Published in final edited form as:

J Glaucoma. 2019 May ; 28(5): 368–374. doi:10.1097/IJG.0000000000001222.

Optical coherence tomography segmentation errors of the retinal nerve fiber layer persist over time

Nisha Nagarkatti-Gude, M.D., Ph.D., Stuart K Gardiner, Ph.D., Brad Fortune, O.D., Ph.D., Shaban Demirel, BSci Optom, Ph.D., and Steven L Mansberger, M.D., M.P.H.

Legacy Devers Eye Institute, Portland, OR

Abstract

Purpose: To identify whether optical coherence tomography (OCT) segmentation errors in retinal nerve fiber layer (RNFL) thickness measurements persist longitudinally.

Methods: This was a cohort study. We used spectral domain OCT (Spectralis, Heidelberg Engineering, Heidelberg, Germany) to measure RNFL thickness in a 6-degree peripapillary circle, and exported the native ‘automated segmentation only’ results. In addition, we exported RNFL thickness results after ‘manual refinement’ to correct errors in the automated segmentation, and used the differences in these measurements as ‘error’ in segmentation. We used Bland-Altman plots and linear regression to determine the magnitude, location, and repeatability of RNFL thickness error in all twelve 30-degree sectors and compared the error at baseline to follow-up time points at 6 months, 2 years, 3 years, and 4 years.

Results: We included 406 eyes from 213 participants. The 95% confidence interval for errors at baseline was -6.5 to $+13.2\mu\text{m}$. The correlation between the baseline error and the errors in the follow-up time periods were high ($r > 0.5$, $p < 0.001$ for all). Automated segmentation had a smaller standard deviation of residuals from the longitudinal trend line when compared to manual refinement ($1.56\mu\text{m}$ vs. $1.80\mu\text{m}$, $p < 0.001$), and a higher ability ($p = 0.009$) to monitor progression using an analysis of a longitudinal signal-to-noise ratio.

Conclusions: Errors in automated segmentation remain relatively stable, and baseline error is highly likely to persist in the same direction and magnitude in subsequent time periods. However, automated segmentation (without manual refinement) is more repeatable and may be more sensitive to glaucomatous progression. Future segmentation algorithms could exploit these findings to improve automated segmentation in the future.

Précis:

There are errors in automated segmentation of the retinal nerve fiber layer in glaucoma suspects or patients with mild glaucoma that appear to persist over time; however, automated segmentation has greater repeatability than manual segmentation.

Keywords

peripapillary; retinal nerve fiber layer thickness; optical coherence tomography; retinal nerve fiber layer; glaucoma; glaucoma progression; optic disc; optic nerve

Introduction:

Optical coherence tomography (OCT) provides objective, quantitative measurements of the peripapillary RNFL thickness, which has become one of the most common measurements used to aid both the diagnosis and monitoring of glaucoma.^{1,2} The utility of this measurement is underscored by recent studies demonstrating that rate of thinning is predictive of subsequent visual field loss.^{2,3} However, OCT measurements of RNFL thickness rely on automated segmentation algorithms^{4,5} to determine RNFL thickness. Yet it is known that automated segmentation algorithms are imperfect and are known to err in delineating the boundaries of the RNFL.⁶⁻¹² We recently published a cross-sectional comparison of RNFL thickness results derived by automated OCT segmentation to results derived by manual refinement of OCT segmentation to determine the source and location of errors present in the automated RNFL segmentations.⁶ We found that automated segmentation resulted in thinner RNFL measurements in general, particularly in the superior and inferior temporal sectors; and it classified more eyes as having glaucomatous damage than when manual refinement was applied to correct segmentation errors.⁶

Studies suggest that multiple factors contribute to errors of RNFL thickness measurements such as the presence of the large retinal blood vessels within the tissue layer, media opacities, miotic pupils, long axial length, advanced glaucoma, scan decentration, epiretinal membranes, vitreoretinal traction, and other less common findings.⁷⁻¹² Between 19.9% and 46.3% of OCT scans contain at least one segmentation artifact.⁷⁻¹¹ The presence of large retinal blood vessels within the RNFL complicates assessment of neural loss in part because the signal attenuation caused by flowing blood creates 'shadows' that make delineation of the posterior RNFL boundary more difficult.^{6,13}

Previous studies have not determined whether errors of RNFL segmentation are static or change over time, perhaps as a result of aging or other factors.¹⁴⁻¹⁷ Ideally, a segmentation algorithm should produce measurements that are both accurate and reproducible. This study utilizes a large, longitudinal database to examine differences in RNFL measurements between automated segmentation and manual refinement in individuals with a diagnosis of glaucoma suspect or open-angle glaucoma. If these differences between automated and manual segmentation persist longitudinally, clinicians and researchers could use this information to identify the most common sources of error and develop improved automated segmentation algorithms over time. This information will also help distinguish between transient errors that impact both detection of glaucoma and assessment of progression, or consistent errors that still impact detection but may have less effect on longitudinal comparisons.

Methods:

Study design and study population

This cohort study includes participants enrolled in the ongoing Portland Progression Project at Legacy Devers Eye Institute (Portland, Oregon). The Legacy Health Institutional Review Board (Portland, OR) reviewed and approved the study. Participants provided written informed consent before undergoing any study-related testing. The study adhered to the tenets of the Declaration of Helsinki.

The study included both eyes of participants with diagnoses of glaucoma suspect or open-angle glaucoma and tested them every 6 months with visual acuity, tonometry, gonioscopy, biomicroscopy of the anterior and posterior segments, perimetry, and OCT measurements of the optic nerve.⁶ Participants at the time of enrollment had: 1) optic nerves suspicious for glaucoma including large cup to disc ratio, cup to disc asymmetry >0.2 between eyes, history of disc hemorrhage or rim notching, or nerve fiber layer thinning or defect on ophthalmoscopy; and/or 2) ocular hypertension defined by intraocular pressure (IOP) ≥ 22 millimeters mercury (mmHg); and/or 3) at least one additional risk factor for glaucoma such as a first-degree family history of primary open angle glaucoma (POAG).⁶ Only participants with a refractive error (spherical equivalent) between -4.00 diopters and $+4.00$ diopters were included in this study. Exclusion criteria included participants with visual acuity worse than 20/40 or with other ocular conditions that could affect visual acuity or visual fields.⁶

Data acquisition

The patients performed perimetry using a standard protocol (Humphrey SITA standard algorithm, Size III stimulus, 24–2 test pattern, Carl Zeiss Meditec, Dublin, CA). Trained operators used a spectral domain OCT to obtain peripapillary RNFL thickness measurements (Spectralis, Heidelberg Engineering, Heidelberg, Germany). Briefly, the operator centered the standard circumpapillary scan on the optic nerve head and focused images to optimize the clarity. Each circle scan included 1536 A-scans at a radius of 6 degrees from the center of the optic nerve head. The final image was the average of 9 collected sweeps obtained with real-time eye-tracking. The operator captured follow-up scans at the same location as the first scan using the instrument's real-time eye tracking algorithm. We exported the native OCT B-scans (.vol file format) including the automated RNFL segmentations in the 'align' mode for subsequent offline analysis.

Next, the operators used the Heidelberg Eye Explorer Software tool (Software version 5.6.1.0, Heidelberg Engineering, Heidelberg, Germany) to achieve an accurate delineation of the anterior RNFL boundary (between the vitreous and the internal limiting membrane) and the posterior boundary of the RNFL¹⁸ (between the RNFL and the ganglion cell layer) when there were obvious errors in the automated results.^{19,20} Then, all B-scans using these manually refined segmentation data were also exported in the 'align' mode for subsequent offline analysis.⁶ We excluded eyes with scan quality score less than 15 or when epiretinal membrane or vitreous traction prevented accurate delineation of RNFL.

Data Analysis

We used R (www.R-project.org, version 3.2.1) to perform all analyses. We used a custom program to convert the 1536 thickness samples from the circumpapillary B-scans into twelve 30-degree sectors of RNFL thickness measurements and converted left eye sectors to the corresponding right eye sectors in a clock hour configuration. We define “error” as the difference between automated and manually refined RNFL thickness measurements (manual RNFL thickness minus automated RNFL thickness), where a positive value indicates that manual refinement resulted in a higher value than automated segmentation.

To determine whether the magnitude errors persisted longitudinally in the same location, we used Bland-Altman plots to compare the errors at baseline with those from follow-up visits (the difference in error at that sector between the two time points, plotted against the average error for those time points). We define the persistence of error longitudinally as no significant change in the slope of error over time. Follow-up visits were 6 months, 2 years, 3 years and 4 years post-baseline, or the nearest test date to that time point (subject to being within ± 180 days of the desired date). This was performed firstly for all 12 sectors, hereafter referred to as the “All Sector Analysis”; and secondly restricted to just the sector that had the largest magnitude error on the baseline scan, referred to as the “Worst Sector Analysis”.

To determine the effect of manual refinement on the repeatability of RNFL thickness measurements, we performed linear regression of both series of the measurements for each sector over time and calculated the standard deviation of the residuals from this trend lines. The standard deviations were compared between automated vs. manually refined segmentations using a paired Wilcoxon test adjusted for clustering (two eyes per subject) using the technique by Rosner and colleagues.²¹ We also compared the longitudinal signal-to-noise ratio in the same manner, which we have previously defined as the rate of change divided by the standard deviation of the residuals for that sector,²² to assess the ability to monitor progression.

Results:

We included 405 eyes from 213 participants at baseline – 7 eyes from the original database were not included due to poor quality scans. We examined data from 405 (100%) of eyes at the 6-month follow-up, 386 (95%) eyes at the 2-year follow up, 374 (92%) eyes at the 3-year follow-up, and 345 (85%) eyes at the 4-year follow up. The Table shows the participant demographics.

A cross-sectional analysis of error using the All Sector Analysis showed that 95% of sectors had an error between -6.5 and $+13.2\mu\text{m}$ at baseline. On the 6-month follow-up date, 95% of sectors had an error between -7.7 and $+13.9\mu\text{m}$. Similar results occurred in subsequent time periods, suggesting that the magnitude and range of error remained similar at each time period.

Figure 1 shows the Bland-Altman plots comparing errors at baseline versus subsequent time periods using the ‘All Sector Analysis’. The 95% range for the difference between the error at baseline versus the error at the 6-month follow-up visit was between -10.5 and $+9.9\mu\text{m}$

(Figure 1, top left) with a correlation of 0.553 ($p<0.001$, from a generalized estimating equation model). The correlation between the error at baseline and the error at the 2-year time period was 0.548 ($p<0.001$, Figure 1, top right). The correlation at the 3-year time period was 0.511 ($p<0.001$, Figure 1, bottom left). The correlations at the 4-year time period was 0.542 ($p<0.001$, Figure 1, bottom right). These results suggest that errors were highly likely to persist in direction and magnitude.

The analysis only found 55 sectors (1.15%) in which the absolute error on the baseline scan was greater than $20\mu\text{m}$. It is notable that in 38 (69%) of these sectors, the error in the 6-month follow-up scan remained greater than $10\mu\text{m}$, and in the same direction. 53 of these 55 sectors also had scans two years later, and in 36 of them (68%) the error remained $>10\mu\text{m}$ in the same direction. Similarly, 49 of these sectors had data available from a scan four years later; and in 33 of them (67%) the error remained $>10\mu\text{m}$ in the same direction. This suggests that in a majority of cases wherein a marked discrepancy existed between automated and manual segmentation at baseline, the discrepancy persisted over time.

Figure 2 displays the equivalent results for the Worst Sector Analysis, i.e. restricting the results to the sector with the highest error at the baseline time period. In 95% of eyes, this error at baseline was between -27.0 and $+27.2\mu\text{m}$. On the 6-month follow-up time period, the difference in errors in the same sector (regardless of whether it is now the worst sector or not) was between -16.6 and $+23.4\mu\text{m}$, as seen in Figure 2 (top left), with a correlation of 0.709 ($p<0.001$). The correlation between the error at baseline and the error at the 2-year time period was 0.689 ($p<0.001$, Figure 2, top right). The correlation at the 3-year time period was 0.662 ($p<0.001$, Figure 2, bottom left). The correlation at the 4-year time period was 0.721 ($p<0.001$, Figure 2, bottom right).

We examined repeatability of the testing by examining the standard deviation of the residuals from the longitudinal trend line for the All Sector Analysis. The standard deviation of the residuals was lower for automated segmentation when compared to manual refinement ($1.56\mu\text{m}$ vs. $1.80\mu\text{m}$, $p<0.001$). Similar results occurred when analyzing each of the 12 sectors individually, with automated segmentation having smaller residuals ($p<0.001$ for all). Overall, this suggests that automated segmentation (without manual refinement) has lower variability over time.

The longitudinal signal-to-noise ratio (rate of change divided by standard deviation of residuals in years (y)) for each sector had a mean of $-0.328y^{-1}$ for manually refined segmentations, versus $-0.374y^{-1}$ for automated segmentation ($p=0.009$). The proportion of sectors having a longitudinal signal-to-noise ratio less than $-1y^{-1}$, i.e. with a rate of change worse than one standard deviation of residuals per year, was 10.0% for manually refined segmentations, versus 12.1% for automated segmentation. This suggests that it is easier to detect rapid progression with automated segmentation.

Figure 3 shows an example of an eye in which errors in the automated segmentation resulted in an underestimation of sectoral RNFL thickness values at baseline, which persisted through the longitudinal series to the 4-year follow-up visit. The automated segmentation errors, in this case, occur along the posterior RNFL boundary where the B-scan path crosses

large vessels in the superior quadrant and in the nasal inferior sector at the position of another vessel crossing. These results also illustrate that the automated segmentation is slightly more repeatable, even if less accurate than the manually refined segmentation. However, this example also shows that nearly the entire thickness of the RNFL is occupied by the major vessels at their cross-section, which is an important caveat to consider when assessing “accuracy” versus repeatability because thinning and loss of axon bundles will have little effect on the apparent tissue thickness at these vessel crossings.

Figure 4 shows an example of an eye in which errors of automated segmentation resulted in an overestimation of sector RNFL thickness values at baseline, which persisted through the longitudinal series to the 4-year follow-up visit. In this case, the errors in the automated segmentation occur along the posterior RNFL boundary in the temporal inferior sector, which also may have been influenced by thin vessels (and their faint, narrow shadows).

Discussion:

Clinicians utilize optical coherence tomography (OCT) measurements of RNFL thickness to follow and manage glaucoma patients. These measurements depend on automated segmentation algorithms, which may commonly include significant errors.^{6,13} Our study is the first to examine longitudinal errors in automated segmentation. We defined error as the difference between RNFL thickness using automated segmentation when compared to manual refinement and determined how this changed over a 4-year period.

We found that the errors in both global RNFL and the 30-degree sector with the highest error at baseline persisted into the subsequent scan and continued longitudinally for 2 years, 3 years, and 4 years after the original OCT measuring RNFL. Our previous study⁶ suggests that automated segmentation usually results in a thinner measurement of nerve fiber layer thickness, but we also demonstrate herein that the alternative can occur as well. While automated measurement may be more reproducible, and have a higher longitudinal signal-to-noise ratio, it also may fail to accurately demonstrate thinning of the RNFL as changes in this value may be masked by the relatively larger artifact. Future studies should consider evaluating automated vs. manual refinement in their ability to detect progression using a separate measure of progression (i.e. visual fields) to determine whether automated segmentation is more likely to detect progression.

Prior studies document that automated segmentation measurements can be affected by normal anatomy (e.g. vessels) and artifacts,^{9,10,12} and these errors need to be evaluated and corrected using manual refinement.⁶ The fact that automated segmentation errors persist longitudinally actually resulted in the automated segmentation algorithm producing more consistent values of thickness over the series. The standard deviation of residuals from the trend line was significantly lower for automated segmentation both globally and in all 12 of the sectors. This suggests that manual refinement of segmentation may be preferable for cross-sectional identification of damage (since it is more accurate), but automated segmentation may be preferable for following eyes over time (since it is more longitudinally consistent). Important caveats to this suggestion include that the differences between the longitudinal signal-to-noise ratio for automated versus manually refined segmentations were

small and a consistently false negative error could have an undesirable impact on patient management (failure to detect real progression).

Our study is the first to examine automated segmentation longitudinally; however, there were limitations in this study. While the results demonstrate that the magnitude of error between automated and manually refined segmentation may persist over time in our population, these results may not be generalizable to individuals with more advanced glaucoma wherein the RNFL is thin and measurements may be more affected by artifacts or in those with highly myopic refractive errors/long axial lengths. Several studies have demonstrated that the ability to detect glaucomatous changes on OCT scans is more challenging in high myopes than emmetropes due to thinner peripapillary RNFL thickness and inaccuracies in delineating the RNFL.²³ If eyes with long axial lengths were included in this study, manual segmentation would likely perform better than automated segmentation. Studying RNFL segmentation error in eyes with higher refractive errors will be an interesting area of future study.

Other possible limitations in our study are related to data acquisition. Our data were collected in a research setting by an experienced operator with sufficient time to obtain high-quality images. In the clinical setting, both automated segmentation and manual refinement can be affected by the operator's training and experience in collecting high-quality OCT scans, recognizing errors of the automated segmentation, and appropriately refining RNFL margins (if enabled by the OCT instrument, or offline). Our study may underestimate the rate of RNFL segmentation error due to the fact that we excluded OCT scans with image quality less than 15. Moreover, we used real-time sweep averaging to improve the OCT B-scan image quality (by reducing speckle noise), which means even scans with a low Signal Quality Score (near 15) have better quality than would a single-sweep B-scan with the same quality score and presumably better segmentation accuracy.¹⁴

Future studies should examine specifically the role of blood vessels in causing errors in RNFL segmentation over time. As axon degeneration progresses, the major vessels begin to protrude out of the thinning RNFL (anteriorly toward the vitreous and posteriorly into the distal retina) creating additional challenges for automated segmentation algorithms.^{24,25} One solution is to separately segment out the vessels in order to account for their contribution to the RNFL.^{14,26} If these signals and the error they cause are relatively constant, newer segmentation algorithms could manage blood vessel artifacts to improve the accuracy of RNFL thickness measurements. For example, the information gained from OCT angiography could be used to segment vessels (at least their lumen) and account for (subtract) their contribution to the RNFL thickness parameter, in a manner similar to what other investigators accomplished either manually²⁶ or by automated detection of vessel shadows.¹⁴

In conclusion, the magnitude of error between automated segmentation and manual refinement of segmentation used to derive RNFL thickness measurements persisted from the baseline OCT scan to 4 years later. We found the magnitude of error frequently persisted when examining each 30-degree sector of RNFL thickness over time. The study demonstrated that while automated segmentation resulted in reproducible error, it can fail to

adequately represent RNFL thickness, which makes manual refinement a useful clinical tool. Thus the primary clinical message of this paper remains important: clinicians should carefully review all automated image segmentation and take the segmentation error into consideration in clinical diagnostics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

This work was supported in part by funding from Good Samaritan Foundation (Portland, OR) and The National Institutes of Health (R01 EY019674 Predicting the Rate of Progression in Glaucoma, PI: SD; and R01 EY020922 Functional Testing for Glaucoma, PI:SKG). We would like to acknowledge Cindy Albert at Devers Eye Institute (Portland, OR) for her assistance in data acquisition and skill in OCT imaging.

Disclosures: NNG and BF have no financial disclosures. SKG has the following disclosures: Equipment-Heidelberg Engineering GmbH. SLM has the following disclosures: Allergan (Research Funding), Bausch & Lomb (Consultant), Gore (Consultant), National Eye Institute (Research Funding), Ocular Therapeutix (Research Funding).

References:

1. Shin HY, Park HY, Jung Y, Choi JA, Park CK. Glaucoma diagnostic accuracy of optical coherence tomography parameters in early glaucoma with different types of optic disc damage. *Ophthalmology*. 2014;121(10):1990–1997. [PubMed: 24935284]
2. Miki A, Medeiros FA, Weinreb RN, et al. Rates of retinal nerve fiber layer thinning in glaucoma suspect eyes. *Ophthalmology*. 2014;121(7):1350–1358. [PubMed: 24629619]
3. Gardiner SK, Fortune B, Demirel S. Localized Changes in Retinal Nerve Fiber Layer Thickness as a Predictor of Localized Functional Change in Glaucoma. *Am J Ophthalmol*. 2016;170:75–82. [PubMed: 27491698]
4. Kafieh R, Rabbani H, Kermani S. A review of algorithms for segmentation of optical coherence tomography from retina. *J Med Signals Sens*. 2013;3(1):45–60. [PubMed: 24083137]
5. Tian J, Varga B, Tatrai E, et al. Performance evaluation of automated segmentation software on optical coherence tomography volume data. *J Biophotonics*. 2016;9(5):478–489. [PubMed: 27159849]
6. Mansberger SL, Menda SA, Fortune BA, Gardiner SK, Demirel S. Automated Segmentation Errors When Using Optical Coherence Tomography to Measure Retinal Nerve Fiber Layer Thickness in Glaucoma. *Am J Ophthalmol*. 2017;174:1–8. [PubMed: 27818206]
7. Asrani S, Essaid L, Alder BD, Santiago-Turla C. Artifacts in spectral-domain optical coherence tomography measurements in glaucoma. *JAMA Ophthalmol*. 2014;132(4):396–402. [PubMed: 24525613]
8. Liu Y, Simavli H, Que CJ, et al. Patient characteristics associated with artifacts in Spectralis optical coherence tomography imaging of the retinal nerve fiber layer in glaucoma. *Am J Ophthalmol*. 2015;159(3):565–576 e562. [PubMed: 25498118]
9. Kim NR, Lee H, Lee ES, et al. Influence of cataract on time domain and spectral domain optical coherence tomography retinal nerve fiber layer measurements. *J Glaucoma*. 2012;21(2):116–122. [PubMed: 21173702]
10. Lee SY, Kwon HJ, Bae HW, et al. Frequency, Type and Cause of Artifacts in Swept-Source and Cirrus HD Optical Coherence Tomography in Cases of Glaucoma and Suspected Glaucoma. *Curr Eye Res*. 2016;41(7):957–964. [PubMed: 26431251]
11. ElAlshareef RA, Dumpala S, Rapole S, et al. Prevalence and Distribution of Segmentation Errors in Macular Ganglion Cell Analysis of Healthy Eyes Using Cirrus HD-OCT. *PLoS One*. 2016;11(5):e0155319. [PubMed: 27191396]

12. Rufer F, Bartsch JJ, Erb C, Riehl A, Zeitz PF. Epiretinal membrane as a source of errors during the measurement of peripapillary nerve fibre thickness using spectral-domain optical coherence tomography (SD-OCT). *Graefes Arch Clin Exp Ophthalmol*. 2016;254(10):2017–2023. [PubMed: 27567904]
13. Ye C, Yu M, Leung CK. Impact of segmentation errors and retinal blood vessels on retinal nerve fibre layer measurements using spectral-domain optical coherence tomography. *Acta Ophthalmol*. 2016;94(3):e211–219. [PubMed: 26132774]
14. Fortune B, Reynaud J, Cull G, Burgoyne CF, Wang L. The Effect of Age on Optic Nerve Axon Counts, SDOCT Scan Quality, and Peripapillary Retinal Nerve Fiber Layer Thickness Measurements in Rhesus Monkeys. *Translational vision science & technology*. 2014;3(3):2.
15. Sun C, Wang JJ, Mackey DA, Wong TY. Retinal vascular caliber: systemic, environmental, and genetic associations. *Surv Ophthalmol*. 2009;54(1):74–95. [PubMed: 19171211]
16. Muraoka Y, Tsujikawa A, Kumagai K, et al. Age- and hypertension-dependent changes in retinal vessel diameter and wall thickness: an optical coherence tomography study. *Am J Ophthalmol*. 2013;156(4):706–714. [PubMed: 23876868]
17. Hood DC, Fortune B, Arthur SN, et al. Blood vessel contributions to retinal nerve fiber layer thickness profiles measured with optical coherence tomography. *J Glaucoma*. 2008;17(7):519–528. [PubMed: 18854727]
18. Spectralis HRA+OCT User Guide Software Version 5.1. 2010;26, 55–56.
19. Gardiner SK, Demirel S, Reynaud J, Fortune B. Changes in Retinal Nerve Fiber Layer Reflectance Intensity as a Predictor of Functional Progression in Glaucoma. *Invest Ophthalmol Vis Sci*. 2016;57(3):1221–1227. [PubMed: 26978028]
20. Gardiner SK, Boey PY, Yang H, Fortune B, Burgoyne CF, Demirel S. Structural Measurements for Monitoring Change in Glaucoma: Comparing Retinal Nerve Fiber Layer Thickness With Minimum Rim Width and Area. *Invest Ophthalmol Vis Sci*. 2015;56(11):6886–6891. [PubMed: 26501416]
21. Rosner B, Glynn RJ, Lee ML. The Wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics*. 2006;62(1):185–192. [PubMed: 16542245]
22. Gardiner SK, Fortune B, Demirel S. Signal-to-Noise Ratios for Structural and Functional Tests in Glaucoma. *Transl Vis Sci Technol*. 2013;2(6):3.
23. Chen JJ, Kardon RH. Avoiding Clinical Misinterpretation and Artifacts of Optical Coherence Tomography Analysis of the Optic Nerve, Retinal Nerve Fiber Layer, and Ganglion Cell Layer. *J Neuroophthalmol*. 2016;36(4):417–438. [PubMed: 27636747]
24. Tan O, Liu L, Zhang X, Morrison JC, Huang D. Glaucoma Increases Retinal Surface Contour Variability as Measured by Optical Coherence Tomography. *Invest Ophthalmol Vis Sci*. 2016;57(9):OCT438–443. [PubMed: 27409503]
25. Fortune B, Cull GA, Burgoyne CF. Relative course of retinal nerve fiber layer birefringence and thickness and retinal function changes after optic nerve transection. *Invest Ophthalmol Vis Sci*. 2008;49(10):4444–4452. [PubMed: 18566463]
26. Patel NB, Luo X, Wheat JL, Harwerth RS. Retinal nerve fiber layer assessment: area versus thickness measurements from elliptical scans centered on the optic nerve. *Invest Ophthalmol Vis Sci*. 2011;52(5):2477–2489. [PubMed: 21220552]

Figure 1. Bland-Altman plots of all RNFL sectors combined comparing magnitude of segmentation error at baseline to that at various follow-up time points. (Top left) Baseline versus next visit. (Top right) Baseline versus 2 years. (Bottom left) Baseline versus 3 years. (Bottom right) Baseline versus 4 years. Grey horizontal lines represent the limits of agreement. Long-term retinal nerve fiber layer segmentation study

Figure 2.

Bland-Altman plots of RNFL sector with most error at baseline comparing magnitude of segmentation error at various follow-up time points. (Top left) Baseline versus next visit. (Top right) Baseline versus 2 years. (Bottom left) Baseline versus 3 years. (Bottom right) Baseline versus 4 years. Grey horizontal lines represent the limits of agreement. Long-term retinal nerve fiber layer segmentation study

Automated Segmentations

Manually Refined Segmentations

Figure 3.

Example of automated segmentation errors resulting in an underestimation of RNFL thickness values. (Top) The circumpapillary OCT B-scan acquired at baseline is shown with automated segmentation of the anterior RNFL boundary in yellow and posterior RNFL boundary in red. There are two errors along the posterior boundary segmentation; one occurs at the crossing of the superior branch artery and vein (red arrows), resulting in RNFL thickness values that are 15–20% smaller than those for manually refined segmentations in sectors 3 and 4. The second error (red arrowhead) occurs at a vessel crossing in sector 9 (nasal inferior), resulting in values that are 10–15% below the manually refined values. A magnified view of the first error at baseline (inset) is shown persisting over time to the 4-year follow-up visit. (Bottom) The same scans are shown with manually refined segmentations. Long-term retinal nerve fiber layer segmentation study

Automated Segmentations

Manually Refined Segmentations

Figure 4.

Example of automated segmentation errors resulting in an overestimation of RNFL thickness values. (Top) The circumpapillary OCT B-scan acquired at baseline is shown with automated segmentation of the anterior RNFL boundary in yellow and posterior RNFL boundary in red. There is an error along the posterior boundary segmentation (red arrow), resulting in RNFL thickness values that are 10–25% larger than those for manually refined segmentations in sector 12. A magnified view of this error at baseline (inset) is shown persisting over time to the 4-year follow-up visit. (Bottom) The same scans are shown with manually refined segmentations. Long-term retinal nerve fiber layer segmentation study

Table

Participant and ocular data, Long-term retinal nerve fiber layer segmentation study
Participant Demographics. Data are presented in mean (standard deviation)

Age, years	64.5 (11.2)
Gender, % Female	58.2
Self-Reported Ethnicity, %	
White	93.0
AI/AN	1.9
African American	2.8
Hispanic/Latino	0.9
Asian	1.4
Mean Deviation Right Eyes, dB	−0.86 (2.8)
Mean Deviation Left Eyes, dB	−1.51 (3.5)
Central Corneal Thickness Right Eyes, microns	560.1 (38.5)
Central Corneal Thickness Left Eyes, microns	559.6 (37.3)