Repeatability and Reproducibility of Macular Thickness Measurements from the Cirrus Spectral-Domain Optical Coherence Tomography System

Paul F. Stetson PhD, Mary K. Durbin PhD, and Thomas M. Callan OD
Carl Zeiss Meditec, Inc., Dublin, CA USA

Abstract: Repeatability and Inter-System reproducibility of Cirrus™ HD-OCT macular thickness measurements were studied in a group of three subjects with no known retinal pathology. Method: Retinal scans were performed on the subjects using six production Cirrus HD-OCT systems, and macular thicknesses were determined from these data. Analysis of Variance (ANOVA) was used to evaluate Inter-Subject, Inter-System, Inter-Session, and random variability components of average thicknesses over specified sectors in the macula. Results: Repeatability and reproducibility of the average thickness over the foveal sector had a standard deviation of 2 μm or better. Average thicknesses over other sectors were comparable, also having typical repeatability and reproducibility standard deviations of approximately 2 μm. Consequently, sector-average thickness differences of 5-6 μm are generally significant with 95% confidence. Conclusions: The Cirrus system has a high degree of repeatability and reproducibility, more than adequate for clinical purposes.

INTRODUCTION

Optical Coherence Tomography (OCT) is an established tool for research and clinial management of retinal diseases.[1,2] Spectral Domain OCT (SD-OCT)[3] provides faster scanning than previous forms of OCT, allowing three-dimensional imaging of retinal tissue. Segmentation of 3-D data allows for better visualization of the retinal layers than was possible with earlier OCT systems. Carl Zeiss Meditec [Dublin, CA, USA] has recently introduced an SD-OCT system, the Cirrus HD-OCT, which offers both faster scanning and better axial resolution than its predecessor, the Stratus OCT™.

Prior versions of OCT systems from Zeiss have been characterized to determine the repeatability and reproducibility of measurements of macular thickness[4-7]. The goal of this study is to make a preliminary determination of this repeatability and reproducibility for the Cirrus HD-OCT system. Repeatability and reproducibility are essential for distinguishing normal subjects from subjects with disease as well as for following patients over time. This study uses an analysis of the variance observed across six instruments with three normal subjects to estimate the repeatability for a single subject and to evaluate variability from subject to subject and across systems.

METHODS

Data Collection

An in house study following an IRB approved protocol was conducted at Carl Zeiss Meditec, Inc. Three subject volunteers were recruited and consented to participate in the study.

All three subjects were scanned on six SD-OCT production instruments in three different study sessions. In the first session, four instruments were used. Two instruments were used in both the second and third sessions. Following the recommended procedure for scan acquisition, the subject's pupil was first centered and focused in an Iris Viewing camera on the system data acquisition screen, and then the system's line-scanning ophthalmoscope (LSO) was used to optimize the view of the retina. The OCT scan was aligned to the proper depth and patient fixation, and system polarization was optimized to maximize the OCT signal.

The scan patterns used cover an angular range of 20deg×20deg, which corresponds to approximately 6mm×6mm on the retina of a typical eye. These scan patterns are raster patterns described by the number of horizontal × vertical samples.

For the first session, both eyes of each subject were scanned once with a 200×200 scan and once with a 512×128 scan on each of the four instruments. For the second and third sessions, both eyes of each subject were scanned twice with each scan pattern on each instrument. One eye was randomly selected as the study eye for each scan pattern for data analysis.

All scans were saved and any cause for poor quality was noted.

Data Analysis

After data acquisition, the images were converted to a common multidimensional biomedicalimaging format from the Mayo Clinic Biomedical Imaging Resource. The image data was then analyzed by an algorithm[8] to find the internal limiting membrane (ILM) and Retinal Pigment Epithelium (RPE) layers.

To remove the effect of misalignment of the scan due to patient motion or operator effects, the foveal center was identified as the thinnest point inside a 0.5 mm radius of the center of the scan. The mean displacement between the scan center and the fovea was 0.10 mm for the 200×200 scans and 0.12 mm for the 512×128 scans; the maximum displacement was 0.3 mm.

This allowed the following nine measurements to be generated:

- C: Average central retinal thickness, using the area inside a 0.5-mm radius of the center of the fovea.
- S3, N3, I3, and T3: Average retinal thicknesses of the Superior, Nasal, Inferior, and Temporal quadrants, respectively, of a concentric ring with an inside diameter of 1 mm and an outside diameter of 3 mm.
- S6, N6, I6, and T6: Average retinal thicknesses of the Superior, Nasal, Inferior, and Temporal quadrants, respectively, of a concentric ring with an inside diameter of 3 mm and an outside diameter of 6 mm. Note that since the fovea will generally not be perfectly centered in the scan, the outer edges of these sectors may fall outside the scanned area. Average thicknesses take this variation in true sector area into account.

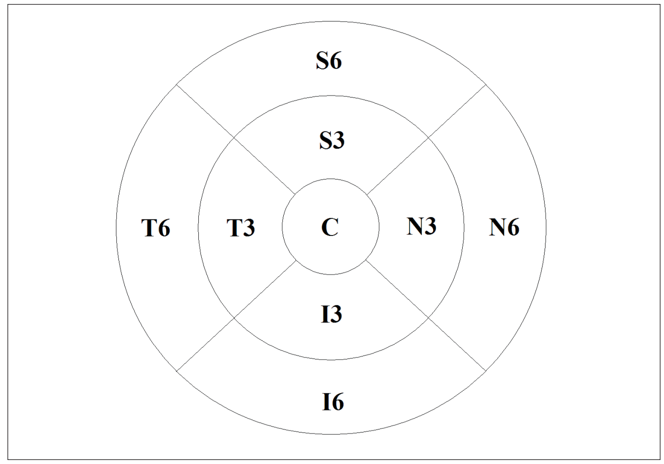

Figure 1. Sectors defined for a thickness map of the right eye. For the left eye, the N and T sectors are reversed.

Mean thickness and standard deviation were calculated as descriptive statistics for each system, and compared across systems. The data from the first two sessions and all six instruments was used for an “Inter-System” analysis, while the “Inter-Session” analysis used data from the second and third sessions (using only two of the instruments). For the Inter-System analysis, only the first of the two scans for each patient was used from the second session. Analysis of variance (ANOVA) was performed to determine random variability as well as the subject and system contributions to variability.

ANOVA is a family of statistical procedures which compare means by splitting the overall observed variance into different parts. For this analysis, each of the measurements was modeled as the linear combination of two main effects (subject and system, or subject and session) plus a constant term and a random error, using a “random effects” model.

RESULTS

Descriptive Statistics

The mean and standard deviation calculated for each sector for each system for both scan types are shown in Table 1.

Table 1. Descriptive statistics for the nine sectors summarizing thickness in the macular region as measured on six Cirrus systems and three volunteer subjects. All units are in µm.

Mean (std. dev.) 200×200 scan	System #1	System #2	System #3	System #4	System #5	System #6
C	251 (15.7)	253 (14.1)	253 (15.1)	252 (15.6)	251 (15.1)	253 (15.5)
S3	325 (3.2)	325 (3.9)	325 (3.4)	326 (4.3)	326 (4.7)	327 (1.5)
N3	330 (3.4)	330 (4.3)	330 (3.9)	330 (3.7)	330 (5.5)	333 (5.0)
I3	324 (8.7)	323 (7.7)	323 (8.4)	323 (7.0)	321 (5.6)	323 (7.0)
T3	313 (2.1)	313 (1.1)	313 (1.3)	314 (1.7)	313 (2.7)	313 (0.5)
S6	277 (11.8)	275 (10.4)	278 (12.7)	278 (11.2)	278 (9.3)	278 (14.3)
N6	301 (13.2)	302 (11.8)	301 (13.2)	301 (11.5)	300 (11.4)	304 (15.0)
I6	275 (23.0)	274 (21.3)	274 (21.5)	276 (23.4)	269 (12.0)	273 (13.6)
T6	261 (8.5)	261 (6.7)	261 (7.3)	263 (7.6)	263 (5.6)	261 (7.8)

Mean (std. dev.) 512×128 scan	System #1	System #2	System #3	System #4	System #5	System #6
C	247 (17.6)	248 (17.1)	252 (18.4)	249 (17.8)	250 (18.2)	248 (13.9)
S3	330 (3.3)	329 (2.4)	332 (1.5)	331 (0.8)	329 (0.7)	328 (3.5)
N3	331 (3.4)	332 (5.6)	335 (7.1)	332 (5.7)	330 (5.2)	330 (5.9)
I3	324 (6.7)	323 (7.8)	327 (12.4)	326 (10.0)	324 (9.5)	324 (7.9)
T3	314 (2.0)	313 (1.6)	316 (3.9)	315 (3.0)	313 (2.6)	313 (1.8)
S6	288 (11.3)	287 (11.0)	287 (13.2)	288 (12.9)	286 (13.0)	285 (7.7)
N6	300 (10.3)	299 (11.0)	302 (12.0)	300 (11.3)	300 (9.4)	297 (9.3)
I6	267 (11.7)	269 (14.6)	269 (15.2)	269 (14.0)	268 (12.9)	268 (10.9)
T6	266 (7.1)	264 (7.5)	265 (8.7)	266 (8.8)	264 (8.9)	264 (6.8)

Variance Components — Inter-System

Variance components were determined by performing an analysis of variance (ANOVA) with two main effects: Subject and System. This variance breakdown is shown in Table 2 and Table 3. Negative components are artifacts of the linear statistical model, as discussed later in this paper. All negativecomponents were statistically insignificant, with $p > 0.05$.

Table 2. Components of variance for nine sectors for 200x200 scan, with no interactions between Subject and System. Units of variance are µm².

Variance Components 200×200 scan	Subject	System	Error
C	229.6	0.2	1.3
S3	11.6	0.0	1.8
N3	14.7	0.5	4.2
I3	52.8	0.0	2.8
T3	0.6	-0.6	2.3
S6	132.5	-0.4	4.9
N6	158.7	0.1	3.8
I6	361.8	-2.5	25.2
T6	51.6	0.7	1.6

Table 3. Components of variance for nine sectors for 512x128 scan, with no interactions between Subject and System. Units of variance are µm².

Variance Components 512×128 scan	Subject	System	Error
C	293.8	3.1*	3.3
S3	1.3	1.2	4.0
N3	27.7	1.7	3.8
I3	80.3	0.3	5.6
T3	5.9	1.0*	0.8
S6	129.4	-1.0	6.6
N6	110.5	1.6*	1.5
I6	173.0	-0.3	4.0
T6	61.2	0.1	3.4

** Inter-System variance was significant, with $p < 0.05$.*

Variance Components — Inter-Session

For the Inter-Session study, two systems were used for two sessions one day apart. Each session included two scans of each type. An analysis of variance was performed with Subject and Session as the main effects – based on the relatively small components of variance for systems in the Inter-System study, the systems were treated as interchangeable for this study. The results are shown in the tables below.

Table 4. Components of variance (Subject and Session) for nine sectors for 200x200 scan. Units of variance are μm².

Variance Components 200×200 scan	Subject	Session	Error
C	247.8	-0.1	3.1
S3	5.4	0.7	4.8
N3	12.2	0.7	11.3
I3	36.3	0.6	4.1
T3	1.9	0.9*	0.8
S6	138.9	1.8*	4.2
N6	179.7	-0.7	8.2
I6	171.1	1.5*	3.7
T6	42.9	2.3*	2.1

Table 5. Components of variance (Subject and Session) for nine sectors for 512x128 scan. Units of variance are μm².

Variance Components 512×128 scan	Subject	Session	Error
C	308.5	-0.3	4.0
S3	11.1	1.5*	2.5
N3	15.8	0.3	4.6
I3	33.2	0.6	10.8
T3	3.2	1.1*	1.7
S6	97.8	2.1*	4.0
N6	98.8	-0.1	4.6
I6	144.2	2.1	9.6
T6	39.1	1.9	7.1

** Inter-Session variance was significant, with $p < 0.05$.*

DISCUSSION

Descriptive Statistics

The descriptive statistics from Table 1 show that there is good agreement among the six systems for all sectors. Since only three subjects were available for this study, the standard deviations could show considerable statistical fluctuation, but are generally consistent across the six instruments.

The standard deviations for the middle ring of sectors (S3, N3, I3, and T3) are markedly lower in general than for the center and the outer ring of sectors. This could indicate a lower thickness variation in these regions among normal subjects, but this study was not designed to present sufficient numbers of subjects to make this conclusion.

Variance Components — Inter-System

The Inter-System components of variance were all found to be insignificant for the 200×200 scans. For the 512×128 scan pattern, some sectors (includingthe central region) had significant Inter-System variance, but these terms were small, in all cases corresponding to a standard deviation of less than 2 μm (variance components less than 4 μm²).

Some components of variance were reported with values less than zero. A variance component can appear negative as an artifact of the mathematical ANOVA model used. Negative components are always not significant, but may be required to make the model consistent. These values can be treated as zero contributions to the overall variability.

The contribution of the system to total variance is less than 3% for all sectors. In general, imaging a subject across multiple systems will contribute less than 1 μm of standard deviation to the total variability.

Repeatability and Reproducibility — Inter-Session

Some sectors had significant Inter-Session variance, but in all cases these terms corresponded to standard deviations of less than 2 μm (variance components less than 4 μm^2).

For regions 0.5 – 1.5 mm distant from the foveal center (sectors S3, N3, I3, and T3), we see very little variation between our three normal subjects in Table 1. For the central and outer sectors, the variability of these thicknesses seen in normals is high compared to variations in measurement from visit to visit.

The Inter-Session repeatability and reproducibilitycan be calculated from the variance components. Repeatability is reported as the standard deviation as calculated from the random component of variance (the error terms in Tables 4 and 5).

Reproducibility is reported as the standard deviation as calculated from all sources of variability. The Inter-Session reproducibility includes the random variance plus the positive Inter-Session variance components reported in Tables 4 and 5.

Repeatability and reproducibility standard deviations are shown in Tables 6 and 7. If two scan measurements differ by less than 2.77 times the reproducibility as reported here, then the difference is not statistically significant[10]. If they differ by more than that amount, the difference is significant at the 95% confidence level.

Table 6. Intra-session repeatability and reproducibility standard deviations for nine macular sectors using the 200x200 scan. All units are in μm.

200×200 scan	Repeatability	Reproducibility
C	1.8	1.8
S3	2.2	2.4
N3	3.4	3.5
I3	2.0	2.2
T3	0.9	1.3
S6	2.1	2.5
N6	2.9	2.9
I6	1.9	2.3
T6	1.5	2.1

Table 7. Intra-session repeatability and reproducibility standard deviations for nine macular sectors using the 512x128 scan. All units are in μm.

512×128 scan	Repeatability	Reproducibility
C	2.0	2.0
S3	1.6	2.0
N3	2.2	2.2
I3	3.3	3.4
T3	1.3	1.7
S6	2.0	2.5
N6	2.2	2.2
I6	3.1	3.4
T6	2.7	3.0

For the average retinal thickness over the central region, the repeatability and reproducibility are the same, because the Inter-Session component of variance was negative. As described above, negativecomponents are a result of the statistical model and may be neglected in such calculations. All negative components encountered in this study were statistically insignificant.

For the 200×200 scan, central thickness averages differing by more than 5.0 μm are significant with 95% confidence; for the 512×128 scan, the differences must be more than 5.5 μm. For the other sectors, the thresholds of significance tend to be slightly higher, typically about 6 μm.

Conclusion

Average thicknesses over regions of the macula were studied for three subjects using six Cirrus SD-OCT systems. There is generally good agreement across multiple systems for the macular thicknesses averaged over the three subjects studied.

Analysis of variance from the six systems showed Inter-System variation to be a small component of overall variance. Repeat scans performed using two of the systems showed Inter-Session variation also to be a small component of overall variance.

Repeatability and reproducibility of thickness measurements, with a typical standard deviation of approximately 2 μm, is more than adequate from a clinical perspective. Performance of different scan patterns was similar for all aspects of the study.

REFERENCES

1. J. S. Schuman, C.A. Puliafito, and J.G. Fujimoto, *Optical Coherence Tomography of Ocular Diseases, Second Edition*, SLACK Incorporated, Thorofare, NJ, USA, 2004.
2. R. Brancato and B. Lumbroso, *Guide to Optical Coherence Tomography Interpretation*, INC Innovation-News-Communication, Rome, Italy, 2004.
3. A. F. Fercher, C. K. Hitzenberger, W. Drexler, G. Kamp, I. Strasser, H. C. Li, "In vivo Optical Coherence Tomography in Ophthalmology", in *Medical optical tomography: functional imaging and monitoring*, G. J. Müller, ed., 355-370, SPIE (1993).
4. D. Koozekanani, C. Roberts, S. E. Katz, and E. E. Herderick, "*Intersession Repeatability of Macular Thickness Measurements with the Humphrey 2000 OCT*", Invest. Ophthalmol. Vis. Sci; 41: 1486 - 1491., May 2000.
5. S. Muscat, S. Parks, E. Kemp and D. Keating, "*Repeatability and Reproducibility of Macular Thickness Measurements with the Humphrey OCT System*" Invest. Ophthalmol. Vis. Sci; 43(2): 490 - 495, February 1, 2002.
6. L. A. Paunescu, J. S. Schuman, L. L. Price, P. C. Stark, S. Beaton, H. Ishikawa, G. Wollstein, and J. G. Fujimoto, "*Reproducibility of Nerve Fiber Thickness, Macular Thickness, and Optic Nerve Head Measurements Using StratusOCT*", Invest. Ophthalmol. Vis. Sci., 45(6): 1716 – 1724, June 1, 2004.
7. A. Polito, M. Del Borrello, M. Isola, N. Zemella, and F. Bandello, "*Repeatability and Reproducibility of Fast Macular Thickness Mapping With Stratus Optical Coherence Tomography*", Arch. Ophthalmol. 123:1330-1337 (2005).
8. G. Gregori, R.W. Knighton, S. Jiao, et al., "*3-D OCT Maps of Retinal Pathologies*", Invest. Ophthalmol. Vis. Sci. 46: E-Abstract 1055 (2005).
9. J. M. Bland, D. G. Altman, "*Statistics Notes: Measurement Error*", BMJ 312: 1654 (1996).

The authors would like to acknowledge Melissa Horne, Kelly Soules, and Alex Thorne for their assistance in acquiring data for this study.

Carl Zeiss Meditec, Inc.
5160 Hacienda Drive
Dublin, CA 94568
USA

Phone: 1-800-342-9821
Fax: 1-925-557-4101
info@meditec.zeiss.com
www.meditec.zeiss.com

CIR.1682