Computer Simulation of Progressive Retinal Nerve Fiber Layer Loss in Glaucoma: Performance of Event and Trend Analyses

Marco Yu,[1] *Robert N. Weinreb,*[2] *Cedric Yiu,*[3] *Shu Liu,*[1] *Ming Keung Or,*[3] *Cong Ye,*[1] *Dennis Shun-Chiu Lam,*[1] *and Christopher Kai-Shun Leung*[1]

Purpose. Although event analysis (EA) and trend analysis (TA) have been widely adopted to evaluate glaucoma progression in clinical trials, there is poor agreement between the strategies and no consensus on strategy selection in clinical practice. With computer simulation of progressive loss of the retinal nerve fiber layer (RNFL), the authors compared the performance of TA and EA for the detection of glaucoma progression.

Methods. RNFL progression was modeled with reference to the individual's test-retest variability and the pattern and rate of progression. The sensitivity and specificity of each scenario were computed from 5000 simulated datasets. Simulation results were validated with longitudinal RNFL measurements obtained from 107 glaucoma and glaucoma suspect patients who had a median follow-up period of 38 months.

Results. TA generally attained a sensitivity ≥80% earlier than EA, although EA with a group reproducibility coefficient had a higher sensitivity than TA for eyes with a large test-retest variability in the early follow-up period, albeit at a lower specificity. The specificity of TA was 95% and ranged between 80% and 100% for EA. Independent of test-retest variability and the pattern and rate of progression, TA had an accuracy ≥80% earlier than EA. In the longitudinal study, the detection rate was 42%, 35%, and 3% for TA, whereas it was 11% to 40%, 12% to 28%, and 3% to 23% for EA at 36 months of follow-up in eyes with small, average, and large test-retest variabilities, respectively.

Conclusions. Although test-retest variability is an important determinant in progression analysis, TA generally outperformed EA for the detection of RNFL progression in glaucoma. (*Invest Ophthalmol Vis Sci.* 2011;52:9674–9683) DOI:10.1167/iovs.11-8052

Glaucoma is a neurodegenerative disease characterized by progressive degeneration of retinal ganglion cells, with gradual deterioration of visual function.[1] As with other neurodegenerative disorders, monitoring disease progression is essential for formulating a treatment plan and evaluating disease prognosis. Detection of glaucoma progression has been largely based on assessment of serial changes of the optic disc with stereophotographs and of visual field changes with standard automated perimetry. With the advent of noninvasive imaging technologies, including optical coherence tomography,[2] scanning laser polarimetry[3] and confocal scanning laser ophthalmoscopy,[4] it is now feasible to measure the optic disc and the retinal nerve fiber layer (RNFL) with high repeatability and reproducibility,[5–13] thereby facilitating the detection of change over time.[14–18] A major challenge remains, however, in discriminating genuine disease progression from inherent measurement variability.

Two strategies, event-based analysis and trend-based analysis, have been devised to detect disease progression. In event analysis (EA), progression is commonly defined when the difference of a parameter of interest between the baseline and the follow-up visits is greater than the test-retest variability (or the reproducibility coefficient). In trend analysis (TA), progression is commonly defined when a significant negative trend is detected with linear regression between the parameter of interest and time. The key outcome measure in major clinical trials in glaucoma treatment has been largely based on EA of visual field measurements. In the Advanced Glaucoma Intervention Study (AGIS) and the Collaborative Initial Glaucoma Treatment Study (CIGTS), 20-interval (0 = normal, 20 = advanced visual field loss) visual field scoring systems were developed based on the extent and the depth of defects in total deviation data in the visual field printouts.[19–21] In AGIS, 95% of glaucomatous eyes showed a test-retest variation of up to 3-scale intervals.[19] Progression was thus defined as an increase in visual field score ≥4 confirmed with two consecutive examinations. In CIGTS, progression was defined when a score increase ≥3 was detected with two consecutive examinations.[21] In the Early Manifest Glaucoma Trial (EMGT), analysis was evaluated at individual test locations.[22] A change in visual sensitivity at a test location was considered significant when it was outside the 95% test-retest distribution interval. Progression was defined if ≥3 test locations showed significant deterioration and was confirmed by two subsequent tests.

Trend analysis was less popular in clinical trials for glaucoma treatment. Nevertheless, a number of statistical packages performing linear regression of mean deviation (MD), visual field index (VFI) (Guided Progression Analysis, Carl Zeiss Meditec, Dublin, CA), pointwise threshold data (Progressor, OBF Laboratories UK Ltd, Wiltshire, UK), and RNFL measurements (Guided Progression Analysis, Carl Zeiss Meditec) with follow-up duration are clinically available to evaluate glaucoma progression. Yet, the agreement of progression detection between event and trend analyses for both function and structure often is poor,[23,24] and there is no consensus regarding the

From the [1]Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong, People's Republic of China; [2]Hamilton Glaucoma Center and the Department of Ophthalmology, University of California, San Diego, California; and the [3]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, People's Republic of China.

Submitted for publication June 14, 2011; revised August 11 and September 19, 2011; accepted October 1, 2011.

Disclosure: **M. Yu**, None; **R.N. Weinreb**, Carl Zeiss Meditec (F, C, R), Optovue (C, R), Heidelberg Engineering (R), Topcon (R), Nidek (R); **C. Yiu**, None; **S. Liu**, None; **M.K. Or**, None; **C. Ye**, None; **D.S.-C. Lam**, None; **C.K.-S. Leung**, Carl Zeiss Meditec (C, R), Optovue (R)

Corresponding author: Christopher Kai-Shun Leung, Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong, People's Republic of China; tlims00@hotmail.com.

Investigative Ophthalmology & Visual Science, December 2011, Vol. 52, No. 13

strategy to be used for progression evaluation. In fact, it is not uncommon to observe different conclusions derived from different strategies using the same set of longitudinal data (Supplementary Fig. S1, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental). This dilemma cannot be easily reconciled because the specificity and sensitivity of most progression analyses are largely unknown given the lack of a reference standard.

In this study, we simulated and modeled different patterns (linear, episodic, exponential, and parabolic loss) and different rates (2 μm/year, 4 μm/year reduction of average RNFL thickness) of RNFL progression using a computer model that takes the inherent test-retest variability into consideration. We compared the performances of event and trend analyses and validated the results of the simulation with prospective, longitudinal data collected from 107 patients followed up for at least 30 months, with the objective of investigating the impact of strategy selection on progression analysis.

Subjects and Methods

Various temporal patterns (linear, episodic, exponential, and parabolic) and rates (−2 μm/year, −4 μm/year) of average RNFL thickness reduction were modeled with reference to different values of intervisit, test-retest variability (spatial patterns of progression were not modeled in this study). These values were obtained from a group of 46 subjects (19 healthy subjects, 27 glaucoma patients) who had weekly RNFL evaluations with a spectral-domain OCT for 8 consecutive weeks. Simulation results were then validated with prospective, longitudinal RNFL measurements obtained from a total of 175 eyes of 107 glaucoma and glaucoma suspect patients who were followed up every 4 months for at least 30 months (range, 30–43 months; median, 38 months).

Subjects

All subjects were observed from June 2007 to March 2011 at the University Eye Center, the Chinese University of Hong Kong. They were enrolled for the Evaluation of RNFL Progression in Glaucoma study, which was designed to investigate the roles of RNFL imaging for monitoring glaucoma. All subjects underwent full ophthalmic examination, including measurement of visual acuity, refraction and intraocular pressure, gonioscopy, and fundus examination. Eyes were included if visual acuity was at least 20/40. Eyes were excluded if they had evidence of macular disease, refractive or retinal surgery, neurologic disease, or diabetes. Glaucoma patients were identified based on the presence of visual field defects with corresponding optic disc and RNFL changes (narrowing of neuroretinal rim or thinning of the RNFL) in at least one eye. Glaucoma suspects were patients without evidence of visual field defects on standard automated perimetry but with glaucomatous optic disc and/or RNFL changes and/or intraocular pressure >22 mm Hg for at least three visits. Healthy subjects had a normal visual field, a normal optic disc and RNFL examination, and no history of intraocular pressure >22 mm Hg. Both eyes were imaged (Cirrus HD-OCT; Carl Zeiss Meditec) every 4 months for at least 30 months. The study was conducted in accordance with the ethical standards stated in the 1964 Declaration of Helsinki and was approved by the local research ethics committee, and informed consent was obtained.

Visual Field Examination

Visual field testing was performed using static automated white-on-white threshold perimetry (SITA Standard 24-2, Humphrey Field Analyzer II; Carl Zeiss Meditec). A visual field was defined as reliable when fixation losses, false-positive, and false-negative errors were <20%. Average visual field sensitivity was expressed in VFI and MD, as calculated by the perimetry software. A visual field defect was defined as the presence of three or more significant ($P < 0.05$) non-edge contiguous points, with at least one at the $P < 0.01$ level on the same side of the horizontal meridian in the pattern deviation plot, and was confirmed with at least two consecutive examinations. Although this definition may exclude glaucoma patients with generalized loss of visual sensitivity, there are far more patients with generalized loss of visual sensitivity as a consequence of cataract. To ensure all patients had glaucomatous damage, we only included patients with visual field defects evident in the pattern deviation plot.

Spectral-Domain OCT RNFL Imaging

Spectral-domain OCT imaging was performed (Cirrus HD-OCT, software version 5.0; Carl Zeiss Meditec). The details of the principles of spectral-domain OCT have been described.[25] Briefly, a superluminescent diode laser with a center wavelength of 840 nm served as a broadband light source for the generation of back-reflections from different intraretinal depths represented by different wavelengths. The acquisition rate of this HD-OCT (Cirrus HD-OCT, software version 5.0; Carl Zeiss Meditec) is 27,000 A-scans per second. The transverse and axial resolutions were 15 μm and 5 μm, respectively. An "optic disc cube" scan protocol was used to measure the RNFL thickness in a 6 × 6 mm^2 area consisting of 200 × 200 axial scans (pixels) at the optic disc region. The RNFL thickness at each pixel was measured, and an RNFL thickness map was generated. A built-in algorithm located the center of the optic disc even if it was not well centered in the scan image.[26] The disc center was identified by the location of a dark spot, whose shape and size were consistent with a range of optic discs, near the center of the scan. A calculation circle of 3.46-mm diameter consisting of 256 A-scans was then automatically positioned around the optic disc. The average RNFL thickness derived from the circle scan was analyzed.

Images were captured by operators with at least 1 year's experience using the Cirrus HD-OCT (Carl Zeiss Meditec). The pupils were not routinely dilated during RNFL imaging. However, dilation with tropicamide 0.5% and phenylephrine 0.5% each was performed when the pupil size was too small for images of the required quality to be obtained. Images with poor centration, motion artifact, poor focus, or missing data were detected by the operator at the time of imaging, with rescanning performed in the same visit. Each OCT scan included in the study had a signal strength of at least 6. Saccadic eye movement was detected with the line-scanning ophthalmoscope overlaid with OCT en face during OCT imaging. Images with motion artifact were rescanned in the same visit. Four RNFL scans from two patients were excluded from the study because of low signal strength. One eye was excluded because of persistent vitreous opacities in the image series.

Measurement of Intervisit Test-Retest Variability

The test-retest variability of average RNFL thickness was calculated for 46 eyes of 46 subjects (19 glaucoma patients and 27 healthy subjects) who had weekly RNFL measurements for 8 consecutive weeks. The mean within-subject SD of average RNFL thickness was 1.77 μm (95% confidence interval, 1.59–1.93 μm), and the range was between 0.74 μm and 3.50 μm (Table 1). The minimum (0.74 μm), median (1.71 μm), and maximum (3.50 μm) within-subject SD, representing eyes with small, average, and large RNFL measurement variability, were used to derive random intervisit, test-retest variabilities in the simulation.

Simulation of Retinal Nerve Fiber Layer Progression

For each simulated dataset, 16 average RNFL thickness measurements, representing RNFL measurements collected every 4 months over 60 months, were randomly generated by a computer from 16 independent normal distributions with a specific value of within-subject SD. A graphical illustration of the simulation process is shown in Supplementary Figure S2, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental. RNFL progression was modeled in four different patterns (linear, episodic, exponential, and parabolic) (Fig. 1) at an average rate of change of −2 μm/year or −4 μm/year over 60 months. The rate estimates were selected with reference to a previous

Table 1. Demographics, Visual Field, and RNFL Measurements of 46 Subjects Who Underwent Weekly RNFL Evaluations for 8 Consecutive Weeks

	Glaucoma Group	Normal Group	All
Sample size	19	27	46
Spherical error, D	−2.7 ± 3.7	−1.3 ± 2.4	−1.9 ± 3.0
Age, y	48.0 ± 13.8	42.9 ± 13.4	45.0 ± 13.6
VFI, %	81.47 ± 24.19	99.48 ± 0.85	92.04 ± 17.75
MD, dB	−7.55 ± 8.08	−0.87 ± 0.84	−3.63 ± 6.13
Average RNFL thickness, μm	68.04 ± 13.59	99.88 ± 9.76	86.73 ± 19.50
Signal strength	8.3 ± 1.0	8.5 ± 0.8	8.4 ± 0.9
Test-retest variability SD, μm			
Median	1.39	1.91	1.71
Minimum	0.83	0.74	0.74
Maximum	2.43	3.50	3.50
Group	1.80	1.87	1.77

D, diopter; VFI, visual field index; MD, mean deviation; RNFL, retinal nerve fiber layer; SD, standard deviation.

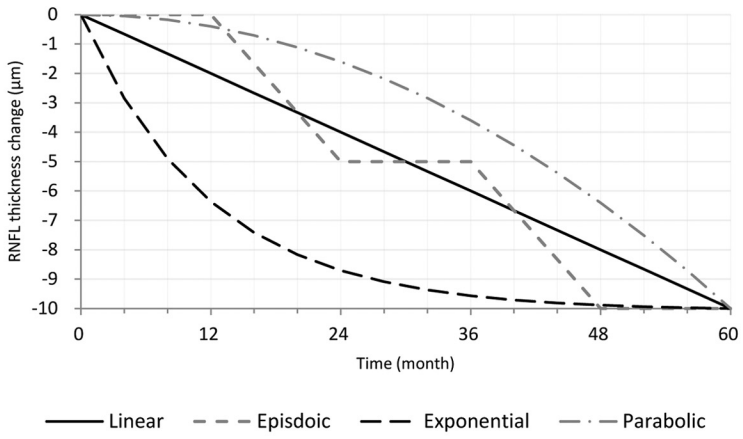
prospective study showing that the rate of average RNFL thickness progression ranged between −1.52 μm/year and −5.03 μm/year.[27] It is likely that −2 μm/year and −4 μm/year would capture the average rate of slow and fast progressors, respectively. Five thousand simulated datasets (each with 16 RNFL measurements) generated with reference to a specific intervisit, test-retest variability, a particular progression pattern, and a specific rate of change were analyzed by five different strategies for detection of progression (The simulation sample size was selected to minimize the confidence intervals of the specificity estimates of the analysis strategies. With a sample size of 5000, all the SDs of the specificity estimates were <1%). These included EA with individual reproducibility coefficient (RC), EA with group RC, EA with individual RC confirmed with a consecutive test, EA with group RC confirmed with a consecutive test, and TA.

Event Analysis with Individual Reproducibility Coefficient

Progression was defined when the difference in average RNFL thickness between the first and the last measurements was greater than the individual's RC (RC was defined as $2 \times \sqrt{2} \times$ within-subject SD[28]).

Event Analysis with Group Reproducibility Coefficient

Progression was defined when the difference in average RNFL thickness between the first and the last measurement was greater than the RC derived from a group of reference individuals ($2 \times \sqrt{2} \times$ mean within-subject SD [1.77 μm]).



Model	RNFL thickness change function
Linear:	$\mu_m - \mu_0 = -2 \times \frac{m}{12}$
Episodic:	$\mu_m - \mu_0 = \begin{cases} 0 & if\ m = 0 \\ (\mu_{m-4} - \mu_0) & if\ m = 4, 8, 12, 28, 32, 36, 52, 56, 60 \\ (\mu_{m-4} - \mu_0) - \frac{10}{6} & if\ m = 16, 20, 24, 40, 44, 48 \end{cases}$
Exponential:	$\mu_m - \mu_0 = -10 \times \frac{e^{-m/12} - 1}{e^{-60/12} - 1}$
Parabolic:	$\mu_m - \mu_0 = -10 \times \left(\frac{m/12}{60/12}\right)^2$

Figure 1. Progressive average retinal nerve fiber layer (RNFL) thickness reduction modeled in linear, episodic, exponential, and parabolic patterns at an average rate of −2.0 μm/year over 60 months.

Event Analysis with Individual Reproducibility Coefficient Confirmed with a Consecutive Test

Progression was defined when the differences in average RNFL thickness between the baseline and the last two measurements were both greater than the individual's RC.

Event Analysis with Group Reproducibility Coefficient Confirmed with a Consecutive Test

Progression was defined when the differences in average RNFL thickness between the baseline and the last two visits were both greater than the group's RC.

Trend Analysis

Progression was defined when a significant negative trend ($P < 0.05$) was detected between average RNFL thickness and time.

Calculation of Sensitivity, Specificity, and Accuracy

Sensitivity, specificity, and accuracy (sensitivity × specificity) (while the area under the receiver operating characteristic curve is an alternative index to combine sensitivity and specificity when the threshold of cutoff [progression vs. no progression] is uncertain, "accuracy" is a better option when the cutoff is well defined) were computed at 12 months and then every 4 months until 60 months of the simulation. Sensitivity was calculated as the proportion of datasets (out of 5000) with an imposed progression pattern detected with progression. Specificity was calculated as the proportion of datasets without an imposed progression pattern detected with no progression. For EA, the last or the latest two consecutive measurements available at a specific time point were compared with the baseline to determine progression. For TA, all available data collected at a specific time point were analyzed. The mathematical formulation of sensitivity and specificity of each analysis strategy are presented in the Appendix.

Results

Sensitivity of Progression Detection

Supplementary Figure S3 (http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental) depicts the sensitivities for the detection of progression with average RNFL thickness reduction modeled at a rate of −2 μm/year in four progression patterns (1, linear; 2, episodic; 3, exponential; 4, parabolic) over 60 months analyzed by five strategies (1, EA with individual RC; 2, EA with group RC; 3, EA with individual RC confirmed with a consecutive test; 4, EA with group RC confirmed with a consecutive test; 5, TA) for eyes with small (SD = 0.74 μm), average (SD = 1.71 μm), and large (SD = 3.50 μm) intervisit, test-retest variabilities.

The sensitivity of progression detection depended on the duration of follow-up, pattern of progression, rate of RNFL loss, individual's test-retest variability, and analysis strategy. With increasing RNFL loss over time, the sensitivity of progression detection increased with the duration of follow-up independent of other factors. Exponential loss of the RNFL exhibited a more dramatic reduction of RNFL thickness in a shorter duration. It attained sensitivity ≥80% for the detection of progression earlier than the other progression patterns. By contrast, parabolic RNFL loss required a longer duration to reach the same sensitivity. The sensitivity profiles of linear and episodic progressions were between those of exponential and parabolic progressions. For the same progression pattern modeled at the same progression rate analyzed by the same strategy, it took longer to achieve sensitivity ≥80% for subjects with large test-retest variability than for those with small test-retest variability.

Although the sensitivities between EA with individual RC and EA with group RC were almost identical for patients with an average test-retest variability (Supplementary Figs. S3E–H, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental), EA with individual RC had a higher sensitivity than EA with group RC for eyes with a small test-retest variability (Supplementary Figs. S3A–D, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental) and vice versa for eyes with a large test-retest variability (Supplementary Figs. S3I–L, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental). EA confirmed with a consecutive test always attained sensitivity ≥80% later than those without, independent of the progression pattern and individual test-retest variability. EA with individual SD, with or without confirmation, had the lowest sensitivity to detect progression for eyes with large test-retest variability (Supplementary Figs. S3I–L, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental).

TA attained sensitivity ≥80% earlier than all forms of EA, although the differences in sensitivity were minimal compared with EA with individual RC for eyes with small test-retest variability in the episodic and exponential progression patterns (Supplementary Figs. S3B, S3C, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental). In general, TA required a shorter duration than EA to reach a high sensitivity for progression detection. Figure 2 shows the minimum duration required to detect RNFL progression at a sensitivity of 80% at various linear rates of change. As an example, for a linear reduction of average RNFL thickness at a rate of

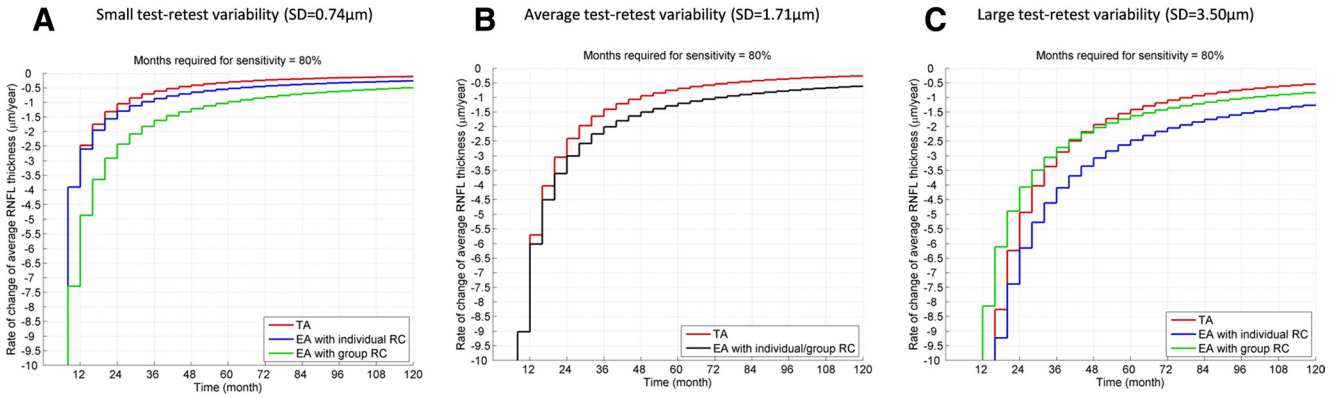


Figure 2. Relationship between the rate of linear change of average RNFL thickness and the minimum duration required for TA and EA to detect progression at a sensitivity of 80% in eyes with small (**A**), average (**B**), and large (**C**) test-retest variabilities. The performance profile of EA with individual RC was identical with that of EA with group RC for eyes with average test-retest variability.

−1.5 μm/year in eyes with large test-retest variability, it took 60 months for TA but 100 months for EA with individual RC and 68 months for EA with group RC to attain a sensitivity of 80% (Fig. 2C). For eyes with small test-retest variability, the minimum duration to attain sensitivity of 80% for the same rate of RNFL reduction was 20, 24, and 40 months, respectively (Fig. 2A). The only scenario in which EA outperformed TA was when the test-retest variability was large and the rate of progression was <−2.5 μm/year and the group RC was used to define progression (Fig. 2C).

The results of simulation modeled at a rate of average RNFL thickness loss at −4 μm/year are shown in Supplementary Figure S4, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental. The pattern of the sensitivity profiles for each analysis strategy was similar to that modeled at a rate of −2 μm/year, though the duration required to attain sensitivity ≥80% was shorter. For eyes with large test-retest variability, EA with group RC attained sensitivity ≥80% earlier than other strategies.

Specificity of Progression Detection

TA and EA with individual RC had a specificity of approximately 95%, independent of test-retest variability (Supplementary Fig. S5, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental). Having a confirmation with a consecutive test increased the specificity of EA with individual RC to 99%. For EA with group RC, eyes with small test-retest variability had a specificity of 100%, whereas eyes with large and average test-retest variability had specificity of 80% and 95%, respectively. Having confirmation with a consecutive test increased the respective specificity to 91% and 99%.

Accuracy of Progression Detection

The accuracies (sensitivity × specificity) of detection of average RNFL thickness progression at a rate of −2 μm/year and −4 μm/year are illustrated in Supplementary Figures S6 and S7, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental, respectively. Independent of the test-retest variability, the pattern and the rate of progression, TA generally reached accuracy ≥80% earlier than EA. Although the accuracy of TA plateaued at 95%, EA with group RC and EA with individual RC confirmed with a consecutive test had accuracy close to 100% for eyes with a small test-retest variability. This is because the specificity for TA was fixed at 95%, whereas it was 100% for the EA.

Validation with Prospective Longitudinal Data

Although no reference standard is available to determine the true sensitivity of TA and EA, it is feasible to compare their relative performance by measuring the proportion of eyes with detectable progression over time. A total of 1680 longitudinal RNFL measurements obtained with a spectral-domain OCT were collected from 175 eyes of 81 glaucoma and 26 glaucoma suspect patients who were prospectively followed up every 4 months for at least 30 months (range, 30–43 months; median, 38 months). Each eye had an average of 10 serial measurements (range, 6–11) for analysis. Demographics, RNFL, and visual field measurements are shown in Table 2.

Each subject's test-retest variability was estimated from the residuals of the least square fit on all the available data points for that subject; 70.3% of eyes (n = 123) had test-retest variability without a significant difference from 1.77 μm (the mean test-retest variability of the reference group [Table 1]) at the 10% level of significance. Figure 3A shows the proportion of these eyes detected with progression by TA and EA computed from 12 months to 36 months. TA and EA with group or individual RC performed similarly in the initial 30 months. At 36 months, TA detected 35%, whereas EA only detected 12% − 28% of eyes with progression. For eyes with a small test-retest variability (arbitrarily defined as the first 50 eyes with the lowest test-retest variability after sorting all eyes in an ascending order), EA with individual RC detected more progressing eyes than EA with group RC (Fig. 3B) and vice versa for eyes with a large test-retest variability (arbitrarily defined as the last 50 eyes with the largest test-retest variability fit after sorting all eyes in an ascending order) (Fig. 3C). For eyes with small test-retest variability, TA detected 42% whereas EA detected 11% − 40% of eyes with progression at 36 months. For eyes with large test-retest variability, EA with group RC detected more progressing eyes (23% at 36 months) than did other strategies (3% − 17% at 36 months) (Fig. 3C). EA confirmed with a consecutive test detected fewer progressing eyes than did EA without confirmation. These patterns closely resembled the simulation results shown in Supplementary Figure S3, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental.

Table 2. Demographics, RNFL, and Visual Field Measurements of 175 Eyes of 107 Glaucoma and Glaucoma Suspect Patients Who Were Monitored Every 4 Months for at Least 30 Months

	Mean ± SD
Spherical error, D	−2.81 ± 4.15
Age, y	51.3 ± 15.1
Signal strength	8.4 ± 1.1
Baseline average RNFL thickness, μm	74.07 ± 14.49
Baseline VFI, %	79.54 ± 24.19
Baseline visual field MD, dB	−7.85 ± 8.22
Average RNFL thickness at final visit, μm	71.99 ± 14.43
VFI at final visit, %	73.01 ± 25.64
Visual field MD at final visit, dB	−10.63 ± 8.27

Duration of follow-up: range, 30–43 months; median, 38 months. D, diopter; y, year; RNFL, retinal nerve fiber layer; VFI, visual field index; MD, mean deviation.

Discussion

With reference to the computer simulation models, we showed that TA generally outperformed EA in attaining a high sensitivity (≥80%) to detect RNFL progression earlier at a comparable level of specificity (95% vs. 80%–100%, respectively). The simulation results were validated with 1680 longitudinal RNFL measurements obtained from 107 glaucoma and glaucoma suspect patients followed up for a median period of 38 months. At 36 months of follow-up, TA detected a greater proportion of eyes with progression (35%) than did EA (12%–28%) in most eyes with average test-retest variability (Fig. 3A). TA may be a preferable strategy for following and detecting disease progression in glaucoma.

EA with progression, defined by a change greater than the reproducibility coefficient calculated from a reference database (i.e., EA with group RC), has been the prevailing approach to analyze glaucoma progression in clinical trials and in clinical practice. However, EA with group RC may fail to detect a small progressive change for patients with small test-retest variability and falsely signify progression for patients with large test-retest variability. Although EA with individual RC offers an individualized approach to detect progression, its performance relative to EA with group RC and TA has not been investigated. Computer simulation revealed that EA with individual RC indeed had a higher sensitivity than EA with group RC to detect change in eyes with small test-retest variability. However, in eyes with large test-retest variability, EA with group RC had

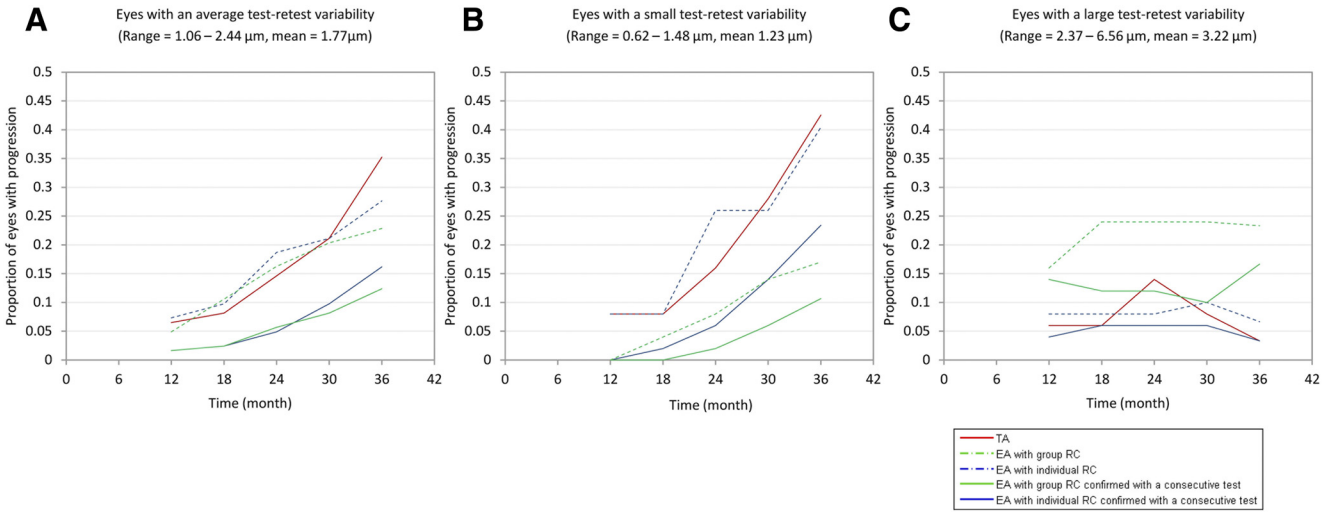


FIGURE 3. Proportion of eyes detected with progression by TA and EA in 175 eyes of 81 glaucoma and 26 glaucoma suspect patients, followed 4 monthly for a median period of 38 months with RNFL thickness measured with spectral-domain OCT. Individual's test-retest variability was estimated from the residuals of the least square fit on all the available data points for that subject. 70.3% of eyes ($n = 123$) had test-retest variability without a significant difference from 1.77 μm (the mean test-retest variability of the reference group) at 10% the level of significance (**A**). Eyes with small test-retest variability were arbitrarily defined as the first 50 eyes with the lowest residuals of the least square fit after sorting all eyes in an ascending order ($n = 50$) (**B**). Eyes with large test-retest variability were arbitrarily defined as the last 50 eyes with the largest residuals of the least square fit after sorting all eyes in ascending order ($n = 50$) (**C**).

higher sensitivity than EA with individual RC, albeit at a lower specificity (80% and 95%, respectively). These findings were confirmed with analysis on longitudinal clinical data (Figs. 3B, 3C). The difference in performance between EA with individual RC and EA with group RC can be explained in mathematical terms, with sensitivity represented by

$$\Phi\left(-z_\alpha - \frac{\mu_n - \mu_1}{\sqrt{2\sigma^2}}\right)$$

and

$$\Phi\left(-z_\alpha\frac{\sigma_g}{\sigma} - \frac{\mu_n - \mu_1}{\sqrt{2\sigma^2}}\right)$$

respectively (see Appendix for annotation) (Supplementary Fig. S8, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental). If the individual's test-retest variability is smaller than the group's test-retest variability, σ_g/σ would be >1, thereby shifting the classification cutoff of EA with group RC more negative and resulting in a lower sensitivity (and a higher specificity) compared with EA with individual RC. By contrast, if the individual's test-retest variability is greater than the group's test-retest variability, σ_g/σ would be <1, thereby shifting the classification cutoff of EA with group RC less negative and resulting in a higher sensitivity (and a lower specificity) compared with EA with individual RC. For most patients with average test-retest variability, the performance of EA with group RC would not be very much different from EA with individual RC (Fig. 3A, Supplementary Figs. S3E–H, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental). Collectively, EA with individual RC is more informative than EA with group RC only when the individual test-retest variability is small. As expected, having a confirmation with a consecutive test increased the specificity but reduced the sensitivity for progression detection.

In the simulation models, TA often attained a high sensitivity earlier than EA with individual RC or group RC. However, EA with group RC was more sensitive than TA in the early follow-up period for patients with large test-retest variability. The superior performance was more remarkable when the rate of progression was fast (Supplementary Figs. S4I–L, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental). The relative sensitivity between TA and EA with group RC can be computed by comparing

$$\Phi\left(-z_\alpha - \frac{\sum(2i - (n+1))\mu_i}{\sqrt{\left(\frac{n(n^2-1)}{3}\right)\sigma^2}}\right)$$

and

$$\Phi\left(-z_\alpha\frac{\sigma_g}{\sigma} - \frac{\mu_n - \mu_1}{\sqrt{2\sigma^2}}\right)$$

(see Appendix). By fixing the specificity at Z_α, TA would be more sensitive when

$$\frac{\sum(2i - (n+1))\mu_i}{\sqrt{\left(\frac{n(n^2-1)}{3}\right)}} < z_\alpha(\sigma_g - \sigma) + \frac{\mu_n - \mu_1}{\sqrt{2}}.$$

This indicates that the selection between the 2 strategies depends on the relative difference between the group and individual test-retest variability (σ_g and σ) and the number of follow-up visits (n) (or follow-up duration). TA would be more sensitive than EA with group RC when the patient's test-retest variability is small and the follow-up duration is long. Although EA with group RC had a higher sensitivity than TA in the early follow-up period for subjects with large test-test variability, EA with group RC had lower specificity (80%). Accuracy (sensitivity × specificity) combines sensitivity and specificity and provides a more comprehensive analysis to evaluate the performance of different strategies. In fact, TA attained accuracy ≥80% earlier than all forms of EA independently of the test-retest variability, the pattern, and the rate of progression (Sup-

plementary Figs. S6, S7, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental).

An ideal strategy for the detection of progression should demonstrate high sensitivity at high specificity. It is not surprising to observe from the simulation that the specificity of TA and EA with individual RC was 95% because the alpha selected to define a significant change was fixed at 5%. Given that the specificity of EA with group RC was defined with alpha fixed at 5% with reference to the group RC, the specificity would be 95% for patients with average test-retest variability when the individual RC is similar to the group RC. Patients with small test-retest variability would have higher specificity (99%), whereas those with large test-retest variability would have a lower specificity (80%) (Supplementary Fig. S5, http://www.iovs.org/lookup/suppl/doi:10.1167/iovs.11-8052/-/DCSupplemental).

TA is superior not only in attaining high accuracy for detecting progression earlier than EA but also in providing a rate estimate that is useful to guide treatment and evaluate disease prognosis. In contrast to EA, however, multiple measurements are always required to compute a reliable slope estimate in TA. An important question is: how often should an imaging or a visual field test be scheduled? In other words, how many observations are needed to reliably derive the slope estimate? This question has been previously discussed with visual field testing, though little is known for structural assessment.[29,30] By specifying the level of sensitivity and the rate of change, it is possible to work out the impact of the number of observations per year on the minimum duration required to detect progression (Fig. 4) (see Appendix). Assuming a rate of change of average RNFL thickness of -2 μm per year to be clinically significant, it takes 2.7 years to detect this change at 70% sensitivity and 2.9 years at 80% sensitivity for a patient with average test-retest variability (SD = 1.71 μm) if three measurements are obtained per year. Increasing the frequency to six measurements per year slightly shortens the duration to 2.1 and 2.3 years, respectively, whereas decreasing the frequency to one observation per year significantly lengthens the duration to 4.5 and 4.8 years, respectively. Patients with larger test-retest variability require longer duration to detect the same rate of change, particularly when the number of measurements obtained per year is small. Collectively, the number of measurements required depends on the rate of change to be considered as significant, the desired level of sensitivity, the patient's test-retest variability, and the acceptable duration for its detection, which may vary individually according to the severity of disease and the life expectancy. In general, the optimal number of measurements required per year is approximately four (i.e., the turning point of the curves) (Fig. 4). The benefit of obtaining additional measurements in shortening the duration required to detect the same rate of change at the same level of sensitivity is small.

The next question to address is whether scheduling follow-up at regular intervals is important to measurements of the slope estimate. In ordinary least square (OLS), we reject the null hypothesis,

$$H_0: \beta = 0,$$

if

$$\frac{\hat{\beta}}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta})}} < -t_{\alpha,n-2}$$

where

$$\hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

is the OLS estimate of β and

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \frac{\frac{1}{n-2}\sum(\varepsilon_i - \bar{\varepsilon})^2}{\sum(x_i - \bar{x})^2}$$

is the estimated variance of $\hat{\beta}$ (see Appendix). Because $\hat{\beta}$ is an unbiased estimator (the mean of $\hat{\beta}$ is the same as the true parameter β, i.e., $E(\hat{\beta}) = \beta$), changing x_i (the time interval between tests) has no effect on the estimation of β. Yet, maximizing $\sum(x_i - \bar{x})^2$ (the denominator of $\widehat{\mathrm{Var}}(\hat{\beta})$) can minimize $\widehat{\mathrm{Var}}(\hat{\beta})$, thus improving the sensitivity. In other words, obtaining measurements only at the beginning and at the end of a defined follow-up period provides the highest sensitivity to detect change. This wait-and-see approach concurs with the visual field simulation model proposed by Crabb and Garway-Heath (*IOVS* 2009;50:ARVO E-Abstract 1669). It is important to note the assumptions of the OLS are that measurements at each time point have the same variance and are time independent. These assumptions, however, may not be legitimate in a real-time clinical setting, especially when the follow-up duration is long. Progression of cataract, surgical interventions including trabeculectomy and cataract extraction, and instrument instability (e.g., lack of regular calibration) could substantially affect the quality of data collection and, hence, the reliability of measurement. Maximizing $\sum(x_i - \bar{x})^2$ by collecting data only at the extreme ends may result in a more biased estimate than spacing the measurements at regular time intervals. Another shortcoming is that patients experiencing rapid progression could be missed if investigations are separated by a relatively long period. Further studies are needed to examine the effect of time-dependency factors on progression analysis.

Although the actual rate of age-related RNFL loss is largely unknown, cross-sectional analysis suggests that the average RNFL

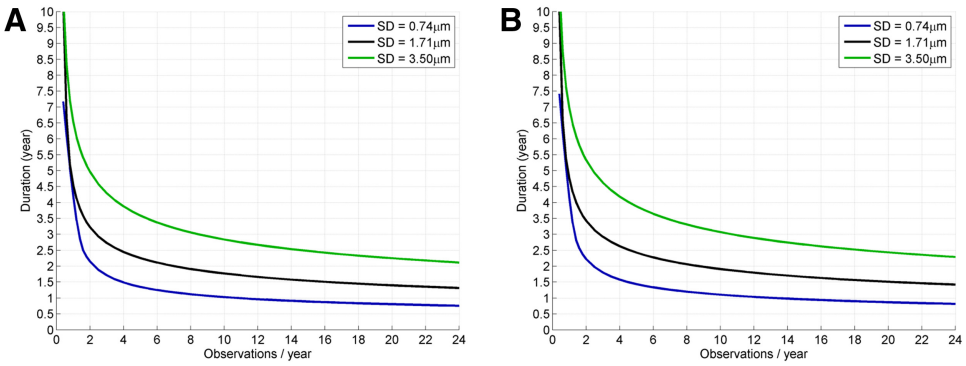


Figure 4. Relationship between the number of observations required per year and the minimum duration needed to detect average RNFL thickness progression at a rate of -2 μm/year at a sensitivity of 70% (**A**) and 80% (**B**) for eyes with small (SD = 0.74 μm), average (SD = 1.71 μm), and large (SD = 3.50 μm) test-retest variabilities.

thickness reduces at a rate of approximately −0.2 μm/year.[31,32] If the follow-up duration is long enough, significant reduction in RNFL thickness might be detected, particularly for patients with small test-retest variability. Thus, the specificity of trend and event analyses would be reduced over time. Examining the rate estimate derived from TA is pertinent to differentiate glaucoma-related from age-related RNFL loss because the rate of glaucoma-related RNFL loss would be expected to exceed the rate of age-related changes. Studies investigating the rate of age-related loss of RNFL are eminently needed.

RNFL progression was modeled in this study because it has been well documented that RNFL measured by spectral-domain OCT has relatively low test-retest variability, thereby facilitating the detection of progression.[5,33] It is feasible to apply the simulation to other imaging or visual field testing modalities provided that the range of test-retest variabilities and the expected rate of change of a particular parameter of interest are available. For tests with inherently large test-retest variability and relatively rapid rate of change, EA with group RC would be preferable to TA for progression detection. Customizing the analysis strategy may be necessary to track disease progression in different types of structural and functional tests. Of note, the present study is limited in evaluating only global change in RNFL thickness derived from a circle scan with the objective of comparing trend and event analyses. Although it is plausible that local change of the RNFL may follow a similar pattern, further investigation is needed to fully address the differences in performance between TA and EA for the detection of local change.

There are different approaches to calculate sensitivity and specificity in defining progression. Although sensitivity may be given by the probability of any of the first n tests detected with a statistically significant progression (i.e., $1 - (1\text{-}P_1)(1\text{-}P_2)(1\text{-}P_3)..(1\text{-}P_n)$, where P_n is the probability of detecting a statistically significant progression at visit n), a more common approach in clinical practice is to compare the baseline with the latest available measurements without taking other tests obtained in between into consideration. For example, in the visual field Guided Progression Analysis (Carl Zeiss Meditec, Dublin, CA), progression is defined when the same three or more locations in the pattern deviation plot had change exceeding the limits of normal variability in the two latest consecutive tests (i.e., the same criteria as in the EMGT). In the Guided Progression Analysis of the Cirrus HD-OCT (Carl Zeiss Meditec) and the GDx ECC/VCC (Carl Zeiss Meditec), retinal nerve fiber layer thickness progression is defined as "possible loss" when the change in the last follow-up measurement exceeds the limits of normal variability and as "likely loss" when the change is evident in the two latest consecutive tests. Taking any of the tests obtained in the follow-up measurements to define progression will overestimate sensitivity and underestimate specificity.

In conclusion, the sensitivity to detect change in glaucoma progression is dictated by the pattern of progression, the rate of change of disease, the test-retest variability, and the analysis strategy. Although it is not possible to modify the pattern or the rate of disease progression, selecting an appropriate strategy to analyze progression is germane to maximizing the probability to detect change. For most patients, TA attains high sensitivity earlier than EA at a comparable level of specificity. For eyes with large test-retest variability, EA with group RC could be more sensitive than TA but at the expense of reduced specificity.

References

1. Weinreb RN, Khaw PT. Primary open-angle glaucoma. *Lancet.* 2004;363:1711–1720.
2. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science.* 1991;254:1178–1181.
3. Zhou Q, Weinreb RN. Individualized compensation of anterior segment birefringence during scanning laser polarimetry. *Invest Ophthalmol Vis Sci.* 2002;43:2221–2228.
4. Webb RH, Hughes GW, Delori FC. Confocal scanning laser ophthalmoscope. *Appl Opt.* 1987;26:1492–1499.
5. 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.
6. Budenz DL, Fredette MJ, Feuer WJ, et al. Reproducibility of peripapillary retinal nerve fiber thickness measurements with stratus OCT in glaucomatous eyes. *Ophthalmology.* 2008;115:661–666.
7. Paunescu LA, Schuman JS, Price LL, et al. Reproducibility of nerve fiber thickness, macular thickness, and optic nerve head measurements using StratusOCT. *Invest Ophthalmol Vis Sci.* 2004;45:1716–1724.
8. Weinreb RN, Lusky M, Bartsch DU, et al. Effect of repetitive imaging on topographic measurements of the optic nerve head. *Arch Ophthalmol.* 1993;111:636–638.
9. Leung CK, Cheung CY, Lin D, et al. Longitudinal variability of optic disc and retinal nerve fiber layer measurements. *Invest Ophthalmol Vis Sci.* 2008;49:4886–4892.
10. Hoh ST, Ishikawa H, Greenfield DS, et al. Peripapillary nerve fiber layer thickness measurement reproducibility using scanning laser polarimetry. *J Glaucoma.* 1998;7:12–15.
11. Chauhan BC, LeBlanc RP, McCormick TA, et al. Test-retest variability of topographic measurements with confocal scanning laser tomography in patients with glaucoma and control subjects. *Am J Ophthalmol.* 1994;118:9–15.
12. Mikelberg FS, Wijsman K, Schulzer M. Reproducibility of topographic parameters obtained with the heidelberg retina tomograph. *J Glaucoma.* 1993;2:101–103.
13. Kruse FE, Burk RO, Völcker HE, et al. Reproducibility of topographic measurements of the optic nerve head with laser tomographic scanning. *Ophthalmology.* 1989;96:1320–1324.
14. Leung CK, Cheung CY, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in glaucoma: a study on optical coherence tomography guided progression analysis. *Invest Ophthalmol Vis Sci.* 2010;51:217–222.
15. Leung CK, Liu S, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in glaucoma: a prospective analysis with neuroretinal rim and visual field progression. *Ophthalmology.* 2011;118:1551–1557.
16. Wollstein G, Schuman JS, Price LL, et al. Optical coherence tomography longitudinal evaluation of retinal nerve fiber layer thickness in glaucoma. *Arch Ophthalmol.* 2005;123:464–470.
17. Medeiros FA, Alencar LM, Zangwill LM, et al. Detection of progressive retinal nerve fiber layer loss in glaucoma using scanning laser polarimetry with variable corneal compensation. *Invest Ophthalmol Vis Sci.* 2009;50:1675–1681.
18. Artes PH, Chauhan BC. Longitudinal changes in the visual field and optic disc in glaucoma. *Prog Retin Eye Res.* 2005;24:333–354.
19. AGIS Investigators. Advanced Glaucoma Intervention Study, 2: visual field test scoring and reliability. *Ophthalmology.* 1994;101:1445–1455.
20. AGIS Investigators. The Advanced Glaucoma Intervention Study (AGIS), 1: study design and methods and baseline characteristics of study patients. *Control Clin Trials.* 1994;15:299–325.
21. Musch DC, Lichter PR, Guire KE, et al. The Collaborative Initial Glaucoma Treatment Study: study design, methods, and baseline characteristics of enrolled patients. *Ophthalmology.* 1999;106:653–662.
22. Leske MC, Heijl A, Hyman L, et al. Early Manifest Glaucoma Trial: design and baseline data. *Ophthalmology.* 1999;106:2144–2153.
23. Casas-Llera P, Rebolleda G, Muñoz-Negrete FJ, et al. Visual field index rate and event-based glaucoma progression analysis: comparison in a glaucoma population. *Br J Ophthalmol.* 2009;93:1576–1579.
24. Fayers T, Strouthidis NG, Garway-Heath DF. Monitoring glaucomatous progression using a novel Heidelberg Retina Tomograph event analysis. *Ophthalmology.* 2007;114:1973–1980.
25. Nassif N, Cense B, Park BH, et al. In vivo human retinal imaging by ultrahigh-speed spectral-domain optical coherence tomography. *Opt Lett.* 2004;29:480–482.
26. *Cirrus HD-OCT User Manual*. Dublin, CA: Carl Zeiss Meditec Inc.; 2008; Revision A:4–18-9.

27. Leung CK, Chiu V, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in glaucoma: a comparison between spectral-domain and time-domain optical coherence tomography. *Ophthalmology*. 2011;118:1558–1562.
28. Bland JM, Altman DG. Measurement error. *BMJ.* 1996;313:744.
29. Nouri-Mahdavi K, Zarei R, Caprioli J. Influence of visual field testing frequency on detection of glaucoma progression with trend analyses. *Arch Ophthalmol.* In press.
30. Chauhan BC, Garway-Heath DF, Goñi FJ, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol.* 2008;92:569–573.
31. Budenz DL, Anderson DR, Varma R, et al. Determinants of normal retinal nerve fiber layer thickness measured by Stratus OCT. *Ophthalmology.* 2007;114:1046–1052.
32. Parikh RS, Parikh SR, Sekhar GC, et al. Normal age-related decay of retinal nerve fiber layer thickness. *Ophthalmology.* 2007;114:921–926.
33. Shin CJ, Sung KR, Um TW, et al. Comparison of retinal nerve fibre layer thickness measurements calculated by the optic nerve head map (NHM4) and RNFL3.45 modes of spectral-domain optical coherence tomography (RTVue-100). *Br J Ophthalmol.* 2010;94:763–767.

APPENDIX

The sensitivity and specificity of each of the progression detection strategies can be computed and expressed in mathematical terms. A list of symbol annotations is summarized as follows:

Y_i = An independent average RNFL thickness measurement obtained at a particular visit and distributed normally with homogenous variance (i.e., $Y_i \sim N(\mu_i, \sigma^2)$).
μ_i = The average RNFL thickness in the ith visit.
σ = The SD of an individual patient.
σ_g = The SD of a group of patients.
z_α = The critical value satisfying $\Phi(-z_\alpha) = \alpha$ (the corresponding z_α for $\alpha = 0.05$ is 1.645).
$\Phi(\cdot)$ = The cumulative distribution function of standard normal distribution.
$\Phi_{2;\Sigma}(\cdot)$ = The cumulative distribution function of a bivariate normal with mean $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and variance-covariance matrix Σ.

Event Analysis with Individual Reproducibility Coefficient

Under the normal distribution assumption, $D_n = (Y_n - Y_1) \sim N(\mu_n - \mu_1, 2\sigma^2)$ progression was defined when $D_n < -z_\alpha\sqrt{2\sigma^2}$.

Denoting z_α to be the critical value satisfying $\Phi(-z_\alpha) = \alpha$ and $\Phi(\cdot)$ to be the cumulative distribution function of standard normal distribution, the specificity is $1 - \Phi(-z_\alpha) = (1 - \alpha)$, and the sensitivity is

$$P(D_n < -z_\alpha\sqrt{2\sigma^2} | \mu_n < \mu_1) = \Phi\left(-z_\alpha - \frac{\mu_n - \mu_1}{\sqrt{2\sigma^2}}\right).$$

Event Analysis with Group Reproducibility Coefficient

Progression was defined when

$$D_n < -z_\alpha\sqrt{2\sigma_g^2}$$

As

$$-z_\alpha\sqrt{2\sigma_g^2} = -z_\alpha\frac{\sigma_g}{\sigma}\sqrt{2\sigma^2},$$

the specificity is

$$1 - \Phi\left(-z_\alpha\frac{\sigma_g}{\sigma}\right) = \Phi\left(z_\alpha\frac{\sigma_g}{\sigma}\right)$$

and the sensitivity is

$$\Phi\left(-z_\alpha\frac{\sigma_g}{\sigma} - \frac{\mu_n - \mu_1}{\sqrt{2\sigma^2}}\right)$$

Event Analysis with Individual Reproducibility Coefficient Confirmed with a Consecutive Test

Under the independent normal distribution assumption of Y_i,

$$\begin{bmatrix} D_{n-1} \\ D_n \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_{n-1} - \mu_1 \\ \mu_n - \mu_1 \end{bmatrix}, 2\sigma^2\Sigma\right),$$

where

$$\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

with the correlation between the two differences come from the common term, Y_1.

Denoting the cumulative distribution function of a bivariate normal with mean $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and variance-covariance matrix Σ to be $\Phi_{2;\Sigma}(\cdot)$. To control the specificity at $(1 - \alpha)$, progression was defined if $D_{n-1} < -z_\alpha\sqrt{2\sigma^2}$ and $D_n < -z_\alpha\sqrt{2\sigma^2}$.

The specificity is $1 - \Phi_{2;\Sigma}(-z_\alpha, -z_\alpha)$ and the sensitivity is

$$P(D_{n-1} < -z_\alpha\sqrt{2\sigma^2}, D_n < -z_\alpha\sqrt{2\sigma^2} | \mu_n < \mu_1) = \Phi_{2;\Sigma}\left(-z_\alpha - \frac{\mu_n - \mu_1}{\sqrt{2\sigma^2}}, -z_\alpha - \frac{\mu_{n-1} - \mu_1}{\sqrt{2\sigma^2}}\right).$$

Event Analysis with Group Reproducibility Coefficient Confirmed with a Consecutive Test

Similarly, by replacing the individual SD σ with the group SD σ_g, the classification cutoff becomes

$$-z_\alpha\sqrt{2\sigma_g^2} = -z_\alpha\frac{\sigma_g}{\sigma}\sqrt{2\sigma^2}$$

Therefore, the specificity and sensitivity are

$$1 - \Phi_{2;\Sigma}\left(-z_\alpha\frac{\sigma_g}{\sigma}, -z_\alpha\frac{\sigma_g}{\sigma}\right)$$

and

$$\Phi_{z;\Sigma}\left(-z_\alpha\frac{\sigma_g}{\sigma} - \frac{\mu_n - \mu_1}{\sqrt{2\sigma^2}}, -z_\alpha\frac{\sigma_g}{\sigma} - \frac{\mu_{n-1} - \mu_1}{\sqrt{2\sigma^2}}\right),$$

respectively.

Trend Analysis

Simulated serial RNFL measurements were analyzed by ordinary least square regression. Progression was defined if the slope β was <0.

The estimate of β is given by

$$\hat{\beta} = \frac{\widehat{\mathrm{Cov}}(X,Y)}{\widehat{\mathrm{Var}}(X)} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}.$$

In the model, each RNFL measurement was simulated at regular intervals. The summation of x can be expressed by an arithmetic series $1+2+3\ldots.+ n$, where n represents the nth measurement.

$$x_i = i, \sum x_i = \frac{n(n+1)}{2}, \sum x_i^2 = \frac{n(n+1)(2n+1)}{6}$$

and

$$\hat{\beta} = \sum\left(i - \frac{n+1}{2}\right) y_i \Big/ \left(\frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4}\right) = \frac{\sum(2i-(n+1))y_i}{\frac{n(n^2-1)}{6}}.$$

Under the normal distribution assumption,

$$\hat{\beta} \sim N\left(\frac{\sum(2i-(n+1))\mu_i}{\frac{n(n^2-1)}{6}}, \frac{2\sigma^2}{\frac{n(n^2-1)}{6}}\right)$$

To control the specificity at $(1 - \alpha)$, progression was defined when

$$\hat{\beta} < -z_\alpha \sqrt{\frac{2\sigma^2}{\frac{n(n^2-1)}{6}}}$$

and the sensitivity is given by

$$P\left(\hat{\beta} < -z_\alpha \sqrt{\frac{2\sigma^2}{n(n^2-1)/6}} \,\middle|\, \mu_n < \mu_1\right) = \Phi\left(-z_\alpha - \frac{\Sigma(2i-(n+1))\mu_i}{\sqrt{(n(n^2-1)/3)\sigma^2}}\right)$$

The duration (years), δ, required to detect progression at a sensitivity β with θ observations per year with a yearly reduction of average RNFL thickness at γ is given by:

$$\delta(\theta,\gamma,\beta) = \arg\min_d \left\{ d \,\middle|\, \Phi\left(-z_\alpha - \frac{\gamma \sum_{i=0,1,\ldots,\lfloor \theta d \rfloor}(2i - (\lfloor \theta d \rfloor + 1))\frac{i}{\theta}}{\sqrt{\left(\frac{\lfloor \theta d \rfloor(\lfloor \theta d \rfloor^2 - 1)}{3}\right)\sigma^2}}\right) \geq \beta \right\}.$$