Published in final edited form as:

Br J Ophthalmol. 2012 April ; 96(4): 514–518. doi:10.1136/bjophthalmol-2011-300044.

Variation in optical coherence tomography signal quality as an indicator of retinal nerve fibre layer segmentation error

Lindsey S Folio^{1,2}, Gadi Wollstein¹, Hiroshi Ishikawa^{1,2}, Richard A Bilonick^{1,3}, Yun Ling^{1,3}, Larry Kagemann^{1,2}, Robert J Noecker¹, James G Fujimoto⁴, and Joel S Schuman^{1,2,5}

¹UPMC Eye Center, Eye & Ear Institute, Ophthalmology and Visual Science Research Center, Department of Ophthalmology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

²Department of Bioengineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

³Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

⁴Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

⁵The McGowan Institute for Regenerative Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

Abstract

Purpose—Commercial optical coherence tomography (OCT) systems use global signal quality indices to quantify scan quality. Signal quality can vary throughout a scan, contributing to local retinal nerve fibre layer segmentation errors (SegE). The purpose of this study was to develop an automated method, using local scan quality, to predict SegE.

Methods—Good-quality (global signal strength (SS) ≥ 6; manufacturer specification) peripapillary circular OCT scans (fast retinal nerve fibre layer scan protocol; Stratus OCT; Carl Zeiss Meditec, Dublin, California, USA) were obtained from 6 healthy, 19 glaucoma-suspect and 43 glaucoma subjects. Scans were grouped based on SegE. Quality index (QI) values were computed for each A-scan using software of our own design. Logistic mixed-effects regression modelling was applied to evaluate SS, global mean and SD of QI, and the probability of SegE.

Results—The difference between local mean QI in SegE regions and No-SegE regions was −5.06 (95% CI −6.38 to 3.734) ($p < 0.001$). Using global mean QI, QI SD and their interaction term resulted in the model of best fit (Akaike information criterion=191.8) for predicting SegE. Global

Correspondence to Dr Gadi Wollstein, UPMC Eye Center, 203 Lothrop Street, Pittsburgh, PA 15213, USA; wollsteing@upmc.edu.

Competing interests Dr Fujimoto is a scientific advisor and has stock options in Optovue. Drs Fujimoto and Schuman receive royalties for intellectual property licensed by Massachusetts Institute of Technology to Carl Zeiss Meditec.

Ethics approval University of Pittsburgh.

Contributors Lindsey S Folio: study conception and design, data collection, interpretation of data and manuscript preparation. Gadi Wollstein: study conception and design, interpretation of data and manuscript preparation. Hiroshi Ishikawa: study conception and design, software development pertinent for the study and critical revision of the manuscript for important intellectual content. Richard A Bilonick: statistical analysis and critical revision of the manuscript for important intellectual content. Yun Ling: statistical analysis. Larry Kagemann: critical revision of the manuscript for important intellectual content. Robert J Noecker: critical revision of the manuscript for important intellectual content. James G Fujimoto: critical revision of the manuscript for important intellectual content. Joel S Schuman: study conception and design, and critical revision of the manuscript for important intellectual content.

Provenance and peer review Not commissioned; externally peer reviewed.

mean QI 20 or SS 8 shows little chance for SegE. Once mean QI<20 or SS<8, the probability of SegE increases as QI SD increases.

Conclusions—When combined with a signal quality parameter, the variation of signal quality between A-scans provides significant information about the quality of an OCT scan and can be used as a predictor of segmentation error.

Glaucoma is characterised by typical cupping of the optic nerve head (ONH) and irreversible loss of retinal ganglion cell axons in a characteristic pattern.¹ Imaging technologies provide structural measurements of the retina that can be used for diagnosing and monitoring glaucoma.^{2–6} Optical coherence tomography (OCT) objectively measures retinal nerve fibre layer (RNFL) thickness, where micron-scale changes in RNFL can be used to identify glaucoma. OCT non-invasively generates cross-sectional images using low coherence light.⁷ The images obtained from the retina can be analysed with segmentation algorithms to determine RNFL thickness. To ensure the most accurate RNFL measurements, good-quality OCT images and precise image analysis must be achieved.

We previously demonstrated that image quality significantly affects OCT's ability to discriminate between healthy and glaucomatous eyes.⁸⁹ Other studies have also shown how image quality can impact RNFL thickness measurements¹⁰¹¹ and how RNFL thickness may be directly related to image quality at mediocre levels of signal strength (SS).¹² Image quality can be diminished by dry eyes,⁸ media opacities,¹³ small pupils or advanced disease, resulting in segmentation error (SegE). We hypothesise that considerable image quality variability exists within an OCT cross-sectional image that is poorly represented by the lone image quality parameter currently available—SS parameter. Subjective assessment of OCT scans clearly demonstrates fluctuations in scan quality, at the A-scan level, and SegE of the RNFL borders (figure 1), even in scans with acceptable global SS, as recommended by the device manufacturer.

The ability to better detect regions of SegE would help improve the consistency in scans deemed as acceptable between consecutive tests, technicians and various imaging sites, thus improving the accuracy of RNFL thickness measurements and the diagnostic capabilities of OCT. Moreover, detection of SegE might improve longitudinal OCT assessment by removing a major source of measurement variability. The purpose of this study was to determine if SegE could be predicted by assessing the signal quality variability between A-scans.

METHODS

Subjects

Six healthy, 19 glaucoma-suspect and 43 glaucoma subjects were randomly selected from the Pittsburgh Imaging Technology Trial, a prospective longitudinal study at the University of Pittsburgh Medical Center Eye Center. The study was approved by the University of Pittsburgh Institutional Review Board and adhered to the tenants of the Declaration of Helsinki and Health Insurance Portability and Accountability Act regulations. Informed consent was obtained from all subjects.

Inclusion criteria for healthy eyes were no history of intraocular surgery or retinal disease, normal-appearing ONH and visual field (VF) with pattern standard deviation (PSD) within 95% of the normal population and glaucoma hemifield test (GHT) within normal limits. Glaucoma suspects were characterised by a borderline GHT or PSD between 5% and 10% cut-offs. Glaucoma subjects had a GHT outside normal limits or PSD outside 95% of the normal population in at least two consecutive and reliable VFs.

Exclusion criteria for the study included history of ocular trauma or surgery other than uncomplicated cataract surgery, best-corrected visual acuity worse than 20/40, refractive error $> +3.0$ D or < -7.0 D. Both eyes were used if eligible for the study.

Study participants received a full eye examination including intraocular pressure measurement, gonioscopy, central corneal pachymetry, anterior and posterior segment biomicroscopy, VF testing and imaging with time-domain (TD)-OCT. All subjects underwent Swedish interactive thresholding algorithm (SITA; Humphrey Field Analyser; Carl Zeiss Meditec, Dublin, California, USA) 24–2 standard perimetry. Reliable VF tests were designated as those with $<30\%$ fixation losses, false-positive responses and false-negative responses.

OCT analysis

All subjects were scanned with TD-OCT (Stratus OCT; Carl Zeiss Meditec, Dublin, California, USA; software V.4.0.1). Although TD-OCT does not represent the most recent iteration of OCT, evaluation of this device remains critical because of its large installed base of users. Current studies have not been able to detect a significant difference between the glaucoma discrimination ability of TD-OCT and spectral domain (SD)-OCT devices.^{14–17} Additionally, there is a need for backward compatibility between devices in longitudinal glaucoma evaluation, with its characteristic slow disease progression. Moreover, the principle of the retinal segmentation software used to determine RNFL thickness is similar among TD- and SD-OCT, and the results of this study can be expected to be similar.

The fast RNFL scan was used in this study, consisting of three 3.4 mm diameter circles acquired sequentially after initial ONH scan centring. SS given by the device is for an entire B-scan only; therefore, quality index (QI) values for each individual A-scan were calculated using a previously described method.¹⁸ Briefly, QI is the product of two parameters, tissue signal ratio and intensity ratio, acquired from a signal intensity histogram of an OCT scan. The histogram represents the pixel distribution in the retinal tissue signals' range of pixel reflectivity. QI for each A-scan and the A-scan QI SD were computed.

Study protocol

Consecutive subjects were selected from the Pittsburgh Imaging Technology Trial cohort with RNFL scans with global SS ≥ 6 , the manufacturer-recommended quality cut-off criterion. Each subject's highest SS scan was used in the study. The SS and RNFL thickness measurements, from the three individual circle scans, were used in the analysis. SS was masked, and each circle scan was subjectively evaluated for SegE. The highly reflective inner and outer borders of the RNFL were traced, interpolating regions hidden by blood vessel shadowing. If the segmentation lines defined by TD-OCT deviated from the traced lines, excluding blood vessel regions, SegE was noted. The specific A-scan boundary locations where SegE occurred were recorded within B-scans. A scan was classified in the SegE group if the SegE region represented consecutively 15% or cumulatively 20% of the image. Scans were independently assessed for SegE twice to determine the evaluation reproducibility.

Statistical analysis

Cohen's κ statistic, in addition to raw agreement, was used to evaluate the two separate evaluations of SegE.^{19,20} Mixed-effects regression modelling accounted for using three circle scans for each subject and both eyes of some subjects. Linear mixed-effects models determined the SegE/No-SegE group difference for the tested variables. A logistic mixed-effects model evaluated the difference between mean QI in local regions of SegE and regions of No-SegE, within scans containing SegE. Logistic mixed-effects models assessed

SS, global mean and QI SD, and the probability of SegE. Models were evaluated using the Akaike information criterion (AIC), where a lower AIC means a smaller amount of information loss²¹ and thus a more plausible model given the observed data. A difference in AIC 10 indicates that the model with the higher AIC should not be considered as a plausible model.

Statistical analysis was performed using R Language and Environment for Statistical Computing package *nlme*.^{22,23} $p < 0.05$ was considered statistically significant.

RESULTS

Mean age was 63 years (range: 43–83), with 46 female participants. Mean global SS for all scans was 8.06 (95% CI 7.81 to 8.30), mean QI was 18.68 (17.76 to 19.59), and mean global RNFL thickness was 82.27 μm (78.32–86.22). Cross-section QI SD varied from 3.80 to 11.20. The SegE group included 22 glaucoma and 7 glaucoma-suspect eyes. The No-SegE group included 51 glaucoma, 27 glaucoma-suspect and 11 healthy eyes. Table 1 shows the mean parameter values for the SegE and No-SegE groups. The differences in mean global RNFL thickness, global SS and QI between the two groups were all statistically significant. The raw agreement between the SegE evaluations was 0.97 and Cohen's κ was 0.80 (0.69–0.91), suggesting substantial agreement allocating SegE.

Signal quality of the segmentation error region evaluated within a scan

The difference between local mean QI in SegE and No-SegE regions, within cross-section images containing SegE, was -5.06 (-6.38 to 3.73 ; $p < 0.001$). Logistic mixed-effects regression modelling to predict sections of A-scans containing SegE using global mean QI and local mean QI (identified based on A-scan SegE boundaries) identified both parameters to be statistically significant ($p = 0.008$ and $p < 0.001$, respectively). As global mean QI increased and local mean QI decreased, the probability of SegE increased.

Segmentation error prediction models

A logistic mixed-effects regression model using the following parameters as predictors of SegE was tested: age, gender, ethnicity, diagnosis, mean RNFL thickness, SS, global mean QI and QI SD. Only global mean QI ($p < 0.001$) and QI SD ($p < 0.001$) were statistically significant predictors of SegE (AIC=203.4). Global mean QI and SS were highly correlated and therefore not included simultaneously in any other model. Global mean QI, QI SD and their interaction term resulted in the model of best fit (AIC=191.8; table 2, Model 1). Table 2 shows all additional models with corresponding AIC values.

Segmentation error probability contour plots

A probability contour plot, estimating the probability of SegE, was created using Model 1 in table 2 (mean QI, QI SD and interaction; figure 2). The probability of SegE is almost negligible when global mean QI ≥ 20 or when QI SD ≤ 4.5 . When global mean QI < 20 , as QI SD increases, the probability of SegE increases.

Figure 3 provides the probability contour plot created by using Model 4 in table 2 (SS, QI SD and interaction). This model, however, has a higher AIC value; therefore, it does not fit the data as well as the model used in figure 2. Figure 3 shows that the lowest probability for SegE occurred when SS ≥ 8 for all values of QI SD or when QI SD ≤ 5 for all SS values.

DISCUSSION

Several studies have determined that RNFL thickness measurements are affected by the signal quality of the OCT scan.^{9–12} In this study, we have assessed how the OCT image analysis algorithm performance is related to the local variability in signal quality within cross-sectional images. The best prediction of SegE is achieved using mean QI, QI SD and their interaction term. Additionally, we have shown an increase in the probability for SegE as mean QI decreases while QI SD increases.

Healthy, glaucoma-suspect and glaucoma subjects were used to maximise the number of scans containing SegE while simulating the typical glaucoma clinic population in an academic institute. QI was used as the quality parameter because, unlike SS, it allowed the evaluation of signal quality on an individual A-scan level and the determination of the variability in signal quality across each cross-sectional image. Small pupils, decentralised scans, media opacities, highly damaged RNFL and dry eyes, among other causes, most likely created the local regions of low signal quality found in the cross-sectional images. Further analysis that associates the appearance and/or distribution of the local regions of low signal quality with the cause of the low signal may offer improved feedback to imaging technicians.

The SegE group was entirely composed of either glaucoma or glaucoma-suspect eyes, resulting in a statistically significant thinner RNFL thickness compared to the No-SegE group (table 1). No healthy eyes contained SegE, suggesting that diseased eyes are more prone to RNFL segmentation failure. A thick, healthy and highly reflective RNFL allows the image analysis algorithm to easily differentiate the retinal layers, thus resulting in fewer segmentation failures. Similarly, global SS and mean global QI were statistically significantly greater for the No-SegE group (table 1). An image with a lower signal quality will tend to have a greater chance of containing SegE.

Our results show that, within a cross-sectional image, regions with local SegE have a statistically significantly lower QI than local regions without SegE. OCT segmentation analysis depends on identifying certain typical features in the signal profile in each A-scan. Attenuated and deformed signal creates errors in segmentation algorithm function, and therefore, it is not surprising that regions with SegE had lower QI. Also of note, there was a statistically significant difference in QI in regions with and without SegE even though all our images were within the manufacturer-defined good-quality limits.

Global QI, QI SD and their interaction resulted in the best-fit model for SegE prediction (table 2). The probability contour plot of this model showed that with high global mean QI (mean QI ≥ 20), SegE is highly unlikely. Additionally, when mean QI is <20 , the probability of SegE increases as QI SD increases. It can therefore be concluded that in the presence of good global mean QI, no assessment of QI SD is needed and the probability of SegE is negligible. At a global mean QI <20 , QI SD should be assessed to predict the probability of SegE. Using the best SS model, a cut-off at the level of SS ≥ 8 can indicate SegE-free scans. However, the higher AIC indicates that the reliability of this model is lower than the QI model, and caution should be taken when using this SS cut-off.

A limitation of the study was the use of TD-OCT technology. We realise the hardware and software improvements in the newer SD-OCT might create a different result in a similar study. However, similar principles of TD-OCT segmentation analysis are used by the current iteration of the technology. Therefore, it can be expected that our major finding, that segmentation analysis is affected by local image quality, is likely to be applicable to SD-OCT. An additional study limitation includes the subjective assessment of SegE. However, this inevitable limitation was mitigated by the assessment of all images by an OCT expert

who was masked to the quantitative quality data. Moreover, a high level of agreement was noted between sessions in defining the SegE location. Additionally, the subjective assessment did not allow us to evaluate how the magnitude of the SegE in each scan was related to signal quality variability. This was not addressed because the current TD-OCT software does not allow the difference in the subjectively perceived and software-designated border to be easily calculated.

Achieving accurate and reproducible image analysis is important in glaucoma diagnosis and progression assessment. Asrani *et al* evaluated the prevalence of different types of artefacts in OCT scans, noting that decentration and SegE were the primary causes of RNFL artefact.²⁴ The prevalence of artefacts for images with SS 4 was 14% but only 4% when evaluating images with SS 6. However, the definition of SegE artefact in that study was not provided, which might explain the potential discrepancy with our results, where SS 8 was associated with negligible probability of SegE. Additionally, all our scans were properly centred, which could also contribute to this discrepancy. For longitudinal assessment, Wu *et al* determined that a consistent SS should be achieved across visits to ensure better reproducibility.¹¹ However, variability exists between the signal quality of individual A-scans and is not reported by the global SS. We suggest that a signal quality variation parameter (QI SD) be provided in addition to the global measure to allow a full evaluation of scan quality and the likelihood of SegE. We present this automated index as a first step in improving the accuracy of RNFL measurements by determining a measurable and reportable cause of variability.

In conclusion, we suggest a new approach to improve the detection of OCT SegE by combining a global signal quality parameter with QI SD.

Acknowledgments

Funding Supported in part by National Institutes of Health contracts R01-EY013178, R01-EY011289, R01-EY013516 and P30-EY008098 (Bethesda, Maryland, USA); Air Force Office of Scientific Research FA9550-070-1-0101; Eye and Ear Foundation (Pittsburgh, Pennsylvania, USA) and unrestricted grants from Research to Prevent Blindness (New York, New York, USA).

REFERENCES

1. Hernandez MR, Pena JD. The optic nerve head in glaucomatous optic neuropathy. *Arch Ophthalmol.* 1997; 115:389–95. [PubMed: 9076213]
2. Schuman JS, Hee MR, Puliafito CA, et al. Quantification of nerve fiber layer thickness in normal and glaucomatous eyes using optical coherence tomography. *Arch Ophthalmol.* 1995; 113:586–96. [PubMed: 7748128]
3. Weinreb RN, Shakiba S, Zangwill L. Scanning laser polarimetry to measure the nerve fiber layer of normal and glaucomatous eyes. *Am J Ophthalmol.* 1995; 119:627–36. [PubMed: 7733188]
4. Badalà F, Nouri-Mahdavi K, Raoof DA, et al. Optic disk and nerve fiber layer imaging to detect glaucoma. *Am J Ophthalmol.* 2007; 144:724–32. [PubMed: 17868631]
5. Stein DM, Wollstein G, Schuman JS. Imaging in glaucoma. *Ophthalmol Clin North Am.* 2004; 17:33–52. [PubMed: 15102512]
6. Wollstein G, Schuman JS, Price LL, et al. Optical coherence tomography longitudinal evaluation of retinal nerve fiber layer thickness in glaucoma. *Arch Ophthalmol.* 2005; 123:464–70. [PubMed: 15824218]
7. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science.* 1991; 254:1178–81. [PubMed: 1957169]
8. Stein DM, Wollstein G, Ishikawa H, et al. Effect of corneal drying on optical coherence tomography. *Ophthalmology.* 2006; 113:985–91. [PubMed: 16751039]

9. Sung KR, Wollstein G, Schuman JS, et al. Scan quality effect on glaucoma discrimination by glaucoma imaging devices. *Br J Ophthalmol*. 2009; 93:1580–4. [PubMed: 19692363]
10. Cheung CY, Leung CK, Lin D, et al. Relationship between retinal nerve fiber layer measurement and signal strength in optical coherence tomography. *Ophthalmology*. 2008; 115:1347–51. 1351.e1–2. [PubMed: 18294689]
11. Wu Z, Vazeen M, Varma R, et al. Factors associated with variability in retinal nerve fiber layer thickness measurements obtained by optical coherence tomography. *Ophthalmology*. 2007; 114:1505–12. [PubMed: 17367862]
12. Wu Z, Huang J, Dustin L, et al. Signal strength is an important determinant of accuracy of nerve fiber layer thickness measurement by optical coherence tomography. *J Glaucoma*. 2009; 18:213–16. [PubMed: 19295375]
13. van Velthoven ME, van der Linden MH, de Smet MD, et al. Influence of cataract on optical coherence tomography image quality and retinal thickness. *Br J Ophthalmol*. 2006; 90:1259–62. [PubMed: 16980644]
14. 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; 116:1257–63. 1263.e1–2. [PubMed: 19464061]
15. Park SB, Sung KR, Kang SY, et al. Comparison of glaucoma diagnostic capabilities of Cirrus HD and Stratus optical coherence tomography. *Arch Ophthalmol*. 2009; 127:1603–9. [PubMed: 20008715]
16. Moreno-Montañés J, Olmo N, Alvarez A, et al. Cirrus high-definition optical coherence tomography compared with Stratus optical coherence tomography in glaucoma diagnosis. *Invest Ophthalmol Vis Sci*. 2010; 51:335–43. [PubMed: 19737881]
17. Sehi M, Grewal DS, Sheets CW, et al. Diagnostic ability of Fourier-domain vs time-domain optical coherence tomography for glaucoma detection. *Am J Ophthalmol*. 2009; 148:597–605. [PubMed: 19589493]
18. Stein DM, Ishikawa H, Hariprasad R, et al. A new quality assessment parameter for optical coherence tomography. *Br J Ophthalmol*. 2006; 90:186–90. [PubMed: 16424531]
19. Cohen J. A coefficient of agreement for nominal scale. *Educ Psychol Measure*. 1960; 20:37–46.
20. Cardillo, G. [(accessed 11 Nov 2010)] Cohen's kappa: compute the Cohen's kappa ratio on a 2x2 matrix. 2007. <http://www.mathworks.com/matlabcentral/fileexchange/15365>
21. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974; 19:716–23.
22. Team RDC. R: a language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2008. <http://www.R-project.org>
23. Pinheiro, J.; Bates, D.; DebRoy, S., et al. nlme: linear and nonlinear mixed effects models. R package version 3.1-89. 2008.
24. Asrani S, Edghill B, Gupta Y, et al. Optical coherence tomography errors in glaucoma. *J Glaucoma*. 2010; 19:237–42. [PubMed: 19661819]

Figure 1. Cross-section image of the retina with signal strength within the manufacturer-recommended acceptable quality range. White lines delineate the retinal nerve fiber layer, with arrow indicating region of algorithm failure.

Figure 2.

Segmentation error (SegE) probability contour plot for mean quality index (QI) and QI SD (QI SD). The pink region represents an approximate probability of 1 for SegE, whereas the blue region represents an approximate probability of 0 for SegE. The values on the contour lines indicate the approximate probability of SegE for a point on a given line.

Figure 3.

Segmentation error (SegE) probability contour plot for signal strength (SS) and quality index SD (QI SD). The pink region represents an approximate probability of 1 for SegE, whereas the blue region represents an approximate probability of 0 for SegE. The values on the contour lines indicate the approximate probability of SegE for a point on a given line.

Table 1

Mean (95% CI) and difference estimates between the SegE and No-SegE groups

	SegE group mean (CI)	No-SegE group mean (CI)	SegE – No-SegE difference (CI)	p Value
Global RNFL thickness (μm)	76.98 (72.59 to 81.36)	83.90 (80.02 to 87.79)	–6.95 (–9.73 to –4.12)	<0.001
Global SS	7.26 (6.99 to 7.53)	8.30 (8.08 to 8.53)	–1.04 (–1.25 to –0.84)	<0.001
Global QI	16.28 (15.25 to 17.32)	19.42 (18.57 to 20.27)	–3.13 (–3.95 to –2.32)	<0.001
QI SD	7.06 (6.65 to 7.48)	6.95 (6.60 to 7.31)	0.11 (–0.19 to 0.41)	0.47

QI, quality index; RNFL, retinal nerve fibre layer; SegE, segmentation error; SS, signal strength.

Table 2

Logistic mixed-effects regression models with corresponding AIC values

Model	Model parameters	AIC
1	Mean QI, QI SD and interaction	191.8
2	Mean QI and QI SD	198.3
3	SS, mean QI and QI SD	200.3
4	SS, QI SD and interaction	204.9
5	SS and QI SD	227.4
6	SS	243.4
7	Mean QI	250.7

AIC, Akaike information criterion; QI, quality index; SS, signal strength.