Effect of Spectrum Bias on the Diagnostic Accuracy of Spectral-Domain Optical Coherence Tomography in Glaucoma

Harsha L. Rao,[1,2] *Tukaram Kumbar,*[1] *Uday K. Addepalli,*[1] *Neha Bharti,*[1] *Sirisha Senthil,*[1] *Nikhil S. Choudhari,*[1] *and Chandra S. Garudadri*[1]

Purpose. To evaluate the influence of a control group on the diagnostic accuracy of spectral-domain optical coherence tomography (SD-OCT) in early glaucoma.

Methods. In a diagnostic, case–control study, 119 eyes of 60 normal subjects with no findings suspicious for glaucoma (control cohort 1); 76 eyes of 41 subjects referred by general ophthalmologists as glaucoma suspects based on optic disc morphology, but found by glaucoma experts to be normal but with physiological variations in their optic nerves (control cohort 2); and 65 eyes of 46 early-glaucoma patients (cases) underwent imaging of the optic nerve head (ONH), retinal nerve fiber layer (RNFL), and ganglion cell complex (GCC) by SD-OCT.

Results. Areas under the receiver operating characteristic curves (AUC) of ONH parameters discriminating glaucomatous eyes from normal eyes of control cohort 2 were significantly lesser ($P < 0.001$) than those discriminating glaucomatous eyes from normal eyes of control cohort 1. AUCs of RNFL parameters discriminating glaucomatous eyes from normal eyes of control cohorts 2 and 1 were comparable. Although the AUCs of GCC thickness parameters were comparable, AUCs of GCC focal and global loss volume in control cohort 2 (0.684 and 0.671. respectively) were significantly less ($P < 0.05$) than in control cohort 1 (0.881 and 0.841, respectively).

Conclusions. The effectiveness of most SD-OCT parameters in detecting glaucoma significantly decreased when evaluated against a clinically relevant control group with suspicious-looking optic nerves compared with that against a control group consisting of normal subjects with no findings suspicious for glaucoma. (*Invest Ophthalmol Vis Sci.* 2012;53:1058–1065) DOI:10.1167/iovs.11-8463

Spectral-domain optical coherence tomography (SD-OCT) is a technique that enables imaging the ocular structures with a faster scan rate than is obtained with the previous version of the technology (Stratus OCT; Carl Zeiss Meditec, Inc., Dublin, CA).[1,2] Several studies have reported good diagnostic performance of SD-OCT in glaucoma.[3–6] These studies, as well as most of the studies with imaging technologies for glaucoma, have been of a case–control design, including glaucoma patients (cases), defined based on the presence of repeatable characteristic glaucomatous visual field defects, and normal subjects (controls), usually recruited from the general population and having normal intraocular pressure (IOP), healthy appearance of the optic nerve, and normal visual fields. However, in clinical practice, a diagnostic test is used to either rule in or rule out disease in subjects with suspected disease and not in subjects with either clear evidence of disease or no suspect findings. Measures of diagnostic effectiveness can be biased if a test is evaluated in a group of patients with clear evidence of disease and a separate group of normal subjects with no suspect findings, rather than in a relevant clinical population that has suspect findings. This bias, introduced when a diagnostic performance study is not conducted in clinically relevant population, is called spectrum bias.[7]

Spectrum bias generally is considered to have three causes.[8] The first is inappropriate selection of cases. Inclusion of cases with advanced disease in the study, for example, is known to cause an overestimation of the accuracy of the diagnostic test in detecting the disease (sensitivity). The second cause of spectrum bias is inappropriate selection of control subjects. Including controls without any signs indicative of the disease, for example, is known to inflate the performance of the test in separating normal from diseased eyes (specificity). The third cause of spectrum bias a change in the prevalence of the disease. Recruiting cases from a referral center where the prevalence of disease is obviously high compared with that in the general population is known to affect both the sensitivity and specificity of the diagnostic test. There have been studies addressing the first cause of spectrum bias by evaluating the effect of disease severity on the diagnostic performance of imaging tests in glaucoma.[9–13] All these studies unequivocally reported that diagnostic effectiveness is significantly overestimated as disease severity increases. Studies addressing the second and third causes of spectrum bias, however, are limited.[14,15]

The purpose of the present study was to evaluate the influence of the control group on the diagnostic performance of SD-OCT in early glaucoma.

Methods

This was an observational, cross-sectional study conducted at a tertiary eye care facility in South India. Informed consent was obtained from each participant, and the Ethics Committee of L. V. Prasad Eye Institute approved all methodology. All protocols complied with the Declaration of Helsinki for research involving human subjects.

Inclusion criteria were age ≥18 years, best corrected visual acuity of 20/40 or better, refractive error within ±5.0 D sphere and ±3 D cylinder, and willingness to participate in the study. Exclusion criteria

From the [1]VST Glaucoma Center and the [2]Center for Clinical Epidemiology and Biostatistics, L. V. Prasad Eye Institute, Banjara Hills, Hyderabad, India.

Supported by Optovue, which provided the RTVue for use in the study.

Submitted for publication August 23, 2011; revised November 10, 2011; accepted December 8, 2011.

Disclosure: **H.L. Rao**, Allergan (C); **T. Kumbar**, None; **U.K. Addepalli**, None; **N. Bharti**, None; **S. Senthil**, None; **N.S. Choudhari**, None; **C.S. Garudadri**, Allergan (C), Alcon (C), Merck (C), Optovue (F)

Corresponding author: Harsha L. Rao, L. V. Prasad Eye Institute, Banjara Hills, Hyderabad 500034, India; harshabl.76@gmail.com.

Investigative Ophthalmology & Visual Science, February 2012, Vol. 53, No. 2

were presence of any media opacities that prevented good imaging, intraocular surgery within the previous 6 months, and any retinal (including macular) or neurologic diseases other than glaucoma that could confound the results of visual field examination and structural measurements with SD-OCT. All participants underwent a comprehensive ocular examination that included a detailed medical history, best corrected visual acuity measurement, slit lamp biomicroscopy, Goldmann applanation tonometry, gonioscopy, dilated fundus examination, digital optic disc photography, standard automated perimetry (SAP), and SD-OCT imaging (RTVue; Optovue Inc., Fremont, CA).

We included two cohorts of control subjects for the study and analyzed the diagnostic ability of SD-OCT in discriminating glaucomatous eyes from normal eyes of the two control cohorts separately. Subjects for control cohort 1 were recruited from those who came for a routine eye examination, patients' relatives, and hospital staff. They had normal findings in an ocular examination, intraocular pressure (IOP) less than 22 mm Hg in both eyes, no history of increased IOP, no family history of glaucoma, no optic disc morphology suspicious for glaucoma (see below), and normal visual fields.

Subjects in control cohort 2 consisted of consecutive subjects who were referred to our center by general ophthalmologists as glaucoma suspects based on optic disc morphology. All these subjects had IOP less than 22 mm Hg in both eyes, no history of increased IOP, no family history of glaucoma, and normal visual fields. Although found suspicious by the general ophthalmologists, their optic discs were confirmed on clinical examination by the glaucoma experts to be nonglaucomatous, but physiological variations of normal.

Consecutive early glaucoma patients defined on the basis of Hodapp-Anderson-Parrish classification[16] formed the study group. Glaucoma patients had characteristic optic disc changes (focal or diffuse neuroretinal rim thinning, localized notching or nerve fiber layer defects) with correlating visual field changes. All glaucoma patients had previous experience in undergoing visual field examination.

Classification into the above groups was also evaluated independently on digital optic disc photographs (Visupac 4.2.2; Carl Zeiss Meditec Systems, GmbH, Pirmasens, Germany). The photographs were evaluated by two experts who were masked to the clinical examination results of the subjects and also the results of visual field and SD-OCT examinations. Discrepancies between the two experts were resolved by consensus or adjudication by a third expert.

SAP was performed with a retinal perimeter (Humphrey Field analyzer, model 750; Carl Zeiss Meditec, Inc., Dublin, CA), with the Swedish interactive threshold algorithm (SITA) standard 24-2 algorithm. Reliability criteria were fixation losses and false-positive or -negative response rates of less than 20%. Abnormal SAP results were considered to be related to glaucoma if they correlated with the optic disc findings and there were no other abnormalities to explain the defect. Glaucomatous visual field defects were defined by two of the following three criteria on SAP: the presence of a cluster of three points on a pattern deviation probability plot with $P < 5\%$, one of which had $P < 1\%$; a pattern standard deviation (PSD) with $P < 5\%$; or a glaucoma hemifield test result outside normal limits.[17]

Instrumentation

SD-OCT examination was performed with the RTVue (software version 4.0.5.39; Optovue). The instrument uses a super luminescent laser diode to provide images of ocular microstructures. The protocols used for imaging in this study were ONH (optic nerve head) and GCC (ganglion cell complex). These protocols have been explained earlier.[18] All patients had both protocols as well as the visual field testing performed on the same day. Only well-centered images with a signal strength index (SSI) of ≥30 were used for analysis. Eyes in which the segmentation algorithm failed were excluded.

ONH Scan

The ONH protocol consists of 12 radial scans 3.4 mm in length and six concentric ring scans ranging from 2.5 to 4.0 mm in diameter, all centered on the optic disc. Retinal pigment epithelium (RPE) tips are automatically detected by the software, with detection refined manually by the operator. The system software (RTVue; Optovue) then delineates the optic disc margin by joining the RPE tips. The optic cup is automatically defined by the software by fitting a plane 150 μm parallel to and above a plane that fits the coordinates of the RPE tips by the least squares error method. Rim tissue above the cup line and within the perpendicular lines drawn from the RPE tips to the surface of the retina is used to calculate the rim area and volume. ONH protocol calculates various parameters that describe the ONH and also generates a polar RNFL thickness map, which is the RNFL thickness measured along a circle 3.45 mm in diameter centered on the optic disc. The parameters generated are the average RNFL thickness in the temporal, superior, nasal, and inferior quadrant, as well as the overall average along the entire measurement circle.

GCC Scan

The ganglion cell complex (GCC) protocol was used to obtain the macular measurements. The GCC scan is designed to automatically measure the inner retinal thickness, which includes the nerve fiber layer, ganglion cell layer, and inner plexiform layer, collectively called the GCC. The parameters generated by the GCC analysis are the average superior and inferior GCC thicknesses, average superior minus inferior GCC thickness, and superior minus inferior thickness SD. In addition to the above parameters, the GCC protocol provides three other parameters called GLV (global loss volume), FLV (focal loss volume), and RMS (root mean square). These have been explained elsewhere.[6,19] In brief, GLV measures the average amount of GCC loss over the entire GCC map. FLV detects focal loss using a pattern deviation map to correct for the overall absolute changes, much like the corrected PSD in the visual fields. RMS is a parameter that provides a summary of how well the deviation maps of an individual fit the normal pattern.

In addition, a few RTVue parameters are compared with the normative database within the software and a diagnostic categorization of outside normal limits is provided if the value falls lower than the 99% confidence interval (CI) of the healthy, age-matched population. A borderline result indicates that the value is between the 95% and 99% CI, and a within-normal-limits result indicates that the value is within the 95% CI. The normative database of RTVue has data on 1081 subjects of different ethnicities, which are used to derive the normal, borderline, and outside-normal-limits cutoffs for these parameters.

Statistical Analysis

Descriptive statistics included mean and SD for normally distributed variables and median, first quartile, and third quartile values for nonnormally distributed variables. Analysis of variance or Kruskal-Wallis tests were used to compare the parameters between the three groups of participants. As measurements from both eyes of a subject are likely to correlate, a nested model was used for comparison between the groups. Eye was nested within subject, and subject as a variable was nested within group. ROC curves were used to describe the ability of each RTVue software–provided parameter to discriminate glaucomatous eyes from normal eyes of control cohorts 1 and 2. The ROC curve provides the tradeoff between the sensitivity and 1 − specificity. An area under the ROC curve (AUC) of 1.0 represents perfect discrimination, whereas an area of 0.5 represents chance discrimination. Sensitivities at fixed specificities of 80% and 95% were determined for all the parameters. ROC curves were adjusted for differences in age and optic disc size between cases and subjects of control cohort 1 and in optic disc size between cases and subjects of control cohort 2, using covariate-adjusted ROC curves, as proposed by Pepe.[20] To obtain confidence intervals for AUC, a bootstrap resampling procedure was used ($n = 1000$ resamples). As measurements from both eyes of the same subject are likely to correlate, the standard statistical methods for parameter estimation lead to underestimation of standard errors and to confidence intervals that are too narrow.[21] Therefore, the cluster of

data for the study subject was regarded as the unit of resampling and the bias-corrected SE was calculated during the AUC estimations. This procedure has been used in literature to adjust for the presence of multiple correlated measurements from the same unit.[22,23] The *z*-test was used to compare the AUCs of SD-OCT parameters obtained with control cohorts 1 and 2.[24,25]

LRs were reported for diagnostic categorization (outside normal limits, borderline, or within normal limits) provided after comparison with the instrument's normative database. An LR of 1 or close to 1 would mean that the test provides no additional information about the posttest probability of the disease. LRs higher than 10 or lower than 0.1 would be associated with large effects on posttest probability, LRs from 5 to 10 or from 0.1 to 0.2 would be associated with moderate effects, LRs from 2 to 5 or from 0.2 to 0.5 would be associated with small effects.[26] The 95% CIs for LRs were calculated according to the method proposed by Simel et al.[27]

Statistical analyses were performed with commercial software (Stata ver. 11.0; StataCorp, College Station, TX and MedCalc ver 11.5.1.0; MedCalc Software, Mariakerke, Ghent, Belgium). Differences reaching $P < 0.05$ were statically significant.

Results

We included 119 eyes of 60 normal subjects for control cohort 1 and 76 eyes of 41 subjects with physiological variations of optic nerves for control cohort 2. Sixty-five eyes of 46 patients with early glaucoma formed the glaucoma cohort. Table 1 shows the demographic, visual field, and SD-OCT parameters of the three cohorts. The glaucoma patients were significantly older than the subjects of control cohort 1. MD and PSD of visual fields were similar in subjects of control cohorts 1 and 2, and both were significantly different from those of the glaucoma patients. Optic disc size was significantly smaller in the glaucoma group than that of both control cohorts. Except for the optic disc area, all other ONH parameters were significantly different between the two control cohorts. Inferior and average RNFL, as well as most GCC parameters, were significantly different between the control cohorts. Comparing the SD-OCT parameters of control cohort 1 with the early-glaucoma cohort, all the parameters were significantly different between the groups. Comparing the SD-OCT parameters between control cohort 2 and the early glaucoma cohort, most of the ONH parameters were similar, whereas most of the RNFL and GCC parameters were significantly different.

Table 2 shows the AUCs of all SD-OCT parameters discriminating glaucomatous eyes from normal eyes of control cohorts 1 and 2 separately. The ONH, RNFL, and GCC parameters with the highest AUCs for discriminating glaucomatous eyes from normal eyes of control cohort 1 were cup-to-disc area ratio, inferior quadrant RNFL thickness, and GCC FLV. AUC of cup-to-disc area ratio was significantly greater than that of inferior quadrant RNFL thickness ($P = 0.01$) but was comparable to that of GCC FLV ($P = 0.29$). AUC of inferior quadrant RNFL thickness was comparable to that of GCC FLV ($P = 0.22$). The ONH, RNFL, and GCC parameters with the highest AUCs for discriminating glaucomatous eyes from eyes of control cohort 2 were vertical cup-to disc-ratio (CDR), inferior quadrant RNFL thickness, and GCC RMS. The AUC of vertical CDR was comparable to that of both inferior quadrant RNFL thickness ($P = 0.15$) and GCC RMS ($P = 0.37$). AUC of inferior quadrant RNFL thickness was also similar to that of GCC RMS ($P = 0.87$). Table 2 also shows the sensitivities of SD-OCT parameters to differentiate glaucomatous eyes from eyes of control cohorts 1 and 2 at fixed specificities of 95% and 80%. The table also shows

Table 1. Demographic, Visual Field, and SD-OCT Parameters of the Study Cohorts

Parameters	Control Cohort 1 (119 Eyes of 60 Subjects)	Control Cohort 2 (76 Eyes of 41 Subjects)	Glaucoma Cohort (65 Eyes of 46 Patients)	*P**	*P*†	*P*‡
Demographic features						
Age, y	47.1 ± 12.8	50.2 ± 14.7	51.9 ± 13.2	0.26	0.05	0.52
Sex, male:female	32:28	29:12	33:13	0.08	0.06	0.92
Visual field parameters						
MD, dB	−1.3 ± 1.1	−1.6 ± 1.4	−3.2 ± 1.5	0.15	<0.001	<0.001
PSD, dB	1.6 ± 0.4	1.8 ± 1.0	2.8 ± 1.8	0.10	<0.001	<0.001
Optic nerve head parameters						
Disc area, mm^2	2.39 ± 0.39	2.43 ± 0.52	2.16 ± 0.60	0.63	0.03	0.02
Cup area, mm^2	0.71 ± 0.49	1.56 ± 0.53	1.52 ± 0.60	<0.001	<0.001	0.72
Rim area, mm^2	1.68 ± 0.45	0.87 ± 0.44	0.64 ± 0.54	<0.001	<0.001	0.04
Cup-disc area ratio	0.27 ± 0.18	0.65 ± 0.18	0.71 ± 0.21	<0.001	<0.001	0.13
Horizontal CDR	0.54 ± 0.26	0.86 ± 0.12	0.88 ± 0.13	<0.001	<0.001	0.39
Vertical CDR	0.48 ± 0.22	0.80 ± 0.12	0.86 ± 0.14	<0.001	<0.001	0.06
Retinal nerve fiber layer thickness parameters						
Temporal quadrant, μm	75.8 ± 8.5	73.8 ± 9.0	71.7 ± 12.1	0.18	0.05	0.34
Superior quadrant, μm	134.4 ± 16.5	129.7 ± 16.6	117.2 ± 20.6	0.14	<0.001	0.002
Nasal quadrant, μm	83.6 ± 14.0	79.9 ± 12.0	72.2 ± 12.1	0.12	<0.001	0.003
Inferior quadrant, μm	139.6 ± 16.5	129.5 ± 17.2	112.1 ± 23.1	0.002	<0.001	<0.001
Average thickness, μm	108.4 ± 10.1	103.2 ± 10.4	92.3 ± 14.9	0.01	<0.001	<0.001
Ganglion cell complex parameters						
Average, μm	96.5 ± 6.7	93.6 ± 8.2	88.0 ± 10.1	0.05	<0.001	0.003
Superior, μm	95.7 ± 6.7	94.0 ± 9.2	89.5 ± 9.5	0.27	<0.001	0.02
Inferior, μm	97.4 ± 7.2	93.3 ± 8.5	86.6 ± 12.5	0.01	<0.001	0.003
FLV, %	0.42 ± 1.2	1.5 ± 1.7	4.0 ± 3.3	0.001	<0.001	<0.001
GLV, %	2.4 ± 4.5	6.3 ± 4.6	11.6 ± 8.0	<0.001	<0.001	<0.001
RMS	0.08 ± 0.03	0.09 ± 0.04	0.13 ± 0.04	0.15	<0.001	<0.001

SI, superior minus inferior. All values are reported as the mean ± SD.

* Comparison between control cohorts 1 and 2.

† Comparison between control cohort 1 and early glaucoma group.

‡ Comparison between control cohort 2 and early glaucoma group.

Table 2. Diagnostic Accuracy of SD-OCT Parameters in Discriminating Glaucomatous Eyes from Eyes with No Findings Suspicious for Glaucoma (Control Cohort 1) and from Eyes with Nonglaucomatous, but Suspect, Optic Nerves (Control Cohort 2)

	Control Cohort 1			Control Cohort 2			
Parameters	**AUC**	**Sensitivity at 95% Specificity (%)**	**Sensitivity at 80% Specificity (%)**	**AUC**	**Sensitivity at 95% Specificity (%)**	**Sensitivity at 80% Specificity (%)**	***P****
Optic nerve head parameters							
Cup area	0.916 (0.838–0.961)	72.3	87.7	0.608 (0.486–0.740)	18.5	40.0	<0.001
Rim area	0.916 (0.844–0.964)	72.3	87.7	0.608 (0.485–0.740)	18.5	40.0	<0.001
Cup disc area ratio	**0.931 (0.863–0.975)**	78.5	89.2	0.600 (0.443–0.732)	10.8	43.1	<0.001
Horizontal CDR	0.915 (0.843–0.965)	60.0	84.6	0.575 (0.431–0.712)	13.8	30.8	<0.001
Vertical CDR	0.920 (0.848–0.969)	75.4	87.7	0.621 (0.477–0.756)	10.8	47.8	<0.001
Retinal nerve fiber layer thickness parameters							
Temporal quadrant	0.627 (0.500–0.728)	21.5	44.6	0.573 (0.447–0.702)	12.3	27.7	0.54
Superior quadrant	0.705 (0.591–0.813)	26.2	58.5	0.663 (0.540–0.776)	21.5	41.5	0.60
Nasal quadrant	0.707 (0.590–0.810)	20.0	49.2	0.676 (0.542–0.791)	18.5	35.4	0.71
Inferior quadrant	**0.812 (0.713–0.891)**	47.7	67.7	0.718 (0.586–0.825)	38.5	50.8	0.21
Average thickness	0.784 (0.679–0.877)	43.1	64.6	0.714 (0.595–0.832)	32.3	60.0	0.38
Ganglion cell complex parameters							
Average	0.740 (0.620–0.851)	40.0	60.0	0.642 (0.521–0.764)	40.0	46.2	0.23
Superior	0.685 (0.559–0.807)	27.7	44.6	0.586 (0.456–0.712)	29.2	38.5	0.26
Inferior	0.743 (0.621–0.842)	44.6	60.0	0.629 (0.494–0.759)	36.9	44.6	0.18
FLV	**0.881 (0.768–0.948)**	53.8	84.6	0.684 (0.544–0.814)	38.5	52.3	0.02
GLV	0.841 (0.718–0.916)	46.2	75.4	0.671 (0.536–0.783)	41.5	47.7	0.04
RMS	0.763 (0.648–0.854)	32.3	60.0	0.706 (0.551–0.825)	24.6	55.4	0.52

SI: superior minus inferior.

* Comparison of the AUCs with control cohorts 1 and 2.

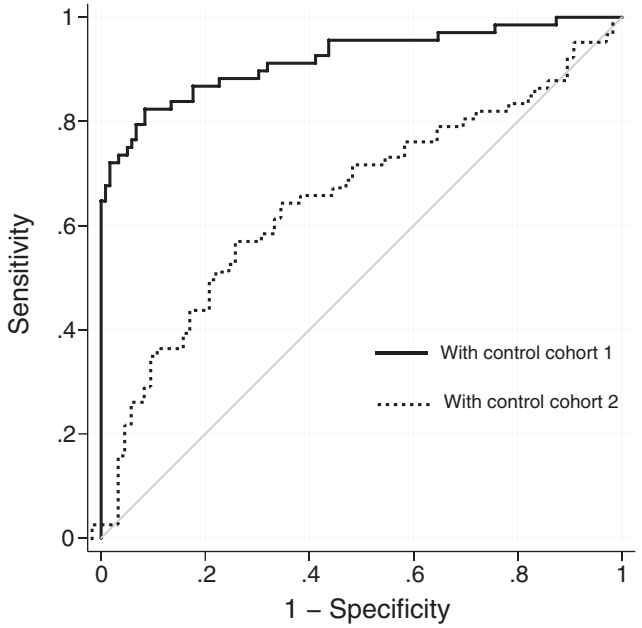

FIGURE 1. ROCs of ONH rim area in discriminating glaucomatous eyes with no findings suspicious for glaucoma (control cohort 1) and eyes with nonglaucomatous, but suspect, optic nerves (control cohort 2).

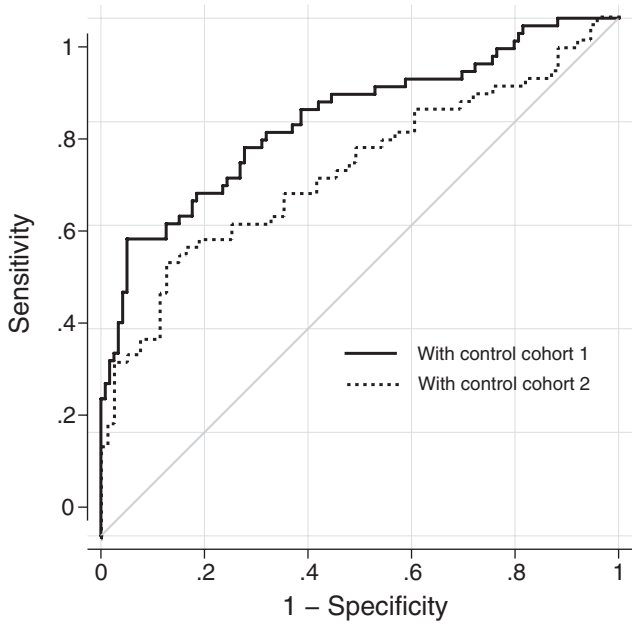

FIGURE 2. ROCs of average RNFL thickness in discriminating glaucomatous eyes from those with no findings suspicious for glaucoma (control cohort 1) and from eyes with nonglaucomatous, but suspect, optic nerves (control cohort 2).

the comparison between the AUCs of SD-OCT parameters in differentiating glaucomatous eyes from eyes of control cohorts 1 and 2. AUCs of ONH parameters discriminating glaucomatous eyes from eyes of control cohort 2 were significantly lesser than those discriminating glaucomatous eyes from eyes of control cohort 1. Figure 1 shows the ROC curves of the ONH rim area discriminating glaucomatous eyes from eyes of control cohorts 1 and 2. AUCs of RNFL parameters discriminating glaucomatous eyes from eyes of control cohorts 1 and 2 were similar. Figure 2 shows the ROC curves of the average RNFL parameter discriminating glaucomatous eyes from eyes of control cohorts 1 and 2. AUCs of GCC FLV and GLV discriminating glaucomatous eyes from eyes of control cohort 2 were significantly lesser than those discriminating glaucomatous eyes from eyes of control cohort 1, whereas those of GCC RMS and other GCC thickness measurements were similar. Figure 3 shows the ROC curves of GCC FLV for discriminating glaucomatous eyes from eyes of control cohorts 1 and 2.

Tables 3 and 4 show the LRs associated with the diagnostic categorization of the parameters when control cohorts 1 and 2, respectively, were used. The outside-normal-limits category of ONH parameters, which were associated with large effects on the posttest probability of disease when control cohort 1 was used, were associated with small effects when control cohort 2 was used. The outside-normal-limits category of RNFL parameters that were associated with large effects on the posttest probability of disease when control cohort 1 was used was associated with moderate effects when control cohort 2 was used. The outside-normal-limits category of GCC parameters was associated with moderate to large effects on the posttest probability of disease in control cohorts 1 and 2. The within-normal-limits category of all SD-OCT parameters was associated with small or no effects on the posttest probability of disease in controls cohorts 1 and 2.

DISCUSSION

This study showed the effect of the control group on the diagnostic accuracy of SD-OCT in glaucoma. The effectiveness of most parameters of SD-OCT in detecting glaucoma significantly decreased when evaluated against a clinically relevant control group with suspicious-looking optic nerves compared with that against a control group consisting of normal subjects with no findings suspicious for glaucoma.

When we evaluated the diagnostic accuracies of SD-OCT parameters in discriminating glaucoma patients with early visual field loss from a control cohort consisting of normal subjects with no findings suspicious for glaucoma, we found

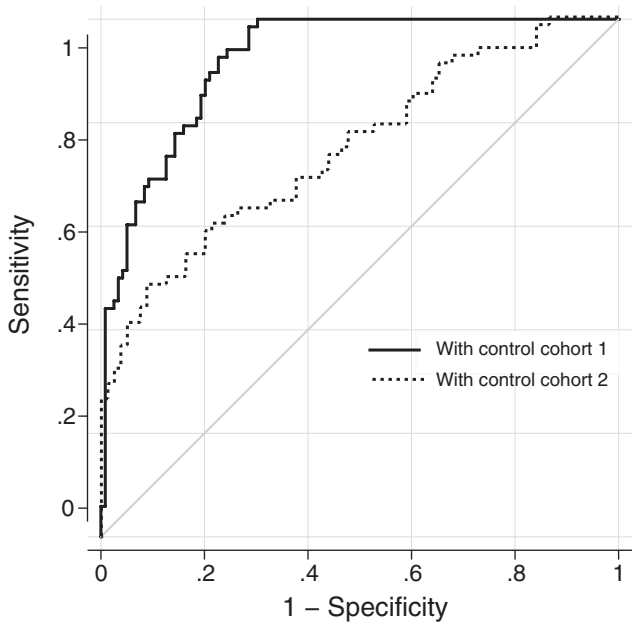

FIGURE 3. ROCs of GCC FLV in discriminating glaucomatous eyes from those with no findings suspicious for glaucoma (control cohort 1) and from those with nonglaucomatous, but suspect, optic nerves (control cohort 2).

TABLE 3. LRs (with 95% CI) of the Normative Database Classification of SD-OCT Parameters in Discriminating Glaucomatous Eyes from Eyes with No Findings Suspicious for Glaucoma (Control Cohort 1)*

Parameter	Within Normal Limits	Borderline	Outside Normal Limits
Cup area	0.38 (0.28–0.53)	12.82 (4.98–33.01)	147 (9.2–2357)
Rim area	0.42 (0.30–0.58)	12.82 (4.98–33.01)	7.01 (3.8–12.9)
Cup disc area ratio	0.41 (0.30–0.56)	30.8 (1.81–525)	143 (8.91–2288)
Average RNFL thickness	0.63 (0.51–0.77)	3.20 (0.97–10.53)	11.4 (4.14–31.4)
Superior RNFL thickness	0.77 (0.67–0.89)	3.20 (0.97–10.53)	13.9 (3.26–59.0)
Inferior RNFL thickness	0.57 (0.45–0.71)	0.92 (0.29–2.94)	24.6 (6.05–99.8)
Average GCC thickness	0.76 (0.66–0.88)	3.05 (1.16–8.01)	55.0 (3.33–908)
Superior GCC thickness	0.84 (0.75–0.95)	6.41 (1.37–29.96)	10.10 (2.28–44.5)
Inferior GCC thickness	0.64 (0.52–0.78)	4.88 (1.34–17.76)	42.7 (5.90–310)
FLV	0.50 (0.37–0.66)	0	4.33 (2.58–7.25)
GLV	0.60 (0.47–0.76)	2.09 (0.79–5.50)	5.11 (2.65–9.84)

* Analysis based on number of eyes.

that the ONH, RNFL, and GCC parameters with the best AUCs were cup-to-disc area ratio, inferior quadrant RNFL thickness, and GCC FLV. Estimates of AUCs obtained in this analysis are similar to those reported in other studies that included patients with early glaucomatous visual field defects and normal subjects, with those with no findings suspicious for glaucoma serving as controls.[13,28] However, the diagnostic accuracy of ONH parameters appears to be better in Indian eyes with early glaucoma compared with that reported in a study by Rao et al.[13] which included participants of Caucasian and African descent. The true diagnostic ability of a test, however, is established only when evaluated in real-life situations. In day-to-day clinical practice, diagnostic tests are used to rule in or rule out the presence of disease in patients suspected of having glaucoma, but not in patients with clearcut signs of the disease or in subjects without any suspicious findings. In fact, a study design similar to the one that we used has been shown to substantially overestimate the performance of the test.[14,29]

Therefore, we evaluated the performance of SD-OCT in detecting glaucomatous from control cohorts consisting of subjects referred by general ophthalmologists as glaucoma suspects based on the presence of a suspect appearance of the optic nerve. These subjects of control cohort 2, however, were diagnosed by the glaucoma experts as being normal subjects with physiological variations of the optic nerve. In other words, the subjects were not true glaucoma suspects, but still were the ones who caused a diagnostic uncertainty among general ophthalmologists. There are bound to be differences in the optic nerve evaluation by general ophthalmologists and glaucoma experts, and earlier reports have shown better evaluation of optic nerves by glaucoma experts compared with general ophthalmologists.[30,31] When we evaluated the ability of SD-OCT to discriminate patients with early glaucomatous visual field defects from the subjects of control cohort 2, we found that the diagnostic effectiveness of all ONH parameters was significantly lesser than that in normal subjects with no findings suspicious for glaucoma. This result is similar to the that of the study by Medeiros et al.,[14] who evaluated the diagnostic ability of the Heidelberg retina tomograph (HRT; Heidelberg Engineering, Heidelberg, Germany) in glaucoma suspects. However, there are two differences between our study design and that of Medeiros et al. Subjects of control cohort 2 in our study were those referred as glaucoma suspects by general ophthalmologists, but diagnosed as normal by glaucoma specialists, and cases were those with early glaucomatous visual field defects. Medeiros et al. recruited glaucoma suspects who had been observed over a period with stereophotographs; those eyes determined by expert graders to be progressing were cases, whereas those determined to be nonprogressing were controls.[14] The decreased diagnostic accuracy of the ONH parameters in our control cohort 2 was expected and is probably related to the subjects' suspect optic nerves, which narrowed down the differences in ONH parameters and decreased discriminatory performance between control cohort 2 and the early-glaucoma cohort.

The AUCs of RNFL parameters in control cohort 2, although not significantly different, were smaller than those in control cohort 1. Although the AUCs of RNFL parameters were inferior to ONH and GCC parameters in control cohort 1, they were comparable to that of ONH and GCC parameters in control

TABLE 4. LRs (with 95% CI) of the Normative Database Classification of SD-OCT Parameters in Discriminating Glaucomatous Eyes from Eyes with Nonglaucomatous, but Suspect, Optic Nerves (Control Cohort 2)*

Parameter	Within Normal Limits	Borderline	Outside Normal Limits
Cup area	0.61 (0.42–0.90)	0.63 (0.27–1.48)	1.63 (1.12–2.37)
Rim area	0.61 (0.42–0.90)	0.63 (0.27–1.48)	1.63 (1.12–2.37)
Cup disc area ratio	0.55 (0.39–0.77)	0.94 (0.39–2.24)	2.28 (1.44–3.59)
Average RNFL thickness	0.66 (0.54–0.83)	0.94 (0.39–2.24)	5.09 (2.06–12.60)
Superior RNFL thickness	0.78 (0.67–0.91)	0.91 (0.36–2.31)	8.09 (1.92–34.10)
Inferior RNFL thickness	0.65 (0.51–0.83)	0.43 (0.14–1.29)	3.20 (1.64–6.24)
Average GCC thickness	0.78 (0.67–0.90)	1.95 (0.75–5.07)	16.5 (2.23–123)
Superior GCC thickness	0.84 (0.74–0.95)	8.18 (1.03–64.8)	12.90 (1.70–98.1)
Inferior GCC thickness	0.66 (0.54–0.81)	1.04 (0.43–2.54)	8.23 (2.59–26.2)
FLV	0.54 (0.40–0.73)	0	2.77 (1.68–4.56)
GLV	0.57 (0.45–0.73)	0.94 (0.39–2.24)	10.0 (3.21–31.4)

* Analysis based on number of eyes.

cohort 2. The result suggests that the AUCs of RNFL parameters are less affected by the physiologic variability of optic nerves, and the outcome is similar to the diagnostic accuracy of the RNFL parameters of HRT in the study by Medeiros et al.[14]

Although the diagnostic accuracies of GCC thickness parameters in control cohort 2 were comparable to that in control cohort 1, AUCs of GCC FLV and GLV were significantly lesser in control cohort 2. This finding is probably related to the manner in which FLV and GLV are calculated.[6,19] GLV is calculated from the fractional deviation map, which is the map of the percentage of GCC thickness decrease at each pixel location compared with the expected or normal value at each pixel determined by the instrument's built-in normative database. FLV detects focal loss using a pattern deviation map, which is a map of the difference between the normalized map (calculated by dividing the GCC thickness values at each location by the average GCC thickness value from the entire map) of an individual and the average normalized map of the instrument's built-in normative database. In short, both FLV and GLV are calculated after comparison with the instrument's built-in normative database. The instrument's normative database is likely to be very similar to the subjects of control cohort 1. Significant differences in the FLV and GLV between the subjects of control cohorts 1 and 2 may have driven the values in control cohort 2 closer to that in the glaucoma cohort, thereby decreasing the discriminating ability.

Diagnostic tests are usually summarized in terms of sensitivity, specificity, AUC, and LR. The advantages and limitations of these measures are described elsewhere.[10,32,33] LR appears to be higher than the previous measures in hierarchy. It expresses the magnitude by which the probability of a diagnosis in a given patient is modified by the results of the test. In other words, the LR indicates how much a given diagnostic test result will raise or lower the pretest probability of the disease in question. We therefore evaluated the LRs associated with the diagnostic categorization of SD-OCT parameters when control cohorts 1 and 2 were used. The magnitude of the LRs associated with the outside-normal-limits category of the ONH and RNFL parameters decreased when control cohort 2 was used instead of control cohort 1. LRs associated with the diagnostic categorization of GCC parameters were similar with both control cohorts 1 and 2. Within-normal-limits results of all SD-OCT parameters were associated with either small or no effect on the posttest probability of glaucoma, which would mean that within-normal-limits results were of little use in ruling out the diagnosis.

Our study demonstrates that the estimates of diagnostic accuracy of SD-OCT in glaucoma reported by studies using a control cohort with no findings of a suspect ONH are higher in real-life clinical situations where SD-OCT is used in subjects with suspicious-looking optic nerves. Also, RNFL parameters of SD-OCT are comparatively better than the ONH parameters in diagnosing glaucoma if the suspicion is based on the appearance of the optic disc. The results, however, do not contradict the importance of imaging tests in suspicious-looking optic nerves, because information from the imaging tests (such as the magnitude of LRs, for example), though associated with small effects on posttest probability, may be relevant and useful, depending on other clinical information and the pretest probability of disease. Our results also do not contradict the utility of imaging tests in clinical situations where glaucoma is suspected, because of findings other than suspect optic nerves.

A limitation of the present study is the possible inclusion of a few preperimetric glaucoma cases in control cohort 2, which may have caused a reduction in diagnostic accuracy. A few false-positive results on SD-OCT may actually be true glaucoma cases, which both the glaucoma experts and SAP may have failed to recognize. This limitation, however, applies to all diagnostic accuracy studies in which a cross-sectional design is used, because there is a lack of reference standard for diagnosing glaucoma. It is not possible to rule out the diagnosis of glaucoma completely in suspect eyes unless there is a reasonable period of follow-up. Medeiros et al.[14] have therefore used progressive optic disc change over follow-up examinations as the reference standard for glaucoma. Although control cohort 2 of our study was predominantly based on a cross-sectional evaluation, 12 of the 41 subjects of that cohort had had a follow-up of more than 2 years, with three or more visual field examinations during the follow-up period, and showed no signs of progression both on optic disc examination and visual field perimetry. Another limitation of the study is the presence of significant differences among the cohorts in age and optic disc size. Previous studies have reported age and optic disc size to affect the SD-OCT measurements.[18,34] Although we adjusted for these differences by using statistical techniques, the possibility of a residual bias in an AUC estimate caused by these differences cannot be totally ruled out.

In conclusion, the diagnostic accuracy of SD-OCT in glaucoma is influenced by the control group. When a clinically relevant control group, consisting of subjects with suspicious-looking optic nerves was used instead of a control group consisting of completely normal subjects, the diagnostic accuracy of most of the SD-OCT parameters decreased significantly. There is a definite need for a clinically relevant control group in glaucoma imaging devices.

References

1. Nassif N, Cense B, Park B, et al. In vivo high-resolution video-rate spectral-domain optical coherence tomography of the human retina and optic nerve. *Opt Express.* 2004;12:367–376.
2. Wojtkowski M, Srinivasan V, Ko T, Fujimoto J, Kowalczyk A, Duker J. Ultrahigh-resolution, high-speed, Fourier domain optical coherence tomography and methods for dispersion compensation. *Opt Express.* 2004;12:2404–2422.
3. Leung CK, Cheung CY, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology.* 2009;116:1257–1263.
4. Knight OJ, Chang RT, Feuer WJ, Budenz DL. Comparison of retinal nerve fiber layer measurements using time domain and spectral domain optical coherent tomography. *Ophthalmology.* 2009;116:1271–1277.
5. Sung KR, Kim DY, Park SB, Kook MS. Comparison of retinal nerve fiber layer thickness measured by Cirrus HD and Stratus optical coherence tomography. *Ophthalmology.* 2009;116:1264–1270.
6. Rao HL, Zangwill LM, Weinreb RN, Sample PA, Alencar LM, Medeiros FA. Comparison of different spectral domain optical coherence tomography scanning areas for glaucoma diagnosis. *Ophthalmology.* 2010;117:1692–1699.
7. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med.* 1978;299:926–930.
8. Willis BH. Spectrum bias: why clinicians need to be cautious when applying diagnostic test studies. *Fam Pract.* 2008;25:390–396.
9. Medeiros FA, Zangwill LM, Bowd C, Sample PA, Weinreb RN. Influence of disease severity and optic disc size on the diagnostic performance of imaging instruments in glaucoma. *Invest Ophthalmol Vis Sci.* 2006;47:1008–1015.
10. 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;48:2653–2660.
11. Garudadri CS, Rao HL, Parikh RS, et al. Effect of optic disc size and disease severity on the diagnostic capability of glaucoma imaging

technologies in an Indian population. *J Glaucoma.* Published online April 25, 2011.

12. Leite MT, Zangwill LM, Weinreb RN, et al. Effect of disease severity on the performance of Cirrus spectral-domain OCT for glaucoma diagnosis. *Invest Ophthalmol Vis Sci.* 2010;51:4104–4109.
13. Rao HL, Leite MT, Weinreb RN, et al. Effect of disease severity and optic disc size on diagnostic accuracy of RTVue spectral domain optical coherence tomograph in glaucoma. *Invest Ophthalmol Vis Sci.* 2011;52:1290–1296.
14. Medeiros FA, Ng D, Zangwill LM, Sample PA, Bowd C, Weinreb RN. 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; 48:214–222.
15. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol.* 2009;62:5–12.
16. Hodapp E, Parrish RK, Anderson DR, eds. *Clinical Decisions in Glaucoma*. St. Louis: Mosby; 1993;53.
17. Anderson DR, Patella VM. *Automated Static Perimetry*. 2nd ed. St. Louis: Mosby; 1999.
18. Rao HL, Kumar AU, Babu JG, Kumar A, Senthil S, Garudadri CS. Predictors of normal optic nerve head, retinal nerve fiber layer, and macular parameters measured by spectral domain optical coherence tomography. *Invest Ophthalmol Vis Sci.* 2011;52: 1103–1110.
19. Tan O, Chopra V, Lu AT, et al. Detection of macular ganglion cell loss in glaucoma by Fourier-domain optical coherence tomography. *Ophthalmology.* 2009;116:2305–2314.
20. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford, UK: Oxford University Press; 2003;130–167.
21. Glynn RJ, Rosner B. Accounting for the correlation between fellow eyes in regression analysis. *Arch Ophthalmol.* 1992;110:381–387.
22. Zhou XH, Obuchowski NA, McClish DK. Analysis of correlated ROC data. In: Zhou XH, Obuchowski NA, McClish DK, eds. *Statistical Methods in Diagnostic Medicine.* New York: John Wiley & Sons, Inc.; 2002;274–306.
23. Alonzo TA, Pepe MS. Distribution-free ROC analysis using binary regression techniques. *Biostatistics.* 2002;3:421–432.
24. Zhou XH, Obuchowski NA, McClish DK. Comparing the accuracy of two diagnostic tests. In: Zhou XH, Obuchowski NA, McClish DK, eds. *Statistical Methods in Diagnostic Medicine*. New York: John Wiley & Sons; 2002;165–194.
25. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44: 837–845.
26. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA.* 1994;271:703–707.
27. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol.* 1991;44:763–770.
28. Seong M, Sung KR, Choi EH, et al. Macular and peripapillary retinal nerve fiber layer measurements by spectral domain optical coherence tomography in normal-tension glaucoma. *Invest Ophthalmol Vis Sci.* 2010;51:1446–1452.
29. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA.* 1999;282: 1061–1066.
30. Abrams LS, Scott IU, Spaeth GL, Quigley HA, Varma R. Agreement among optometrists, ophthalmologists, and residents in evaluating the optic disc for glaucoma. *Ophthalmology.* 1994;101:1662–1667.
31. Vessani RM, Moritz R, Batis L, Zagui RB, Bernardoni S, Susanna R. Comparison of quantitative imaging devices and subjective optic nerve head assessment by general ophthalmologists to differentiate normal from glaucomatous eyes. *J Glaucoma.* 2009;18:253–261.
32. Langlotz CP. Fundamental measures of diagnostic examination performance: usefulness for clinical decision making and research. *Radiology.* 2003;228:3–9.
33. Medeiros FA, Zangwill LM, Bowd C, Weinreb RN. Comparison of the GDx VCC scanning laser polarimeter, HRT II confocal scanning laser ophthalmoscope, and stratus OCT optical coherence tomograph for the detection of glaucoma. *Arch Ophthalmol.* 2004;122: 827–837.
34. Bendschneider D, Tornow RP, Horn FK, et al. Retinal nerve fiber layer thickness in normals measured by spectral domain OCT. *J Glaucoma.* 2010;19:475–482.