Scan quality effect on glaucoma discrimination by glaucoma imaging devices

K R Sung, **G Wollstein**, **J S Schuman**, **R A Bilonick**, **H Ishikawa**, **K A Townsend**, **L Kagemann**, and **M L Gabriele for the Advanced Imaging in Glaucoma Study Group**
UPMC Eye Center, Eye and Ear Institute, Ophthalmology and Visual Science Research Center, Department of Ophthalmology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA; www.AIGstudy.net

Abstract

Aim—To evaluate, within ocular imaging scans of acceptable quality as determined by manufacturers' guidelines, the effects of image quality on glaucoma discrimination capabilities.

Methods—One hundred and four healthy and 75 glaucomatous eyes from the Advanced Imaging in Glaucoma Study (AIGS) were imaged with GDx-VCC, HRT II and StratusOCT. Quality score (QS≥8), pixel standard deviation (SD≤50) and signal strength (SS≥5) were used as quality parameter cut-offs, respectively. GDx nerve fibre indicator (NFI) and HRT Moorfields regression analysis (MRA) classifications and OCT mean retinal nerve fibre layer (RNFL) thickness were used as the discriminatory parameters. Logistic regression models were used to model the dichotomous clinical classification (healthy vs glaucoma) as a function of image-quality parameters and discriminatory parameters.

Results—Quality parameter covariates were statistically non-significant for GDx and HRT but had an inverse effect on OCT in predicting disease (a higher SS had a lower probability of glaucoma). Age was a significant covariate for GDx and HRT, but not OCT, while ethnicity and interaction between the image quality and the institute where scans were acquired were significant covariates in the OCT models.

Conclusion—Scan quality within the range recommended as acceptable by the manufacturer of each imaging device does not affect the glaucoma discriminating ability of GDx or HRT but does affect Stratus OCT glaucoma discrimination.

The role of imaging devices in glaucoma diagnosis has been increasing in recent years by providing an automated, objective, reliable and accurate way of quantifying glaucomatous damage.[1–5] Diagnostic performance of imaging devices depends on various factors, such as operator skill, quality of image, quality of analysis, individual anatomical variability and

pathologies of ocular structures.[6–12] Previous studies have shown that these factors affect the measurements obtained by imaging devices and consequently might affect the diagnostic performance.[6 7 10] The manufacturer of the imaging devices defined a range for acceptable scan quality that depends on the appearance of the scan, but the effect of this range on the diagnostic ability has not been tested. The purpose of this study was to test the hypothesis that prediction of healthy and glaucomatous eyes is not affected by image quality for images within the manufacturer recommended and commonly used range that defines good-quality images. Validating this hypothesis can ensure the user that all images considered acceptable quality have a similar diagnostic performance. We examined the effect of image quality on diagnostic performance in three commonly used ocular imaging technologies: scanning laser polarimetry (SLP), confocal scanning laser ophthalmoscopy (CSLO) and optical coherence tomography (OCT).

Methods

All healthy and glaucomatous subjects from the Advanced Imaging in Glaucoma Study (AIGS) were enrolled in this study. AIGS is a multicentre National Institutes of Health funded clinical trial designed to investigate advanced imaging technologies that can improve the detection and management of glaucoma. The study was approved by the Institutional Review Board/Ethics Committee, and adhered to the Declaration of Helsinki and Health Insurance Portability and Accountability Act regulations, with informed consent obtained from all patients.

The study participants underwent a full ocular examination, including intraocular pressure (IOP) measurements and gonioscopy, central corneal pachymetry, anterior and posterior segment biomicroscopy before and after pupil dilation, visual-field (VF) testing, and ocular imaging with SLP, CSLO and OCT, all within a 6-month time period. Pharmacological dilation was provided if the pupil diameter was less than 3 mm. Exclusion criteria for the study were age less than 40 or greater than 80 years, history of ocular trauma or surgery other than uncomplicated cataract surgery, best-corrected visual acuity worse than 20/40, refractive error greater than +3.0 D or less than −7.0 D, central corneal thickness <500 μm, inability to view the optic nerve head or obtain acceptable imaging scans due to media opacity or poorly dilating pupil, ocular disease other than glaucoma or other diseases that might cause VF abnormalities.

Central corneal thickness was measured using ultrasound pachymetry (Pachette 2, DGH Technology, Exton, Pennsylvania).

All subjects had reliable Swedish interactive thresholding algorithm (SITA) 24–2 standard perimetry (Carl Zeiss Meditec, Dublin, California). A reliable VF test was defined as one with <30% fixation losses, false-positive or false-negative responses.

Healthy subjects were friends of patients with normal ocular examination, open anterior chamber angle and IOP ≤21 mm Hg. All healthy subjects had normal VF with mean deviation (MD) and pattern standard deviation (PSD) within 95% of the normal population and glaucoma hemifield test (GHT) within normal limits.

Glaucoma subjects had at least one of the following structural abnormalities detected during clinical examination: diffuse or localised thinning of the neuroretinal rim, vertical cup-to-disc ratio difference between eyes >0.2, disc haemorrhage or retinal nerve fibre layer (RNFL) defect. All glaucoma subjects had reproducible VF abnormalities, defined as MD and/or PSD outside 95% of the normal population or GHT outside normal limits.

Imaging

All participants were imaged with SLP (GDx-VCC; Carl Zeiss Meditec; Software version 5.5.1.5), CSLO (Heidelberg Retina Tomography, HRT II; Heidelberg Engineering, Heidelberg, Germany; Software version 1.4.1.0) and OCT (StratusOCT; Carl Zeiss Meditec; Software version 4.0) devices. Each device provides unique parameters for quality assessment. We selected quality score (QS) for GDx-VCC, pixel standard deviation (SD) for HRT II and signal strength (SS) for StratusOCT as quality parameters to be used for the analysis. These parameters appear in the standard output of each device and are easily accessible in a clinical setting. Eligible eyes had good-quality images as defined by the manufacturer of each device: QS ≥8 (GDx-VCC),[13] SD <50 (HRT II)[14] and SS ≥5 (StratusOCT).[15] If both eyes were eligible for the study, the eye with the lower-quality score was chosen, to give the widest spread of quality scores within good-quality image boundaries. In addition, data were also collected from all qualified eyes regardless of their scan quality to evaluate the effect throughout the quality range. However, due to the uneven and limited number of scans beyond the manufacturer's recommended limits among the technologies (31 scans for GDx, 22 scans for HRT and two scans for OCT), this information was used only as a secondary analysis.

The nerve fibre indicator (NFI) from GDx-VCC analysis, Moorfields regression analysis (MRA) classification from HRT II and average RNFL thickness as obtained by the fast RNFL circumpapillary scan protocol of StratusOCT were used as the discriminatory parameters. These parameters are those most commonly used in clinical practice for glaucoma detection and have been demonstrated to provide good discriminatory ability between healthy and glaucomatous eyes.[1 3 16] The NFI cut-off for outside normal limits was >40, borderline was 28 to 40, and within normal limits was <28, according to the manufacturer criteria. MRA is reported by the HRT as within normal limits, borderline or outside normal limits. The categorical classification of these discriminatory parameters and OCT average RNFL thickness were considered in our analysis.

Statistical analysis

The primary analysis was conducted with all qualified eyes that had image quality within the manufacturer recommended limits. A secondary analysis was conducted with all eyes regardless of their image quality. The Student t test, Wilcoxon test (for skewed data) and χ^2 test were used to compare the demographic, clinical characteristics and imaging devices' measurements of the healthy and glaucoma groups. The area under the receiver operating characteristic curve (AUC) was calculated for each discriminatory parameter alone, as well as for the model with age, ethnicity, site and quality score information. Separately for GDx, HRT and OCT, logistic regression models were used to assess the relation between the scan quality, demographic variables (age, ethnicity and site), and discriminatory parameter and interactions among the various components. The Akaike Information Criterion (AIC),[17] was used to determine whether a significant decrement in goodness of fit occurred after removing effects from the model. The AIC is considered to be the preferable way for choosing the best model of a set of plausible models for a given data set.[18] p Values of <0.05 were considered statistically significant. Specificity and sensitivity were calculated for different image-quality cut-offs. For this calculation, images that were classified as borderline were grouped with outside normal limit results.

Results

One hundred and four healthy eyes (104 subjects) and 75 glaucomatous eyes (75 subjects) were analysed for this study. Detailed demographics and clinical characteristics of the participants are described in table 1. In all three subgroups, the healthy subjects were

significantly younger than the glaucoma subjects. The glaucomatous VF damage was mild to moderate as indicated by the MD (~−4.00 dB). Structural measurements obtained by the imaging devices are summarised in table 2. The distribution of the quality parameters among the technologies is presented in fig 1.

GDx-VCC

One hundred and forty-eight eyes out of the 179 study eyes had eligible GDx, with 88 healthy and 60 glaucomatous eyes. None of the qualified images showed atypical birefringence. Using NFI as the classification criterion, 107 eyes were defined as within normal limits, 15 were borderline, and 26 were outside normal limits. The AUC for NFI alone was 0.894, and when accounting for age, ethnicity, site and quality, the score was 0.916. Among the various models that were tested, the lowest AIC (and therefore the best model to be used) included age (p=0.003) and NFI (p<0.001) only. This model was superior to other tested models that included other variables (including QS) and interactions among the variables. Therefore, scan quality did not add any valuable information or affect the ability to predict glaucoma in scans that were defined by the manufacturer as good-quality scans. Using all images with QS≥8 (88 healthy eyes and 60 glaucomatous eyes) the specificity and sensitivity of GDx NFI were 93.2% and 68.3%, respectively. Including only images with QS≥9 (64 healthy eyes and 37 glaucomatous eyes) the specificity and sensitivity were slightly higher: 95.3% and 73.0%, respectively.

HRTII

One hundred and fifty-seven eyes out of the 179 study eyes had eligible HRT with 95 healthy and 62 glaucomatous eyes. MRA classified 97 eyes as within normal limits, 23 as borderline and 37 as outside normal limits. AUC for MRA alone was 0.782, and when age, ethnicity, site where scanning occurred and quality score were accounted for, AUC was 0.885. The best logistic regression model for glaucoma prediction included age, ethnicity, site, MRA and interactions between sites and MRA. The only statistically significant variables in this model were age (p=0.0004) and an interaction between MRA borderline classification and one of the sites (p=0.04). Since none of the models that included SD was selected by the AIC analysis, it can be concluded that this variable did not add any valuable information in predicting glaucoma. Using all images with pixel SD <50, the specificity of HRT MRA was 83.2% with a sensitivity of 71.0%. When using only images with SD <30, the specificity and sensitivity remained similar (87.7% and 67.4%, respectively).

StratusOCT

One hundred and seventy-seven eyes out of the 179 study eyes had eligible OCT with 104 healthy and 73 glaucomatous eyes. Based on the comparison of mean RNFL thickness with the normative database, 128 were defined as within normal limits, 17 as borderline and 32 as outside normal limits. AUC for average RNFL thickness was 0.918. When signal strength, age, ethnicity and site were accounted for, AUC was 0.942. The best random intercept logistic regression model included age, ethnicity, site, average RNFL thickness, SS and interactions between sites and age. Statistically significantly variables in this model included ethnicity (p=0.03), one of the sites (p=0.01), average RNFL thickness (p<0.001) and interactions between age and two of the sites (p=0.04 and 0.009). SS had a negative effect on discrimination, shown by a negative coefficient in the logistic regression model, meaning that higher SS tended to have a lower probability of glaucoma and vice versa but had a weaker effect than any other covariate that was included in the model (p=0.07). The specificity and sensitivity of OCT mean RNFL thickness when using all images with SS ≥5 were 96.2% and 61.6%, respectively. Using images with SS ≥8, the specificity remained similar (97.1%), but the sensitivity declined to 53.6%.

All three imaging devices showed a linear relationship between the quality parameters and glaucoma prediction on the logit scale. Similar findings were also observed in the secondary analysis where all eyes that were scanned were enrolled regardless of their quality parameters. It is therefore impossible to define quality cut-offs for all three devices where scan quality will affect the ability to detect glaucoma.

Discussion

In this study, we evaluated the effect of scan quality on the glaucoma discriminating ability of commonly used imaging modalities' acceptable quality scans. We found that among "acceptable" scans, image quality did not statistically significantly affect the prediction of disease for GDx-VCC and HRT II, and had an inverse effect on StratusOCT (a higher SS had a lower probability of glaucoma). Age was a consistent statistically significant variable for GDx and HRT, while ethnicity and the institute where the scan was obtained affected only OCT.

When evaluating the relationship with scan quality for the various devices, one should consider the different properties of each of the quality parameters. The QS of GDx-VCC is defined by proprietary software that takes into account alignment, fixation, refraction and others.[13] The HRTII scan comprises three consecutive images, which are averaged into a single mean topographic image. The SD in HRTII standard output is the SD for the height measurement of each pixel for three images. Although this parameter is considered the quality parameter for HRTII, it is more specifically defining a measure of reproducibility or repeatability of three images. The SS of OCT is also a product of proprietary software that takes into account the intensity level of the signal along with the uniformity of the signal within a scan.[15]

GDx QS did not affect the prediction of disease, and so it can be considered a robust indicator for useful scans. However, the narrow range of acceptable QS (8–10) caused a relatively high proportion of inadequate-quality scans (16 healthy and 15 glaucomatous eyes) even in the hands of highly trained and certified operators in this multicentre study. This substantially higher failure rate than with other devices is decreasing the utility of the device in the clinical setting.

Although HRT SD did not show a statistically significant effect on disease detection, this may be due to the limited reflection of this parameter on the true quality properties of the scan; for example, consistently poor quality throughout the three scans may yield an overall low SD interpreted as an indication of a good-quality scan. We expect that our results are valid also for the newer iteration of this technology (HRT 3) as the fundamental properties used for the calculation of SD remain identical, even though some adjustments have been made in the reported stereometric parameters.[19] However, in HRT 3, some additional quality parameters are reported, such as scan centration, illumination, etc, though they are not reflected in a quantitative value. The utility of this additional information should be evaluated further.

The results of the current study using OCT are consistent with those of a previous study that demonstrated that scan-quality deterioration affected RNFL thickness measurements and therefore influenced the discrimination ability as observed in the current study.[6] The previous study evaluated the effect of corneal dryness on scan quality and RNFL thickness measurements with repetitive scanning through the process of corneal drying. Even though the previous study included a large number of scans with poor quality as per the manufacturer definitions, the findings were in line with the current study where only good-quality scans were included.

Given the near linear relationship between scan quality and glaucoma prediction on the logit scale that was observed with all three imaging devices, the bias imposed by the scan quality affects the discriminating ability throughout the quality range. It was expected that a linearly changing relationship would exist for poorer-quality scans but that this would level off at the higher values, where high-quality scans had all similar discrimination capabilities. However, the linear effects of the quality extended through all quality scores, even when poor-quality images were included in the analysis (secondary analysis). Therefore, we cannot recommend any adjustment to the quality cut-off. When arbitrary image-quality cut-offs were selected, we observed a slight improvement in specificity with slight reduction in sensitivity in the subgroup that included only better-quality images for HRT and OCT. For GDx, the subgroup with better image quality showed a slight improvement in both specificity and sensitivity. While image quality had an effect on discrimination ability, it should be noted that the images with higher quality encompassed a different subgroup of the study group for each device, and so caution should be used in a direct comparison of these results.

Age has been shown in numerous studies to be strongly related to the prediction of glaucoma, corresponding to the higher prevalence of the disease in aged populations.[20–25] Therefore, the significant age effect in the model for two of the devices is as expected. However, the reason for the significant contribution of ethnicity and the institute where the OCT scan was acquired is unclear and may be related to the variety of operator skill at different sites on scan quality for OCT compared with HRT and GDx.

Caution should be used when comparing the AUCs among the devices, since the study population varied slightly between the devices due to differences in qualified scans. Another possible limitation of the study stems from the clinical definition that was made by numerous clinicians across the different study sites. However, the clinicians followed a rigorous study protocol, and so this limitation should have only a small effect on our results.

In conclusion, image-quality parameters may play a role in the measurements obtained by ocular imaging devices, but when the images are within the range that the manufacturers defined as good quality, they do not significantly affect the ability of the machine to detect glaucoma for GDx and HRT but influence the ability for OCT. It is therefore recommended that OCT scan quality be accounted for when comparing scans of different quality.

Acknowledgments

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. [PubMed: 12834666]

2. Sihota R, Sony P, Gupta V, et al. Diagnostic capability of optical coherence tomography in evaluating the degree of glaucomatous retinal nerve fiber damage. Invest Ophthalmol Vis Sci 2006;47:2006–10. [PubMed: 16639009]

3. Medeiros FA, Zangwill LM, Bowd C, et al. 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. [PubMed: 15197057]

4. Wollstein G, Schuman JS, Price LL, et al. Optical coherence tomography (OCT) macular and peripapillary retinal nerve fiber layer measurements and automated visual fields. Am J Ophthalmol 2004;138:218–25. [PubMed: 15289130]

5. Jeoung JW, Park KH, Kim TW, et al. Diagnostic ability of optical coherence tomography with a normative database to detect localized retinal nerve fiber layer defects. Ophthalmology 2005;112:2157–63. [PubMed: 16290196]

6. Stein DM, Wollstein G, Ishikawa H, et al. Effect of corneal drying on optical coherence tomography. Ophthalmology 2006;113:985–91. [PubMed: 16751039]

7. 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. [PubMed: 16505035]

8. Stein DM, Ishikawa H, Hariprasad R, et al. A new quality assessment parameter for optical coherence tomography. Br J Ophthalmol 2006;90:186–90. [PubMed: 16424531]

9. Sadda SR, Wu Z, Walsh AC, et al. Errors in retinal thickness measurements obtained by optical coherence tomography. Ophthalmology 2006;113:285–93. [PubMed: 16406542]

10. Ray R, Stinnett SS, Jaffe GJ. Evaluation of image artifact produced by optical coherence tomography of retinal pathology. Am J Ophthalmol 2005;139:18–29. [PubMed: 15652824]

11. Strouthidis NG, White ET, Owen VM, et al. Factors affecting the test–retest variability of Heidelberg retina tomograph and Heidelberg retina tomograph II measurements. Br J Ophthalmol 2005;89:1427–32. [PubMed: 16234446]

12. Zangwill L, Irak I, Berry CC, et al. Effect of cataract and pupil size on image quality with confocal scanning laser ophthalmoscopy. Arch Ophthalmol 1997;115:983–90. [PubMed: 9258219]

13. GDx-VCC operation manual. San Diego: Lase Diagnostics Technologies; 2006.

14. Glaucoma module Heidelberg retina tomograph operating instructions software version 3.0. Heidelberg: Heidelberg Engineering; 2005.

15. StratusOCT user manual, rev A 4.04. Dublin: Carl Zeiss Meditec; 2004.

16. Reus NJ, Lemij HG. Diagnostic accuracy of the GDx VCC for glaucoma. Ophthalmology 2004;111:1860–5. [PubMed: 15465547]

17. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Contr 1974;19:716–23.

18. Burnham, KP.; Anderson, DR. Model selection and multimodel inference: a practical information-theoretical approach. 2nd. New York: Springer; 2002.

19. Gabriele ML, Wollstein G, Bilonick RA, et al. Comparison of parameters from Heidelberg Retina Tomographs 2 and 3. Ophthalmology 2008;115:673–77. [PubMed: 17719642]

20. Le A, Mukesh BN, McCarty CA, et al. Risk factors associated with the incidence of open-angle glaucoma: the visual impairment project. Invest Ophthalmol Vis Sci 2003;44:3783–9. [PubMed: 12939292]

21. Coleman AL, Miglior S. Risk factors for glaucoma onset and progression. Surv Ophthalmol 2008;53(1 Suppl):S3–10. [PubMed: 19038621]

22. He M, Foster PJ, Ge J, et al. Prevalence and clinical characteristics of glaucoma in adult Chinese: a population-based study in Liwan District, Guangzhou. Invest Ophthalmol Vis Sci 2006;47:2782–8. [PubMed: 16799014]

23. Rudnicka AR, Mt-Isa S, Owen CG, et al. Variations in primary open-angle glaucoma prevalence by age, gender, and race: a Bayesian meta-analysis. Invest Ophthalmol Vis Sci 2006;47:4254–61. [PubMed: 17003413]

24. Suzuki Y, Iwase A, Araie M, et al. Risk factors for open-angle glaucoma in a Japanese population: the Tajimi Study. Ophthalmology 2006;113:1613–17. [PubMed: 16828504]

25. Leske MC, Wu SY, Hennis A, et al. Risk factors for incident open-angle glaucoma: the Barbados Eye Studies. Ophthalmology 2008;115:85–93. [PubMed: 17629563]
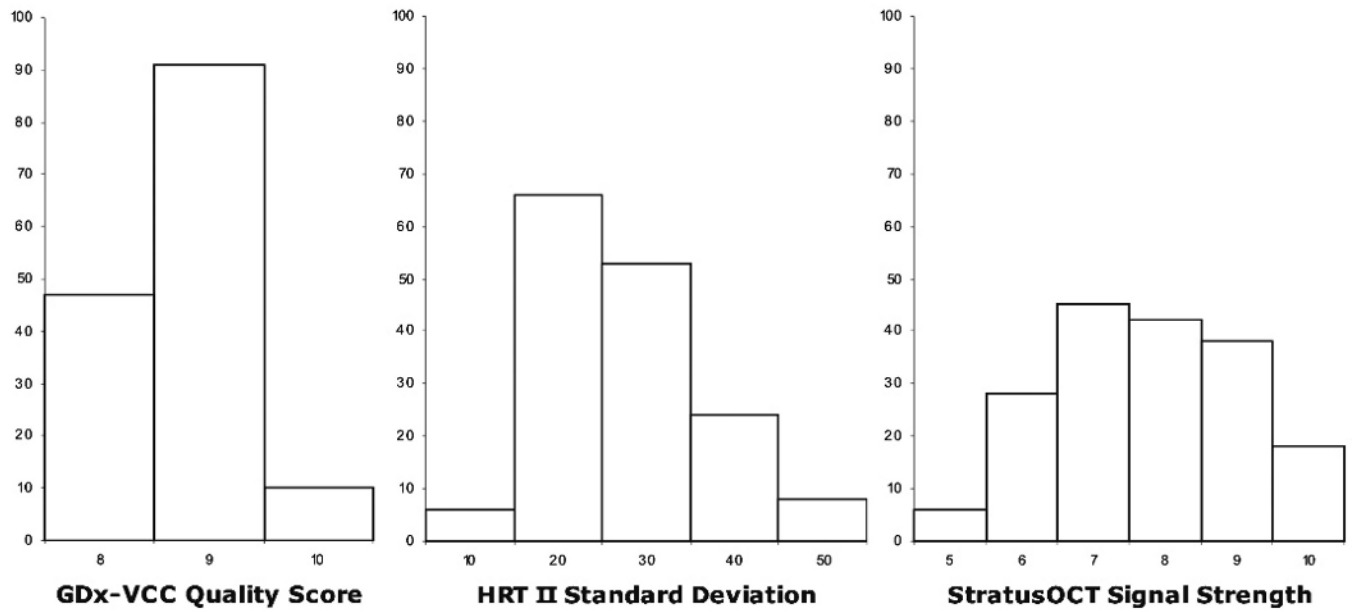
Figure 1.
Histograms of the number of eyes with each of the quality parameters of GDx-VCC (quality score), HRT II (standard deviation) and StratusOCT (signal strength).

Table 1

Demographics of the study population (mean (SD))

	GDx-VCC			HRTII			StratusOCT		
	Healthy	Glaucoma	p Value	Healthy	Glaucoma	p Value	Healthy	Glaucoma	p Value
	n=88	**n=60**		**n=95**	**n=62**		**n=104**	**n=73**	
Age (years)	56.5 (10.7)	64.4 (8.5)	<0.001*	56.2 (10.3)	63.7 (8.1)	<0.001*	56.4 (10.4)	63.7 (8.8)	<0.001*
Gender (male/female)	26/62	23/37	0.26†	28/67	24/38	0.23†	29/75	28/45	0.14†
Ethnicity			0.47†			0.47†			0.35†
Caucasian	79	50		84	51		94	59	
African–American	7	6		8	7		8	7	
Asian	2	2		3	2		2	4	
Others	0	2		0	2		0	3	
Visual-field mean deviation (dB)	−0.10 (1.25)	−3.90 (4.24)	<0.001*	−0.04 (1.28)	−3.94 (4.33)	<0.001*	−0.17 (1.25)	−3.92 (4.13)	<0.001*
Visual-field pattern SD (dB)	1.75 (1.13)	5.14 (4.21)	<0.001§	1.56 (0.41)	5.61 (4.20)	<0.001§	1.62 (0.50)	5.53 (4.17)	<0.001§

*
Student t test.

†
χ^2 test.

§
Wilcoxon test.

Table 2

Representative results of healthy and glaucoma groups obtained with the imaging devices

Device	Parameter	Healthy	Glaucoma	p Value
GDx-VCC n=148	Temporal superior nasal inferior temporal (μm)	58.3 (5.9)	49.0 (7.4)	<0.001[*]
	Nerve fibre indicator	15.3 (8.3)	37.0 (16.8)	<0.001[†]
	Quality score	8.8 (0.6)	8.7 (0.6)	0.14[*]
HRTII n=157	Disc area (mm^2)	1.97 (0.46)	2.13 (0.74)	0.33[†]
	Rim area (mm^2)	1.54 (0.43)	1.20 (0.48)	<0.001[*]
	Cup-shape measure	−0.17 (0.14)	−0.10 (0.09)	<0.001[*]
	Pixel SD	21.6 (8.8)	24.9 (9.4)	0.03[*]
StratusOCT n=177	Average retinal nerve fibre layer (μm)	99.5 (11.9)	74.2 (13.9)	<0.001[*]
	Signal strength	8.1 (1.3)	7.2 (1.2)	<0.001[*]

[*]
Student t test.

[†]
Wilcoxon test.