Diagnostic Accuracy of the Optical Coherence Tomography in Assessing Glaucoma Among Filipinos. Part 1: Categorical Outcomes Based on a Normative Database

Noel de Jesus Atienza, MD, MSc and Joseph Anthony Tumbocon, MD

International Eye Institute
St. Luke's Medical Center
Quezon City, Philippines

Correspondence: Noel Atienza, MD, MSc
St. Luke's Medical Center
Quezon City, Philippines
Email: njatienza@yahoo.com

ABSTRACT

Objective: To determine the accuracy of the optic nerve head (ONH) and retinal nerve fiber layer (RNFL) parameters using the Stratus OCT in diagnosing glaucoma among Filipino patients and to determine the validity of the OCT measurements compared with a reference standard.

Methods: Glaucoma suspects were recruited consecutively from patients undergoing diagnostic tests. The results of the Stratus OCT fast RNFL protocol were analyzed against the independent assessment by glaucoma experts who were blinded. Outcomes included RNFL thickness values of ≤5 percentile and ≤1 percentile of the normative database for each RNFL sector, quadrant, and for the average RNFL thickness. Estimates of diagnostic accuracy with 95% confidence intervals were calculated.

Results: A total of 119 subjects assessed as glaucoma and 397 subjects assessed as normal were included in the analysis. Using the best categorical criterion for abnormality (average RNFL thickness of ≤1 percentile of the normative database), the OCT had a specificity of 98% (95% CI 96 - 99) and a sensitivity of 37% (95% CI 28 - 46). The criterion with the highest sensitivity was the presence of at least one sector with thickness at ≤5 percentile.

Conclusion: The Stratus OCT demonstrated a specificity of 98% and a sensitivity of 37% for the diagnosis of glaucoma using as a criterion an average RNFL thickness of ≤1 percentile of the normative database. It was a poor screening test for glaucoma suspects, although it showed some promise as a confirmatory test.

Keywords: Glaucoma, Glaucoma suspects, Optical coherence tomography, Glaucoma imaging

Glaucomatous optic nerve damage is a result of retinal ganglion cell (RGC) death with progressive loss of axons located in the retinal nerve fiber layer (RNFL). Several clinical studies showed that optic nerve head (ONH) damage and thinning of the RNFL occur earlier than the appearance of abnormalities in the visual field.[1] The European Glaucoma Society (EGS) stated in the 2008 Clinical Practice Guidelines that at least 50% of patients with glaucoma remain undiagnosed while more than 50% of patients currently receiving treatment for glaucoma do not actually have glaucoma. The EGS cited the need to improve on the sensitivity and specificity of diagnostic tests for glaucoma.[2]

The newer diagnostic modalities, such as the optical coherence tomography (OCT), are primarily directed at demonstrating the presence of decreased thickness of the retinal nerve fiber layer (RNFL) around the optic nerve head in glaucoma patients. The OCT is an accurate and reproducible method that measures and analyzes RNFL thickness and ONH parameters to help differentiate glaucomatous eyes from normal eyes.

In its most recent technology assessment published in 2007, the American Academy of Ophthalmology reviewed the evidence from articles related to the diagnosis of glaucoma using the OCT and other imaging devices.[3] A review of 159 articles on diagnostic studies showed that these were mostly Phase II studies with no independent masked comparison with a gold standard. The current best evidence from diagnostic studies on the OCT for glaucoma was level II evidence. There was no level I evidence from current published literature due to the lack of a masked independent comparison on a set of consecutive subjects that represented the target population for OCT testing.[3]

The early models of the OCT had a dilemma of deriving the abnormal value using the quantitative data available from the OCT. A study by Budenz using the fast RNFL protocol on 109 normal and 63 glaucoma subjects determined the sensitivity and specificity of the Stratus OCT RNFL thickness measurements in diagnosing glaucoma using the standard automatic perimetry as the gold standard.[4] One major source of bias from this study was that the normal subjects did not undergo visual field testing. The researchers relied purely on a complete eye examination and the decision to perform the perimetry was determined by the clinical exam. Sensitivity and specificity estimates were determined without an independent blinded comparison with the reference standard.

The sensitivity and specificity of the different parameters in the Stratus OCT are tabulated in Table 1, showing that it has high specificity and moderate sensitivity. At an average RNFL thickness measurement of <1 percentile, the sensitivity was only 68% and the highest sensitivity achieved using either one quadrant or sector at <5 percentile was 89% (Table 1).

Budenz also recommended that studies on the OCT be performed on specific populations since RNFL differences were noted between ethnic groups. Statistical adjustments were recommended for RNFL parameters in order to provide better sensitivity and specificity for glaucoma detection among these specific target populations.[5]

This study, therefore, determined the accuracy of the ONH and RNFL parameters using the Stratus OCT in diagnosing glaucoma among Filipino suspects. It was a cross-sectional diagnostic validation study with a phase 3 design[6] that analyzed the ability of the OCT to assess patients that represented the target population.

METHODOLOGY

This validation study focused on the OCT parameters using the fast optic disc and fast RNFL protocols of the Stratus OCT machine as applied to glaucoma suspects. A prospective recruitment of glaucoma suspects was undertaken from September

Table 1. Sensitivity and Specificity of the Stratus OCT (Budenz, 2005).[4]

Criteria for Abnormality	Sensitivity (95% CI)	Specificity (95% CI)
Average RNFL <5 percentile	84% (75-93)	98% (96-100)
Average RNFL <1 percentile	68% (57-80)	100%
1 quadrant with RNFL <5 percentile	89% (81-97)	95% (90-99)
1 quadrant with RNFL <1 percentile	83% (73-92)	100%
1 clock hour RNFL <5 percentile	89% (81-97)	92% (87-97)
1 clock hour RNFL <1 percentile	83% (73-92)	100%

2008 to October 2010 at the St. Luke's International Eye Institute. Subjects were enrolled based on the inclusion and exclusion criteria enumerated below.

Inclusion Criteria

Patients were Filipinos aged 20 years or older with best corrected visual acuity of at least 20/40. A "glaucoma suspect" was defined as a patient with probable glaucomatous optic neuropathy based on the presence of any of the following findings on clinical examination:

1. Increased intraocular pressure (IOP ≥23 mmHg by applanation tonometry);
2. Optic cup-to-disc ratio greater than 0.5;
3. Optic cup-to-disc ratio asymmetry of >0.2;
4. Suspicious optic disc findings, such as thinning or notching of the neuroretinal rim, bayoneting of the optic nerve head vessels, and optic disc hemorrhage;
5. Loss of the RNFL reflex ("RNFL drop-out"), especially on the superior and inferior areas of the ONH;
6. A history of iridescent vision, ocular pain and redness accompanied by corneal edema, and a mid-dilated pupil.

Exclusion Criteria

Subjects were excluded when any of the following were present:

1. Best corrected visual acuity worse that 20/40;
2. Presence of eye conditions that can affect the visual fields, such as neuroophthalmologic conditions, diabetic retinopathy, chorioretinitis, and maculopathy;
3. Severe glaucomatous cupping of 1.0 and a tunnel vision with remaining visual field of less than 3 degrees;
4. Previously diagnosed cases of chronic glaucoma with established glaucomatous visual field defects on perimetry.

Sample Size Determination

The sample size was determined based on a previous study that reported an overall sensitivity of 77% and a specificity of 77%[1]. With a margin of error of 5%, and a 95% confidence level, the minimum sample size required to estimate the sensitivity was computed as:

$n = (0.77)(1 – 0.77) [1.96]2/(.05)2$ or $n = 272$ subjects per group.

The same number was needed to estimate specificity. A prevalence of 50% was assumed based on the opinion of glaucoma specialists. The total sample size was estimated at 544 subjects:

$n / 0.5 = 272 / 0.5 = 544$ total subjects

Ethical Considerations

The study was approved by the institutional scientific and ethics board of the St. Luke's Medical Center Q.C. The protocol conformed to the Declaration of Helsinki. An informed consent was secured from all patients which ensured confidentiality of data.

Data Collection

Baseline data collected during the screening included the following: age, gender, refractive error (spherical equivalent), snellen visual acuity, intraocular pressure by applanation tonometry, and chamber angle by gonioscopy.

Patients underwent the following diagnostic tests on the same day:

1. Standard automated perimetry (SAP) (Octopus or Humphrey perimetry);
2. Optical coherence tomography (Stratus OCT Model 3000, Carl Zeiss Meditec, Dublin, CA with Version 4.0.1 Software) using the fast optic disc and fast RNFL (3.4) protocols;
3. Optic nerve head photography with a Zeiss fundus camera and VISUPAC system.

Figure 1 shows the flowchart of diagnostic procedures starting from screening to the different glaucoma tests.

Standard automated perimetry (SAP) was done using either the Octopus 101 (G2 program) or the Humphrey Field Analyzer (central 30-2 test, size III white stimulus, SITA-standard strategy). The fast optic disc protocol included six 4-mm radial linear scans centered on the optic disc and acquired in 1.92 seconds which produced two printouts: the Individual Radial Scan Analysis and the Optic Nerve Head Analysis. The fast RNFL protocol included three 3.4-mm circular scans centered on the optic disc and acquired in 1.92 seconds which produced the Fast RNFL Analysis printout.

The reference standard determination was performed by two glaucoma specialists who were blinded to the OCT results and who independently examined the clinical records, the optic disc images
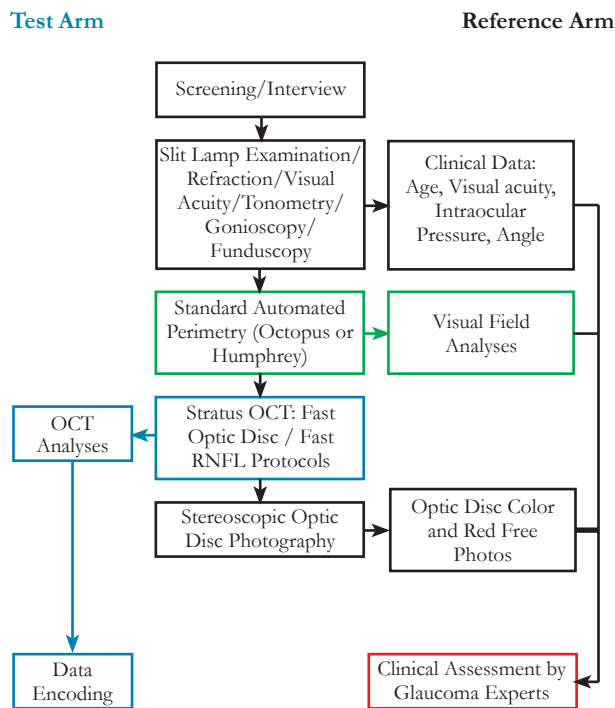
Figure 1. Flowchart of diagnostic procedures.

on standard photos, and the results of the visual field tests. Each assessor gave his assessment without knowledge of the other expert assessment, and a diagnosis of *glaucoma* or *no glaucoma* was made based on a consensus of the two glaucoma specialists. In cases of disagreement, a third glaucoma expert was consulted to resolve the disagreements by an independent examination.

The glaucoma specialists based their assessment of glaucoma on the presence of any of the following features: 1) Optic cupping to the disc margin with associated enlargement of peripapillary atrophy with or without detectable abnormalities on SAP; 2) Abnormalities of the optic nerve head characteristic of glaucomatous excavation, such as notching, disc asymmetry of more than 0.2 between the two eyes, focal or diffuse atrophy of the RNFL, vertical cup/ disc ratio of more than 0.6; and 3) An abnormal visual field on SAP characteristic of glaucoma with a minimum criteria of: a cluster of three or more non-edge points in a location typical for glaucoma, all of which were depressed on pattern deviation plot at p<.05 level and one of which was depressed at p<.01 level. The absence of all of the above conditions was the basis for an assessment of no glaucoma or normal.

One eye per subject was included in the analyses. In cases of unilateral disease, the diseased eye was chosen. In cases where one eye was excluded due to severe glaucoma, the other eye was included. In all other cases of bilateral glaucoma and for normal subjects, one eye was chosen at random using a computer generated code.

Statistical Analysis: Categorical Outcomes

Categorical outcomes for the OCT were encoded based on the color-coded results from the OCT RNFL printouts. Red highlighted values had RNFL thickness values of ≤1 percentile of the normative database. Yellow highlighted values had thickness values of ≤5 percentile but >1 percentile of the normative database.

Six different criteria for abnormality were defined as a basis for a determination of normal or abnormal OCT results. These were, namely;

1. Average RNFL thickness ≤1 percentile
2. Average RNFL thickness ≤5 percentile
3. The presence of at least one quadrant ≤1 percentile
4. The presence of at least one quadrant ≤5 percentile
5. The presence of at least one clock hour sector ≤1 percentile
6. The presence of at least one clock hour sector ≤5 percentile

Outcome results for the OCT were encoded for each of the six criteria for all subjects. An abnormal value at ≤5 percentile included OCT results color-coded yellow or red. An abnormal value at ≤1 percentile included OCT results color-coded red only. All green-coded results were considered within normal.

The baseline data, the categorical OCT outcomes, and the results of the expert assessment were analyzed using the SPSS version 16.0 software. For the categorical outcomes, estimates of diagnostic accuracy were determined with a 95% confidence interval. The following estimates of diagnostic accuracy were computed: sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratio for a positive result, and the likelihood ratio for a negative result.

RESULTS

A total of 547 glaucoma suspects were screened and tested, and 31 subjects were excluded from the final analysis (Figure 2). The reasons for these exclusions were: 10 were less than 20 years; 10 had findings highly suggestive of a neuroophthalmological condition; 9 had retinal disease and pathologic myopia; 2
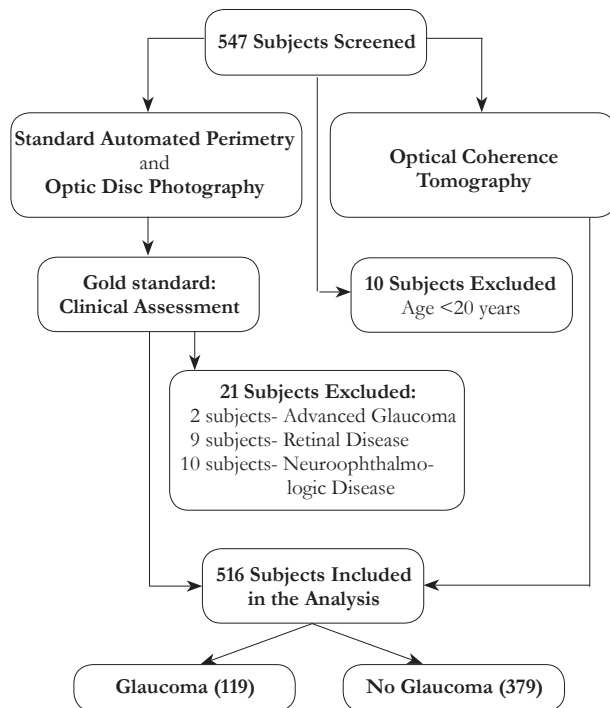
Figure 2. Flowchart of the study recruitment process.

had advanced bilateral glaucoma.

The age of the sample ranged from 20 to 93 years, with a mean of 52.7 years (±12.9). There were 327 female (63.4%) and 189 (36.6%) male subjects. The subjects were also categorized into age brackets similar to the normative database of the OCT. The 40-49, 50-59, and 60-69 year age categories contained 399 subjects which comprised 77.3% of the total sample (Table 1).

All 516 subjects included in the analysis had open anterior chamber angles. There were no cases of acute or chronic angle closure glaucoma. Intraocular pressures were within normal in 471 eyes (91.3%) and were elevated (IOP >22 mmHg) in 45 eyes (8.7%). The refractive errors of the sample had a mean spherical equivalent of -0.07sph (±1.84sph) and median of 0.00sph.

Two glaucoma specialists (AT and JM) examined the datasets of 547 subjects (1094 eyes). There was agreement in the status of 972 eyes (88.8%) and disagreement in 122 eyes. A third glaucoma expert (ML) examined the datasets and gave the final assessment for the 122 eyes.

Of the 516 subjects included in the analyses, 119 eyes (23.1%) were diagnosed to have glaucomatous

optic neuropathy and 397 eyes (76.9%) were classified as normal. The mean age for the glaucoma group was 58.7 years (±10.9) and for the normal group 50.9 years (±12.9); this difference was statistically significant (p=0.03). The glaucoma patients were older than the normal patients by a mean difference of 7.8 years (95% CI:5.2 – 10.3).

Among the 45 eyes with elevated IOP, 19(42.2%) were assessed to have glaucoma. Among the 471 eyes with normal IOP, 100(21.2%) were assessed to have glaucoma. The mean IOP for the glaucoma eyes was 18.1 mmHg (±7.3) and for the normal eyes 16.7 mmHg (±4.2). The t-test for the difference in means was significant with a mean IOP difference of 1.4 mm Hg (95% CI: 0.37 – 2.47).

Table 1. Age Distribution of the Study Sample Divided into Categories by Decades (n=516).

Age Category (years)	Frequency (%)
19 – 29	24 (4.7%)
30 – 39	60 (11.6%)
40 – 49	116 (22.5%)
50 – 59	147 (28.5%)
60 – 69	136 (26.4%)
70 – 79	26 (5.0%)
80 – 94	7 (1.4%)
Total	516

Table 2. Distribution in Age, Gender, and IOP Between the Glaucoma and No Glaucoma Groups.

Characteristic	Glaucoma (%)[a] n=119	No Glaucoma (%)[a] n=397	Totals
Age (years)			
19-29	0 (0%)	24 (100%)	24
30-39	9 (15%)	51 (85%)	60
40-49	15 (13%)	101 (87%)	116
50-59	31 (21%)	116 (79%)	147
60-69	49 (36%)	87 (64%)	136
70-79	13 (50%)	13 (50%)	26
80-94	2 (29%)	5 (71%)	7
Gender			
Male	64 (34%)	125 (66%)	189
Female	55 (17%)	272 (83%)	327
Intraocular pressure			
Elevated IOP	19 (42%)	26 (58%)	45
Normal IOP	100 (21%)	371 (79%)	471

[a] row percentages

The mean refractive error in the glaucoma group was -0.26 sph (±1.73) and in the normal group -0.01 (±1.87). The glaucoma group was slightly more myopic (Mean difference: -0.25 sph, 95% CI: -0.62-0.13) but the difference was not statistically significant (p=0.37).

The average RNFL thickness was the OCT parameter that served as a global index for the fast RNFL protocol. Table 3 shows the 2 x 2 contingency table values for the six different criteria for abnormality. The criterion with the highest sensitivity (or true positive rate) was at least 1 sector at ≤5 percentile with a value of 73%. The criterion with the highest specificity was the average RNFL thickness at ≤1 percentile with a value of 98%.

Table 3. Comparative 2x2 Contingency Table for the Six Different Criteria for Abnormality in RNFL Thickness (N=516).

Criteria for RNFL Thickness Below Normal	Sensitivity (True Positive Rate)	False Positive Rate	False Negative Rate	Specificity (True Negative Rate)
Ave RNFL Thickness ≤1 Percentile	44 (37%)	9 (2%)	75 (63%)	388 (98%)
Ave RNFL Thickness ≤5 percentile	56 (47%)	28 (7%)	63 (53%)	369 (93%)
At least 1 RNFL Quadrant ≤1 Percentile	49 (41%)	18 (5%)	70 (59%)	379 (95%)
At least 1 RNFL Quadrant ≤5 Percentile	64 (54%)	57 (14%)	55 (46%)	340 (86%)
At least 1 RNFL Sector ≤1 Percentile	61 (51%)	35 (9.6%)	58 (49%)	362 (91.2%)
At least 1 RNFL Sector ≤5 Percentile	87 (73%)	91 (23%)	32 (27%)	306 (77%)

Table 4. Statistical Estimates and Predictive Values for the Six Different Criteria for Abnormality in RNFL Thickness (N=516).

OCT Abnormality Criteria	Sensitivity (95% CI)	Specificity (95% CI)	LR Positive (95% CI)	LR Negative (95% CI)
Ave Thickness ≤1 Percentile	37% (28 – 46)	98% (96 – 99)	16.31 (8.20 – 32.43)	0.64 (0.56 – 0.74)
Ave Thickness ≤5 Percentile	47% (38 - 56)	93% (90-95)	6.67 (4.45 – 10.00)	0.57 (0.48 – 0.68)
1 Quadrant RNFL ≤1 Percentile	41% (32 – 50)	95% (93 – 98)	9.08 (5.51 – 14.97)	0.62 (0.53 – 0.72)
1 Quadrant RNFL ≤5 Percentile	54% (45 – 63)	86% (82 – 89)	3.75 (2.8 – 5.02)	0.54 (0.44 – 0.66)
1 Sector RNFL ≤1 Percentile	51% (42 – 60)	91% (88 – 94)	5.81 (4.05 – 8.35)	0.53 (0.44 – 0.64)
1 Sector RNFL ≤5 Percentile	73% (65 – 81)	77% (73 – 81)	3.19 (2.58 – 3.94)	0.35 (0.26 – 0.47)

Sensitivity, specificity, and likelihood ratios were computed for each of the six criteria with 95% confidence intervals (Table 4). Sensitivity values ranged from 37% to 73%, while the specificity values ranged from 77 to 98%. When the criteria for abnormality was ≤1 percentile, the sensitivity was also lower while the specificity became higher when compared with the ≤5 percentile criteria. None of the six criteria exhibited high values for both sensitivity and specificity.

Confidence intervals for the sensitivity were wide in comparison to the narrow confidence intervals of the estimates for specificity. This could be due to the low prevalence of glaucoma (23%) in this study. The sample size was computed with a projected prevalence of 50%, and the narrow intervals for the specificity was due to

the larger percentage of subjects assessed to have no glaucoma.

High likelihood ratios for positive results were seen in the average thickness (LR+ 16.3) and quadrant average (9.08) at ≤1 percentile of the normative database. The outcome with the highest sensitivity of 73% was the presence of at least one clock-hour sector at ≤5% of the normative database, but this was accompanied by a lower specificity of 77%.

DISCUSSION

The Phase 3 diagnostic study design assessed the ability of the OCT to discriminate between glaucoma and normal patients from among glaucoma suspects that best represented the target population. The inclusion and exclusion criteria used for the study were designed to ensure the presence of diagnostic uncertainty. The exclusion of patients with established field defects or with far advanced glaucoma were aimed at producing a population with an "intention to diagnose" glaucoma.

The resulting prevalence of 23% from the sample of glaucoma suspects was lower than the projected 50% used in the determination of the sample size. The relatively low prevalence of glaucoma in the study explained the low accuracy of the estimates of sensitivity and positive predictive value. In contrast, the larger number of normal subjects resulted in narrow confidence intervals for the specificity and the negative predictive values of the OCT.

Accuracy of the OCT in the Diagnosis of Glaucoma

The most salient observation derived from the study was a high specificity of 98% and a low sensitivity of 37% when using the categorical outcome of average thickness at ≤1 percentile as the criterion for abnormality. The high specificity of the OCT was accompanied by a low false-positive rate, with only 9 false-positive readings out of 397 normal subjects. In contrast, there was a high false-negative rate of 63% with 75 diagnosed glaucoma subjects classified as normal based on the stop-sign color scheme of the Stratus OCT.

Considering that the stoplight color coding was applied to the sectors, quadrants, and the average thickness, the OCT will at best be able to confirm glaucoma in three out of four cases. Using the most sensitive criterion (at least one RNFL sector at ≤5 percentile), the OCT will falsely classify 1 in 4 glaucoma cases as being normal and will also misdiagnose 1 in 4 normal subjects as having glaucoma.

The SpPin (Specificity is so high that a Positive result rules in the diagnosis) principle applies to the OCT because with its high specificity a positive result will virtually rule-in the disease. On the other hand, the SnNout (Sensitivity is so high that a Negative result rules out the diagnosis) principle cannot be applied to the OCT because with a low sensitivity a negative result may not rule out the disease. Thus, the OCT is a poor screening test for glaucoma suspects, although it showed some promise as a confirmatory test.

The sensitivity estimates found in our study were much lower than those from previous studies by Hougard (72% sensitivity and 95% specificity)[7] and Budenz (68% sensitivity and 100% specificity)[6]. The decrease in performance of the OCT in our study was partly because of the inclusion criteria that focused on glaucoma suspects. The studies by Hougard and Budenz were Phase 2 studies on known normal and glaucomatous eyes. The normal subjects used in these studies had no suspicious findings of the disease and were generally required to have normal optic disc appearance. In contrast, the normal subjects in our study had suspicious optic discs. A clinician using the OCT software and its stoplight-coded outputs would have difficulty in differentiating the normal subjects from those with definite glaucoma.

The objective of this study was to estimate global test performance using the stoplight color-coded printouts of the OCT. These were based on the normative database derived for the OCT. The estimates of test accuracy and validity may have been affected by the choice of the reference standard. Sensitivity is underestimated most when the prevalence of the condition is low. Miscalculation of the reference standard will tend to result in underestimation of test accuracy.

The gold standard used in this study placed much importance on expert clinical assessment of the optic nerve head and the SAP. The criteria for a positive diagnosis of glaucoma include a combination of structural and functional evidence such that a diagnosis of glaucoma could be made with certainty with or without an accompanying visual field defect. The visual field defect must also be typically

glaucomatous. Thus, the global indices were less useful for this study.

There is no widely accepted gold standard for the diagnosis of glaucoma.[4] In future researches, it is essential that a gold standard for the definition of glaucoma be established. One possible gold standard would be the clinical evidence of progression of glaucomatous damage. Much research has been done on repeated visual field assessment to document evidence of progression in field damage. It is possible that the Stratus OCT and other imaging devices, such as the HRT II and the Cirrus OCT, will be used in future research in glaucoma diagnostics.

In summary, glaucoma suspects undergoing the OCT cannot be assessed for the presence of glaucoma based purely on the results of the OCT. The Stratus OCT using the fast RNFL protocol with its internal software gave categorical results with high specificity but low sensitivity. At the present time, the Stratus OCT cannot replace the gold standard of clinical assessment of structural and functional damage in the diagnosis of glaucoma. Because of its high specificity, the Stratus OCT may be used as a confirmatory test. With its low sensitivity, the OCT may not be useful as a screening test for glaucoma suspects.

REFERENCES

1. American Academy of Ophthalmology. Ophthalmic Procedures Assessment: Optic nerve head and retinal nerve fiber layer analysis - a report by the American Academy of Ophthalmology. *Ophthalmology* 1999;106:1414-1424.
2. European Glaucoma Society (EGS). Terminologies and Guidelines for Glaucoma, 3rd ed, 2008. Downloadable from www.eugs.org.
3. American Academy of Ophthalmology. Ophthalmic Procedures Assessment: Optic nerve head and retinal nerve fiber layer analysis: a report by the American Academy of Ophthalmology. *Ophthalmology* 2007;114:1937–1949.
4. Budenz DL, Michael A, Chang RT, et al. Sensitivity and specificity of the Stratus OCT for perimetric glaucoma. *Ophthalmology* 2005;112:3–9.
5. Budenz DL, Anderson DR, Varma R, et al. Determinants of normal retinal nerve fiber layer thickness measured by Stratus OCT. *Ophthalmology* 2007;114:1046-52.
6. Sackett DL, Haynes RB. Evidence base of clinical diagnosis: The architecture of diagnostic research. *Br Med J* 2002; 321:539-541.
7. Hougaard JL, Heijl A, Bengtsson B. Glaucoma detection by Stratus OCT. *J Glaucoma* 2007; 16:302-306.