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, 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.AlGstudy.net
Correspondence to: Professor Joel S Schuman, UPMC Eye Center, Department of Ophthalmology, University of Pittsburgh School of Medicine, 203 Lothrop Street, Eye and Ear Institute, Suite #16, Pittsburgh, PA 15213, USA; schumanjs@upmc.edu
KRS and GW contributed equally.
Presented at the International Symposium on Imaging of the Eye (ISIE) annual meeting, Fort Lauderdale, Florida, May 2006.
Accepted 26 May 2009
Published Online First 18 August 2009

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.1–3 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.4–12 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.
Glucoma 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),25 SD <50 (HRT II)14 and SS ≥5 (StratusOCT).21 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 (51 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 analysis for OCT, this information was used only as a secondary analysis. 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 limits.

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 57 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 85.2% 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 significant 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 devices.

---

**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))

<table>
<thead>
<tr>
<th></th>
<th>GDx-VCC</th>
<th>HRTII</th>
<th>StratusOCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 88</td>
<td>Healthy</td>
<td>Glaucoma</td>
<td>n = 104</td>
</tr>
<tr>
<td>Age (years)</td>
<td>56.5 (10.7)</td>
<td>64.4 (8.5)</td>
<td>0.001*</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>26/62</td>
<td>23/37</td>
<td>0.26†</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Caucasian</td>
<td>79 50</td>
<td>0.47†</td>
</tr>
<tr>
<td></td>
<td>African–American</td>
<td>7 6</td>
<td>8 7</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>2 2</td>
<td>3 2</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>0 2</td>
<td>0 2</td>
</tr>
<tr>
<td>Visual-field mean deviation (dB)</td>
<td>-0.10 (1.25)</td>
<td>-3.90 (4.24)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Visual-field pattern SD (dB)</td>
<td>1.75 (1.13)</td>
<td>5.14 (4.21)</td>
<td>&lt;0.0015</td>
</tr>
</tbody>
</table>

* Student t test.
† χ² test.
§ Wilcoxon test.
moalities’ 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 OS of GDx-VCC is defined by proprietary software that takes into account alignment, fixation, refraction and others. The HRT II scan comprises three consecutive images, which are averaged into a single mean topographic image. The SD in HRT II standard output is the SD of 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 defined as good quality, they do not significantly affect the 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 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 ability of the machine to detect glaucoma for GDx and HRT but influence the ability for 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 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. 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. The previous study evaluated the effect of corneal dryness on scan quality and RNFL thickness measure-

### Table 2

<table>
<thead>
<tr>
<th>Device</th>
<th>Parameter</th>
<th>Healthy</th>
<th>Glaucoma</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDx-VCC</td>
<td>Temporal superior nasal inferior temporal (μm)</td>
<td>59.3 (5.9)</td>
<td>49.0 (7.4)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>n = 148</td>
<td>Vision fibre indicator</td>
<td>15.3 (8.3)</td>
<td>37.0 (16.8)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>HRT II</td>
<td>Quality score</td>
<td>8.8 (0.6)</td>
<td>8.7 (0.6)</td>
<td>0.14*</td>
</tr>
<tr>
<td>n = 157</td>
<td>Disc area (mm²)</td>
<td>1.97 (0.46)</td>
<td>2.13 (0.74)</td>
<td>0.33*</td>
</tr>
<tr>
<td>Rim area (mm²)</td>
<td></td>
<td>1.54 (0.43)</td>
<td>1.20 (0.48)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Cup-shape measure</td>
<td></td>
<td>−0.17 (0.14)</td>
<td>−0.10 (0.09)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Pixel SD</td>
<td></td>
<td>21.6 (8.8)</td>
<td>24.9 (9.4)</td>
<td>0.03*</td>
</tr>
<tr>
<td>StratusOCT</td>
<td>Average retinal nerve fibre layer (μm)</td>
<td>99.5 (11.9)</td>
<td>74.2 (13.9)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>n = 177</td>
<td>Signal strength</td>
<td>8.1 (1.3)</td>
<td>7.2 (1.2)</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>

*Student t test.
†Wilcoxon test.

OCT scan quality be accounted for when comparing scans of different quality.

**Funding:** Supported in part by NIH grants R01-EY013178-09, R01-EY013516-06, P30-EY00898-23 (Bethesda, Maryland), The Eye and Ear Foundation (Pittsburgh, Pennsylvania) and an unrestricted grant from Research to Prevent Blindness (New York).

**Competing interests:** GW received research funding from Carl Zeiss Meditec and Optovue. JSS receives royalties for intellectual property licensed by Massachusetts Institute of Technology to Carl Zeiss Meditec. GW, JSS and HI receive royalties for intellectual property licensed by the University of Pittsburgh to Biopigen.

**Ethics approval:** Ethics approval was provided by University of Pittsburgh, University of Southern California and University of Miami.

**Patient consent:** Obtained.

The Advanced Imaging for Glaucoma (AIG) project is a bioengineering partnership (NIH 1 R01 EY013516) sponsored by the National Eye Institute to improve quantitative imaging technologies for glaucoma diagnosis and management. Advanced Imaging for Glaucoma Study Group (http://www.AIGStudy.net): University of Miami, Bascom Palmer Eye Institute, Palm Beach, Florida; DS Greenfield, M Sehi, CD Quinn, KS Kishor; University of Pittsburgh Medical Center, UPMC Eye Center, Eye and Ear Institute, University of Pittsburgh School of Medicine, Pittsburgh, PA; JS Schuman, G Wollstein, H Ishikawa, L Pittsburgh Medical Center, UPMC Eye Center, Eye and Ear Institute, University of Pittsburgh School of Medicine, Pittsburgh, PA: DS Greenfield, M Sehi, CD Quinn, KS Kishor; University of Pittsburgh, University of Southern California Keck School of Medicine, Doheny Eye Institute, Los Angeles, California: D Huang (PI), R Varma, V Chopra, B Francis, F Memarzadeh, A Lu, O Tan, Steering Committee: B Francis, DS Greenfield, D Huang (Chair), JS Schuman, R Varma.

**Provenance and peer review:** Not commissioned; externally peer reviewed.

**REFERENCES**


Scan quality effect on glaucoma discrimination by glaucoma imaging devices

K R Sung, G Wollstein, J S Schuman, et al.

Br J Ophthalmol 2009 93: 1580-1584 originally published online August 18, 2009
doi: 10.1136/bjo.2008.152223

Updated information and services can be found at:
http://bjo.bmj.com/content/93/12/1580.full.html

These include:

References
This article cites 21 articles, 9 of which can be accessed free at:
http://bjo.bmj.com/content/93/12/1580.full.html#ref-list-1

Article cited in:
http://bjo.bmj.com/content/93/12/1580.full.html#related-urls

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/