger A, et al. Optical coherence tomography and scanning laser polarimetry in normal, ocular hypertensive, and glaucomatous eyes. Am J Ophthalmol 2000;129:129–35.
8. Greenfield DS. Optic nerve and retinal nerve fiber layer analyzers in glaucoma. Curr Opin Ophthalmol 2002;13:68–76.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. Burgansky-Eliash Z, Wollstein G, Bilonick RA, et al. Glaucoma detection with the Heidelberg Retina Tomograph 3. Ophthalmology 2007;114:466–71.
15. Oddone F, Centofanti M, Rossetti L, et al. Exploring the Heidelberg Retinal Tomograph 3 diagnostic accuracy across disc sizes and glaucoma stages: a multicenter study. Ophthalmology 2008;115:1358–65.

16. Oddone F, Centofanti M, Iester M, et al. Sector-based analysis with the Heidelberg Retinal Tomograph 3 across disc sizes and glaucoma stages: a multicenter study. Ophthalmology 2009;116:1106–11.

17. 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.

18. Zhou XH. Correcting for verification bias in studies of a diagnostic test's accuracy. Stat Methods Med Res 1998;7:337–53.

19. 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.

20. 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.

21. Medeiros FA, Vizzeri G, Zangwill LM, et al. Comparison of retinal nerve fiber layer and optic disc imaging for diagnosing glaucoma in patients suspected of having the disease. Ophthalmology 2008;115:1340–6.

22. Caprioli J. Automated perimetry in glaucoma. Am J Ophthalmol 1991;11:235–9.

23. Wollstein G, Garway-Heath DF, Hitchings RA. Identification of early glaucoma cases with the scanning laser ophthalmoscope. Ophthalmology 1998;105:1557–63.

24. Swindale NV, Stjepanovic G, Chin A, Mikelberg FS. Automated analysis of normal and glaucomatous optic nerve head topography images. Invest Ophthalmol Vis Sci 2000;41:1730–42.

25. Vapnik VN. Statistical Learning Theory. New York: Wiley; 1998:401–40.

26. Ho J, Sull AC, Vuong LN, et al. Assessment of artifacts and reproducibility across spectral- and time-domain optical coherence tomography devices. Ophthalmology 2009;116:1960–70.

27. Wojtkowski M, Srinivasan V, Ko T, et al. Ultrahigh-resolution, high-speed, Fourier domain optical coherence tomography and methods for dispersion compensation. Opt Express 2004; 12:2404–22.

28. Leitgeb R, Drexler W, Unterhuber A, et al. Ultrahigh resolution Fourier domain optical coherence tomography. Opt Express 2004;12:2156–65.

29. Leisenring W, Pepe Ms, Longton G. A marginal regression modelling framework for evaluating medical diagnostic tests. Stat Med 1997;16:1263–81.

30. Medeiros FA, Sample PA, Zangwill LM, et al. A statistical approach to the evaluation of covariate effects on the receiver operating characteristic curves of diagnostic tests in glaucoma. Invest Ophthalmol Vis Sci 2006;47:2520–7.

31. Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression. Am J Ophthalmol 2008;145: 343–53.

32. Brusini P, Salvetat ML, Zeppieri M, et al. Comparison between GDx VCC scanning laser polarimetry and Stratus OCT optical coherence tomography in the diagnosis of chronic glaucoma. Acta Ophthalmol Scand 2006;84:650–5.

33. Reus NJ, Lemij HG. Diagnostic accuracy of the GDx VCC for glaucoma. Ophthalmology 2004;111:1860–5.

34. 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.

## Footnotes and Financial Disclosures

[1] IRCCS Fondazione G.B. Bietti, Rome, Italy.

[2] Glaucoma Unit, University of Tor Vergata, Rome, Italy.

[3] Unità di Oculistica I, Azienda Ospedaliera San Giovanni-Addolorata, Rome, Italy.

Correspondence:

Francesco Oddone, Via Livenza 3, 00198, Rome, Italy. E-mail: oddonef@gmail.com.