Published in final edited form as:

Br J Ophthalmol. 2008 June ; 92(6): 806–809. doi:10.1136/bjo.2007.129312.

Sources of longitudinal variability in optical coherence tomography nerve-fibre layer measurements

L Kagemann^{1,2}, T Mumcuoglu¹, G Wollstein¹, R Bilonick¹, H Ishikawa^{1,2}, K A Townsend¹, M Gabriele^{1,2}, J G Fujimoto³, and J S Schuman^{1,2}

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

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

³ Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

Aims—The purpose of this study was to compare the day-to-day reproducibility of optical coherence tomography (OCT; StratusOCT, Carl Zeiss Meditec, Dublin, CA) measurements of retinal nerve-fibre layer (RNFL) measurements at time points 1 year apart.

Methods—One eye in each of 11 healthy subjects was examined using the StratusOCT fast RNFL scan protocol. Three fast RNFL scans with signal strength ≥ 7 were obtained on each of 3 days within a month. This protocol was repeated after 12 months. A linear mixed effects model fitted to the nested data was used to compute the variance components.

Results—The square root of the variance component that was attributed to the differences between subjects was 7.17 μm in 2005 and 7.28 μm in 2006. The square roots of the variance component due to differences between days within a single subject were 1.95 μm and 1.50 μm , respectively, and for within day within a single subject were 2.51 μm and 2.55 μm , respectively. There were no statistically significant differences for any variance component between the two testing occasions.

Conclusions—Measurement error variance remains similar from year to year. Day and scan variance component values obtained in a cohort study may be safely applied for prediction of long-term reproducibility.

Glaucoma is a slowly progressing chronic disorder presenting clinically as cupping of the optic disc, thinning of the retinal nerve-fibre layer (RNFL) and loss of visual function. Detection of structural changes in glaucoma patients requires precise and accurate measurements of RNFL thickness and optic disc structures over many years. Optical coherence tomography (OCT) is a noninvasive imaging technology that creates high-resolution cross-sectional images of the retina in vivo. RNFL thickness may be measured from these images, providing a quantitative

Correspondence to: G Wollstein, UPMC Eye Center, Department of Ophthalmology, University of Pittsburgh School of Medicine, 203 Lothrop Street, Eye and Ear Institute, Suite 834, Pittsburgh, PA 15213, USA; wollsteing@upmc.edu.
LK and TM contributed equally.

Competing interests: JGF and JSS receive royalties from intellectual property licensed by Massachusetts Institute of Technology to Carl Zeiss Meditec. GW receives grant support from Carl Zeiss Meditec and Optovue.

Ethics approval: Institutional Review Board and Ethics Committee approval were obtained for the study. This study followed the tenets of the Declaration of Helsinki and was conducted in compliance with the Health Insurance Portability and Accountability Act (HIPPA).

Patient consent: Informed consent was obtained from all subjects.

assessment of the presence and progression of glaucoma. Previous studies have demonstrated the ability of the commercially available time domain OCT (StratusOCT, Carl Zeiss Meditec, Dublin, CA) to detect and quantify glaucomatous retinal damage.^{1 2}

Reproducibility of StratusOCT measurements of RNFL thickness have been published;^{1 3–5} however these studies only determined reproducibility at a single point in time. It is then assumed that the device may be used for several years to monitor a chronic disease, and that measurement variability will remain unchanged. Since glaucoma progression occurs slowly, in most cases in a time frame of years, StratusOCT measurements must be reproducible over a similar range in order for RNFL measurements to be clinically useful in comparing measurements over time. The stability of StratusOCT RNFL measurements in a meaningful time frame relative to the disease it is used to manage remains unknown. The purpose of this study was to compare the day-to-day and year-to-year reproducibility of StratusOCT RNFL measurements.

Methods

One eye in each of 11 healthy subjects (hospital staff and family of patients) was used in this study (five men, six women, seven OD, four OS, 39 (SD 10) years of age).

Study design

This study used a two-factor hierarchical design. An identical study was performed 12 months after the initial study using the same eyes, equipment and protocol. Specifically, subjects were scanned three times on each of 3 days within 1 month. This exact protocol was repeated after 12 months.

Inclusion criteria

All subjects were healthy, with no family history of glaucoma or other neurological or microvascular pathologies. The subjects had a comprehensive ocular examination, and good-quality StatusOCT examination of the ONH region. Subjects were also required to have, in the same visit, reliable and normal Swedish Interactive Thresholding Algorithm (SITA) standard 24–2 perimetry. All had best-corrected visual acuity of 20/40 or better, refractive error between –6.00 and +3.00 dioptres, a normal ocular examination and a VF glaucoma hemifield test within normal limits. One eye was randomly selected from each subject if both eyes were eligible.

Scans

The fast RNFL scan protocol was used to measure RNFL thickness. This consisted of a series of three consecutive 3.4-mm-diameter circumpapillary scans with 256 A-scans each after initial setting by the operator. The overall average RNFL thickness, as well as RNFL thickness in the superior, inferior, temporal and nasal quadrants, was automatically generated by the machine's software. All scans were performed through an undilated pupil. Only scans with signal strength ≥ 7 were included in the study.

The StratusOCT used in this study was calibrated on a monthly basis throughout the study. Calibration measurements were performed by clinical imaging staff using a model eye provided by the manufacturer. In the event that calibration adjustments were required, these were performed by a field engineer for the manufacturer.

Statistical analysis

Three factors contributed to the variability in the study: differences between subjects, differences between days within the same subjects, and differences between scans on each day

for the same subject. Each measurement y_{ijk} is considered to be the sum of these three statistically independent and normally distributed components:

$$Y_{ijk} = \mu_i + \varepsilon_{ij} + \varepsilon_{ijk}$$

where μ_i is the true value for subject i (assumed to have a mean of μ and a standard deviation of σ_{sub}), ε_{ij} is the deviation from the true value for subject i due to the j th day effect (assumed to have a mean of zero and an SD of σ_{day}), and ε_{ijk} is the deviation from the j th day effect due to the k th scan-effect (assumed to have a mean of zero and SD of σ_{scan}). The three variances σ_{sub}^2 , σ_{day}^2 and σ_{scan}^2 are referred to as the variance components because their sum equals the total variance of the measurements, s_y^2 . The first variance component, σ_{sub}^2 reflects the natural variation in the population from which the subjects were selected (and will vary from population to population). The remaining variance components represent sources of error due to environmental factors, the measurement devices and how the devices are used, and possibly other sources.

The hierarchical design used in this study provided sample estimates of the subject, day, and scan variance components, namely s_{sub}^2 , s_{day}^2 , and s_{scan}^2 , respectively, using the method of restricted maximum likelihood (REML). The components for days and scans are arguably the most important, as they reflect the characteristics of the measurement device as typically used and can be used to compare the measurement reproducibility of different devices or the same device through time. For ease of comparison on the scale of the measurements, typically the square root of the variance components is reported and represents the SD of the respective population or error distributions described above.

The variance components can also be used to compute various intraclass correlation coefficients (ICCs). As the ICCs are proportions of the total variance, and the total variance depends on the particular population studied, the ICCs are not generally comparable across studies and so should not be used as a measure of reproducibility.

The variance components (subject, day, and scan) were estimated separately for the baseline visit and the 12-month visit. Bootstrap sampling was used to construct 95% CI for various functions of the variance components. The Bias Corrected accelerated (BCa) nonparametric CIs were computed from the estimated sampling distributions.⁶ In a number of cases, the adjustment to the naïve intervals was very large. The 95% BCa intervals were expected to be both more accurate (lower and upper bounds demarcate the central interval) and more correct (true coverage closer to 95%) than the naïve intervals, and therefore closer to the unknown exact 95% intervals. The BCa intervals are reported and indicate the uncertainty in the estimates—the wider the interval, the greater the uncertainty.

BCa CIs were also constructed for the ratio of years (follow-up divided by baseline) for each function of the variance components. Finally, BCa CIs for the difference between years (follow-up minus baseline) were computed.

Results

Eleven subjects were scanned three times on each of 3 days in each of 2 years, producing nine observations per year per subject. There were no missing measurements. Throughout the study, the thickness of a surface within a calibration phantom was 19.39 (0.93) μm . One month during the study, the phantom measurement was 31.19 μm , and the device was calibrated by the manufacturer's field engineer. The baseline average RNFL thickness was 245.64 (9.65) μm , with superior, inferior, nasal, and temporal baseline RNFL thicknesses of 128.91 (11.33) μm ,

131.49 (12.58) μm , 75.68 (15.29) μm , and 72.13 (14.89) μm respectively. The baseline disc and rim areas were 2.32 (0.36) mm^2 and 1.90 (0.36) mm^2 , respectively.

The SDs for the average RNFL thickness are summarised in table 1. SDs for nasal, temporal, superior and inferior quadrants are summarised in tables 2–5. The SD in overall average RNFL thickness between subjects was approximately 7 μm , while the SD within any single subject between days was approximately 1.75 μm , and the SD between multiple scans within any single subject on any single day was approximately 2.5 μm (table 1). SDs associated with differences between subjects, days, and scans were similar for both years of the study. All BCa 95% CIs for ratios between follow-up and baseline visits contained, and were close to 1, and BCa 95% CIs for differences between years contained 0, meaning that there were no statistically detectable differences in any component of the SD between the two years of the study.

SDs due to differences between subjects and differences between scans within the nasal quadrant were approximately two times greater than those for the overall average RNFL thickness (table 2). The SDs due to days for nasal quadrant and average RNFL thickness were similar. There were no significant differences between years based on BCa 95% CIs for ratios and differences. Similarly, the estimate of the difference in SDs between years due to days is close to zero, but the actual difference in SDs could be between -2.6 and 2.2 μm .

SDs due to differences between subjects and differences between scans within the temporal quadrant were approximately four times those for the overall average RNFL thickness (table 3). SDs due to days for temporal quadrant and average RNFL thickness were similar in 2006; however, the temporal RNFL thickness SD in 2005 was very low (0.0005 μm). This resulted in a significant difference in both the ratio and difference between years in SD due to days. The BCa 95% CI was very wide, with values ranging from 2.61 to 14246. The ratio between years was at least 2.61; however, given the size of the CI, the actual ratio is poorly localised in this dataset. The BCa 95% CI for difference, however, suggests that the actual difference in SD between the two years was 1.15 μm to 3.72 μm .

The SD due to differences between scans within the superior quadrant was greater than that for the overall average RNFL thickness (table 4). The estimate of SD for days in the superior quadrant was smaller in the follow-up visit than in the baseline visit. The SD between subjects in the superior quadrant was smaller than that of the average thickness measurement SD in the follow-up visit, and larger than that at baseline. Both the ratio and the difference between follow-up and baseline superior quadrant SDs were statistically significant. The ICC for days was also statistically significant.

The SDs due to differences between subjects, days within a single subject, and differences between scans on a single day within a single subject in the inferior quadrant were greater than those for the overall average RNFL thickness (table 5). There were no significant differences between years based on BCa 95% CIs for ratios or differences.

Discussion

The present study demonstrates that the overall variance in StratusOCT measurements remains consistent after 1 year. The levels of variance in mean RNFL thickness in the present study are similar to the previously published SD of 2.68 μm .¹ Overall, the day and scan variance components of StratusOCT RNFL measurements were low. This agrees with numerous studies that have found acceptable to excellent reproducibility of StratusOCT RNFL measurements.^{2,3,7–12} Isolated from other sources, the inter-subject overall average RNFL SD was slightly greater than 7 μm for both years. Within subjects, the SD due to scans and days was approximately 2 μm for both years.

The variance components for the averaged RNFL measurements were lower than any of the individual quadrant values, as would be expected due to the effect of averaging. As RNFL thickness measurements are related to the scan position around the optic nerve head, even a small shift can induce thickness changes in quadrant measurements. The overall thickness measurement is less sensitive, as a thickness increase in one quadrant is mostly compensated by a decrease in the contralateral quadrant. This was true for subject, day and scan components of variance, confirming previously reported findings.¹

The follow-up to baseline ratio of the day SD in the temporal quadrant was large, with an upper BCa 95% CI limit of more than 14 246. In absolute units, the temporal quadrant day SDs were relatively small for both years, with levels of 0.0005 μm and 2.076 μm in baseline and follow-up, respectively. In this case, the magnitude of the ratio is likely due to the small SD in the temporal quadrant in the baseline visit. The BCa 95% CI for difference suggests that the actual difference in variance between years due to day in the temporal quadrant differs by only 1.15 to 3.72 μm .

Superior quadrant SD for difference among subjects was significantly different between baseline and follow-up (9.799 μm and 6.317 μm , respectively). While subject SDs differed significantly between the two years of the study, there was no significant difference for scans between the two years. The ICCs are functions of the subject variance component, and the subject variance components were much larger than the other components; therefore, the significant difference for ICC within days between baseline and follow-up is in large part due to the difference between the subject variance components.

Subject variance components reflect the heterogeneity of the population and can differ greatly from population to population. Because the ICCs depend on the subject variance component, ICCs should be assessed with caution and are not, in general, comparable from study to study. The scan and day variances components are directly related to the way the devices are constructed and operated so that it would be expected that scan and day variance components would be approximately equivalent for all similar devices when the same protocol is used. However, further testing is required to confirm this hypothesis.

The present study was conducted on healthy subjects and might not be applicable to glaucoma subjects; however, a previous study demonstrates better short-term reproducibility in glaucoma subjects as compared with healthy subjects, thus suggesting that the present findings are applicable to glaucomatous eyes.

In summary, the SDs in StratusOCT measurements of overall average RNFL thickness vary between 2 μm and 3 μm within a single eye, with higher SDs for subsets of the overall scan (quadrants, clock hours). Variance components appear to remain consistent from year to year. Variance component values for days and for scans obtained in a cohort study may safely be used as a surrogate indicator of long-term reproducibility.

Acknowledgments

Funding: National Institutes of Health R01-EY13178-07, R01-EY11289-21, P30-EY008098, The Eye and Ear Foundation (Pittsburgh, PA) and unrestricted grant from Research to Prevent Blindness.

References

1. 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;45:1716–24. [PubMed: 15161831]

2. Schuman JS, Pedut-Kloizman T, Hertzmark E, et al. Reproducibility of nerve fiber layer thickness measurements using optical coherence tomography. *Ophthalmology* 1996;103:1889–98. [PubMed: 8942887]
3. Budenz DL, Chang RT, Huang X, et al. Reproducibility of retinal nerve fiber thickness measurements using the stratus OCT in normal and glaucomatous eyes. *Invest Ophthalmol Vis Sci* 2005;46:2440–3. [PubMed: 15980233]
4. Carpineto P, Ciancaglini M, Aharrh-Gnama A, et al. Custom measurement of retinal nerve fiber layer thickness using STRATUS OCT in normal eyes. *Eur J Ophthalmol* 2005;15:360–6. [PubMed: 15945005]
5. Stein DM, Wollstein G, Ishikawa H, et al. Effect of corneal drying on optical coherence tomography. *Ophthalmology* 2006;113:985–91. [PubMed: 16751039]
6. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci* 1986;1:54–77.
7. Alamouti B, Funk J. Retinal thickness decreases with age: an OCT study. *Br J Ophthalmol* 2003;87:899–901. [PubMed: 12812895]
8. Baumann M, Gentile RC, Liebmann JM, et al. Reproducibility of retinal thickness measurements in normal eyes using optical coherence tomography. *Ophthalmic Surg Lasers* 1998;29:280–5. [PubMed: 9571659]
9. Benson SE, Schlottmann PG, Bunce C, et al. Assessment of the reproducibility and repeatability of a method of grading macular subretinal fluid using optical coherence tomography. *Eye* 2006;20:1030–3. [PubMed: 16151485]
10. Blumenthal EZ, Williams JM, Weinreb RN, et al. Reproducibility of nerve fiber layer thickness measurements by use of optical coherence tomography. *Ophthalmology* 2000;107:2278–82. [PubMed: 11097610]
11. Jones AL, Sheen NJ, North RV, et al. The Humphrey optical coherence tomography scanner: quantitative analysis and reproducibility study of the normal human retinal nerve fibre layer. *Br J Ophthalmol* 2001;85:673–7. [PubMed: 11371486]
12. Klemm M, Rumberger E, Walter A, et al. Reproducibility of measuring retinal nerve fiber density. Comparison of optical coherence tomography with the nerve fiber analyzer and the Heidelberg retinal tomography device. *Ophthalmologie* 2002;99:345–51. [PubMed: 12043288]

Table 1
Average retinal nerve-fibre layer thickness: summary table for SDs and intraclass correlation coefficients (ICCs) (bias-corrected accelerated non-parametric 95% CI)

	Baseline (μm)	Follow-up (μm)	Follow-up/baseline	Follow-up-baseline (μm)
SD				
Subjects	7.170 (5.26 to 8.92)	7.278 (5.26 to 9.29)	1.015 (0.85 to 1.14)	0.109 (−0.91 to 0.86)
Days	1.950 (1.13 to 2.42)	1.502 (0.57 to 2.42)	0.770 (0.29 to 1.48)	−0.448 (−1.63 to 0.67)
Scans	2.506 (2.18 to 3.17)	2.550 (2.31 to 2.93)	1.018 (0.69 to 1.34)	0.045 (−0.84 to 0.67)
Days+scans	3.175 (2.54 to 3.81)	2.959 (2.18 to 3.63)	0.932 (0.66 to 1.29)	−0.215 (−1.18 to 0.77)
ICC				
Between days	0.836 (0.70 to 0.92)	0.858 (0.76 to 0.92)	1.026 (0.94 to 1.17)	0.022 (−0.06 to 0.12)
Within days	0.898 (0.78 to 0.94)	0.895 (0.76 to 0.94)	0.996 (0.94 to 1.08)	−0.003 (−0.06 to 0.07)

Table 2
Nasal retinal nerve-fibre layer thickness: summary table for SDs and
intraclass correlation coefficients (ICCs) (bias-corrected accelerated non-
parametric 95% CI)

	Baseline (μm)	Follow-up (μm)	Follow-up/baseline	Follow-up-baseline (μm)
SD				
Subjects	14.40 (9.39 to 19.60)	13.897 (9.22 to 18.82)	0.965 (0.80 to 1.20)	−0.503 (−2.64 to 2.18)
Days	2.162 (0.00 to 3.43)	2.052 (0.00 to 3.44)	0.949 (0.00 to 2.14)	−0.110 (−2.97 to 2.50)
Scans	6.238 (5.25 to 8.64)	6.805 (6.23 to 7.79)	1.091 (0.86 to 1.45)	0.568 (−0.71 to 2.05)
Days+scans	6.602 (5.77 to 7.97)	7.108 (6.06 to 8.03)	1.077 (0.89 to 1.30)	0.506 (−0.75 to 1.76)
ICC				
Between days	0.826 (0.65 to 0.90)	0.793 (0.63 to 0.89)	0.959 (0.77 to 1.10)	−0.034 (−0.18 to 0.07)
Within days	0.845 (0.64 to 0.90)	0.810 (0.48 to 0.88)	0.959 (0.82 to 1.03)	−0.035 (−0.15 to 0.03)

Table 3
Temporal retinal nerve-fibre layer thickness: summary table for SDs and intraclass correlation coefficients (ICCs) (bias-corrected accelerated non-parametric 95% CI)

	Baseline (μm)	Follow-up (μm)	Follow-up/baseline	Follow-up-baseline (μm)
SD				
Subjects	14.622 (5.20 to 18.75)	14.403 (6.50 to 18.70)	0.985 (0.78 to 1.38)	−0.218 (−2.04 to 2.30)
Days	0.0005 (0.00 to 0.49)	2.076 (1.67 to 1.96)	4505.116 (2.61 to 14246)	2.075 (1.15 to 3.72)
Scans	5.051 (4.20 to 6.09)	5.383 (3.87 to 8.06)	1.066 (0.72 to 1.67)	0.332 (−1.30 to 2.94)
Days+scans	5.051 (4.23 to 5.93)	5.769 (4.37 to 8.40)	1.142 (0.87 to 1.70)	0.719 (−0.67 to 3.07)
ICC				
Between days	0.893 (0.64 to 0.95)	0.862 (0.61 to 0.95)	0.965 (0.80 to 1.22)	−0.032 (20.16 to 0.11)
Within days	0.893 (0.68 to 0.95)	0.880 (0.65 to 0.97)	0.985 (0.82 to 1.07)	−0.014 (−0.15 to 0.05)

Significant values are shown in bold.

Table 4
Superior retinal nerve-fibre layer thickness: summary table for SDs and
intraclass correlation coefficients (ICCs) (bias-corrected accelerated non-
parametric 95% CI)

	Baseline (μm)	Follow-up (μm)	Follow-up/baseline	Follow-up-baseline (μm)
SD				
Subjects	9.799 (7.20 to 11.71)	6.317 (3.22 to 8.65)	0.645 (0.39 to 0.83)	−3.482 (−5.46 to −1.13)
Days	2.933 (0.00 to 3.79)	1.575 (0.00 to 3.28)	0.537 (0.00 to 1.30)	−1.358 (−4.08 to 1.72)
Scans	5.635 (4.87 to 6.98)	5.262 (4.89 to 5.85)	0.934 (0.79 to 1.10)	−0.374 (−1.37 to 0.38)
Days+scans	6.353 (5.51 to 7.21)	5.492 (4.47 to 6.90)	0.865 (0.72 to 1.08)	−0.861 (−1.87 to 0.42)
ICC				
Between days	0.704 (0.57 to 0.77)	0.569 (0.22 to 0.78)	0.809 (0.39 to 1.07)	−0.135 (−0.40 to 0.05)
Within days	0.767 (0.52 to 0.85)	0.605 (0.38 to 0.78)	0.788 (0.55 to 0.96)	−0.162 (−0.35 to −0.03)

Significant values are shown in bold.

Table 5
Inferior retinal nerve-fibre layer thickness: summary table for SD and intraclass correlation coefficients (ICCs) (bias-corrected accelerated non-parametric 95% CI)

	Baseline (μm)	Follow-up (μm)	Follow-up/baseline	Follow-up-baseline (μm)
SD				
Subjects	11.212 (7.64 to 14.43)	12.245 (8.48 to 14.85)	1.092 (0.80 to 1.45)	1.033 (−1.89 to 3.97)
Days	2.652 (0.37 to 3.31)	2.876 (1.38 to 3.33)	1.085 (0.62 to 2.05)	0.224 (−1.56 to 2.20)
Scans	5.992 (5.70 to 6.59)	5.650 (5.31 to 6.46)	0.943 (0.81 to 1.15)	−0.342 (−1.16 to 0.67)
Days+scans	6.553 (5.74 to 7.25)	6.340 (5.73 to 7.20)	0.968 (0.82 to 1.17)	−0.213 (−1.24 to 0.95)
ICC				
Between days	0.745 (0.56 to 0.85)	0.789 (0.61 to 0.85)	1.058 (0.88 to 1.32)	0.043 (−0.08 to 0.19)
Within days	0.787 (0.57 to 0.85)	0.832 (0.56 to 0.89)	1.057 (0.97 to 1.18)	0.045 (−0.03 to 0.12)