Quality and Reproducibility of Retinal Thickness Measurements in Two Spectral-Domain Optical Coherence Tomography Machines

Ilse Krebs,[1,2] *Eva Smretschnig,*[1,2] *Sarah Moussa,*[2] *Werner Brannath,*[3] *Irene Womastek,*[3] *and Susanne Binder*[1,2]

Purpose. To evaluate the accuracy and reproducibility of retinal thickness measurements in exudative age-related macular degeneration (AMD) by the Spectralis (Heidelberg Engineering, Heidelberg, Germany) and the Cirrus (Carl Zeiss Meditec, Dublin, CA) optical coherence tomography (OCT) systems.

Methods. Eyes with exudative age-related macular degeneration were randomly assigned to one of eight groups, each different in the sequence of examiner and OCT system. The 512 × 128 cube program of the Cirrus and the 30° × 25° volume scan containing 32 lines of the Spectralis were performed twice. The correlation between the examinations was expressed by the interclass correlation coefficient (ICC).

Results. Enrolled in the study were 112 patients and 112 eyes (mean age, 76.5 ± 7.9 years; range 51–89), with 14 patients in each group. The mean error scores per line were 0.53 and 0.52 in the Cirrus, significantly ($P < 0.001$) lower than in the Spectralis (0.83 and 0.98). For automatic central retinal thickness (CRT), the ICC for Cirrus (all examinations calculated) was 0.61 for groups 1 to 4 (the same examiner) and 0.65 for groups 5 to 8 (two different examiners); for Spectralis (13.4% not calculated) the ICC was 0.93 for groups 1 to 4 and 0.86 for groups 5 to 8. After error correction, the Cirrus ICC improved to 1.0 and 0.99 and the Spectralis ICC to 1.0 in both groups.

Conclusions. Considerable differences were found between the two systems, both of which incorporate the spectral-domain technology. Different positioning of segmentation lines, control of localization, density of included scan lines, and number of available maps explain the differences in segmentation quality and reproducibility. Manual correction of segmentation and centralization improves the reproducibility. (ClinicalTrials.gov number, NCT00927303.) (*Invest Ophthalmol Vis Sci.* 2011;52:6925–6933) DOI:10.1167/iovs.10-6612

Retinal thickness measurement with optical coherence tomography (OCT) has gained increased importance in the documentation of treatment effects of neovascular age-related macular degeneration (AMD).[1] Providing comparable information concerning the activity of a lesion, OCT has also partly replaced fluorescein angiography in the follow-up of treated neovascular AMD.[2,3] Today, the most frequently applied treatment of neovascular AMD is intravitreal injections of antagonists of the vascular endothelial growth factor (anti-VEGF). A fixed treatment regimen with monthly injections was used in large multicenter studies.[4–6] A flexible treatment regimen based on retinal thicknesses obtained by OCT has recently been introduced.[1] The data of already published studies are based on the Stratus OCT (Carl Zeiss Meditec, Inc., Dublin, CA). In the past, it was the only OCT system available, and it is still the most widely used. Recently, an OCT technology has been developed that is based on the spectral domain (SD), in contrast to the time domain (TD) of the Stratus. SD technology offers a series of advantages, such as higher resolution and faster scan acquisition. Whereas TD OCT is employed only by Stratus, there are several machines on the market that use SD technology.

Retinal thickness measurement with OCT is based on automatically set threshold algorithm lines at the inner limiting membrane (ILM) and in the area of hyperreflectivity of the outer retina (the exact position varies between systems). Choroidal neovascularization is located in the area of the retinal pigment epithelium (RPE); therefore, we also have to assume that pathology-related algorithm line failures influence the measurement. Variable fixation in patients with impaired visual acuity may also affect the quality of retinal thickness measurements. First experiences with SD-OCT systems have already been published.[7–9] However, most of these articles deal with the reproducibility of measurements in healthy subjects. In the present study, the reproducibility of retinal thickness measurement in patients with neovascular AMD was examined with two different SD-OCT systems: the Spectralis (Heidelberg Engineering, Heidelberg, Germany) and the Cirrus (Carl Zeiss Meditec, Dublin, CA). Although both incorporate SD technology, they have different algorithms. The consequences of these differences and the accuracy of positioning of the boundary lines should be examined.

Patients and Methods

In this prospective study, 112 consecutive patients were recruited from September through December 2009. Included were patients aged more than 50 years who were currently under treatment or in whom neovascular AMD was under control after treatment with anti-VEGF. Eyes with macular disease other than neovascular AMD and eyes with myopia of more than 5 D were excluded. Only one eye of each patient was included: either the eye currently under treatment or, when both eyes were under treatment, the eye with the better distance acuity.

Each of two examiners (ES and SM) performed the Spectralis and Cirrus examinations twice. The sequences of the examinations and the

From the [1]The Ludwig Boltzmann Institute for Retinology and Biomicroscopic Laser Surgery, Vienna, Austria; the [2]Department of Ophthalmology, Rudolph Foundation Clinic, Vienna, Austria; and the [3]Core Unit for Medical Statistics and Informatics, Medical University, Vienna, Austria.

Supported by an unrestricted research grant from the L. Boltzmann Institute (SB).

Submitted for publication September 21, 2010; revised December 30, 2010, and April 8, June 14, and July 15, 2011; accepted July 18, 2011.

Disclosure: **I. Krebs**, None; **E. Smretschnig**, None; **S. Moussa**, None; **W. Brannath**, None; **I. Womastek**, None; **S. Binder**, None

Corresponding author: Ilse Krebs, Department of Ophthalmology, Rudolf Foundation Clinic, Juchgasse 25, 1030 Vienna, Austria; ilse.krebs@wienkav.at.

examiner performing them were randomized. Patients were assigned to one of eight groups, by using the web-based randomization program located at the Center of Medical Statistics and Informatics from the Medical University of Vienna. Repeated scans by Cirrus and Spectralis were performed by the same examiner in groups 1 to 4 (repeatability) and by different examiners in groups 5 to 8 (reproducibility).

Examinations

Distance acuity was tested with Early Treatment of Diabetic Retinopathy Study (ETDRS) charts at a 2-m distance. When all the letters were read correctly, the test was repeated at a 4-m distance, and the score was multiplied by 2. Thereafter, the pupils were dilated for OCT.

Correct sitting position, position of the head, focus of the video image, and centralization of the scan area were carefully controlled before recording started. With Cirrus, the cube 512 × 128 program was performed consisting of 128 horizontal lines of 512 A-scans. With Spectralis, 30° × 25° volume scans containing 32 lines of 512 A-scans were chosen. The scans were acquired in high-speed mode. The automatic real-time (ART) mode was activated and set to 5× (creating a mean image of five repeated identical B-scans, for noise reduction). The first examination was marked as the reference. If the second examination was marked as the follow-up, the same setting and position of the scans were automatically applied.

For both examinations, an internal fixation light or, if the patient was unable to see it, an external fixation light was used.

Evaluation

Two examiners (ES and IK) independently evaluated the scans to determine whether the automatically set threshold algorithm lines were set correctly. Examinations differently evaluated by the examiners were reviewed by both. Both OCT systems position the anterior line at the ILM. The posterior line is positioned at the third hyperreflective line of the outer retina corresponding to the RPE in both systems—in Spectralis, even more posterior at the posterior surface of the third line corresponding to Bruch's membrane. As was reported in a prior publication, the algorithm line remains at the base of the detachments of the RPE in the Spectralis, whereas it follows the detached RPE in the Cirrus.[10] The examiners were instructed to judge the lines that followed the above-mentioned criteria to be correct. The manufacturer of Spectralis specifies Bruch's membrane as the posterior reference line. Therefore, only this position was judged to be correct, although Cirrus accounts for RPE detachments and Spectralis does not. An analysis of the position of the automatically set line in the area of the RPE was added.

One of the examiners (IK) graded the failures in a manner similar to that recommended by Sadda et al.[11] for the Stratus. The length, depth, and position of every failure in every line were documented, and thereafter a failure score/line was calculated specifying the accuracy of line segmentation and therefore the quality of retinal thickness measurement. Additionally, a second grading was performed. Failures were noted that affected the central 1-mm-diameter area, which is measured in the calculation (central error score) of central retinal thickness (CRT). The scans bordering the ETDRS subfield were identified, and the scans within these scans (usually the central 1 mm of Cirrus scans 53–75 and of Spectralis scans 14–19) were graded according to the scale that Han et al.[12] introduced: stage 0, no failure; stage 1, a central error one third or less of retinal thickness; stage 2, error more than one third but less than or equal to two thirds of retinal thickness; and stage 3, more than two thirds of retinal thickness. The nature of the threshold algorithm line errors was also described. Motion artifacts, including blinking, saccades, and flipped scans, were recorded, even when they did not cause any failures in line positioning. Spectralis volume scans last up to 5 minutes and are strenuous for elderly patients. Repeating these scans could have influenced the examinations that followed; therefore, repetitions of scans were not allowed. The number of Cirrus scans with a signal strength of <5 (quality levels, 0–10) in and of Spectralis scans <15 dB (0–30 dB) was documented.

Thereafter, the same examiner (IK) corrected the algorithm line errors using the built-in software of both systems. The scans of the first examination of both OCTs were corrected twice, to examine the repeatability of error correction. With the Cirrus, the lines were redrawn at the correct position on the screen with the computer mouse. With the Spectralis, the points were reset until the line was set correctly. Similarly, missing lines were created. With the Cirrus, the center of the ETDRS area was set at the foveal region in both examinations. The corrected values of CRT were recorded and compared to the automatically obtained values.

Statistical Evaluation

Repeatability and reproducibility of retinal thickness measurements were analyzed with the data from groups 1 to 4 and groups 5 to 8, respectively. Differences between the first and second automatic and corrected measurements were assessed by Bland-Altman plots and 95% confidence intervals for the mean differences of the two repeated measurements. The ICC was computed via a variance component analysis.

The influence of the independent variables examiner, measurement, OCT system, and OCT*examiner interaction on CRT (corrected and automatic measurements) was investigated by likelihood ratio tests in groups 1 to 4 and groups 5 to 8 separately, by calculating mixed models with patient number as the random effect.

To investigate differences between the error rates of the two OCT systems, we calculated a logistic regression model with patient number as the random effect and with the dichotomous variable indicating an error as the dependent variable and OCT system, examiner, age, and distance acuity as the independent variables. For this analysis, only the first measurement was considered.

To analyze the effect of the central error (the maximum central error of the two repeated measurements) on the automatic measurements of CRT, we calculated a mixed model with the logarithm of the absolute difference between the two measurements of the Cirrus and Spectralis as the dependent variable. OCT system and the interaction between central error and system were additional independent variables, and patient number was the random effect.

The correlation of the signal strength between the first and second examinations by the Cirrus and Spectralis was analyzed by Spearman correlation. The correlation of the signal strength between the two OCT systems was analyzed in a similar manner, using the minimum signal strength of the two repeated measurements. Spearman correlations were calculated for Cirrus and Spectralis separately, to investigate the correlation between the absolute difference of automatic measurements of CRT and signal strength (the minimum signal strength of the two repeated measurements).

A conversion formula between the corrected CRT measurements of Cirrus and Spectralis was calculated by a linear model, with the mean of the two repeated measurements of the Cirrus as the dependent variable and the mean of the two repeated measurements of the Spectralis as the fixed factor.

For the automatic and corrected measurements of CRT, we calculated the 75th, 80th, 85th, 90th, and 95th quantiles of the absolute difference between measurements, for groups 1 to 4 and groups 5 to 8 separately, as well as for all groups together.

Statistical analyses were conducted with the statistical program R, version 2.11.0. For all analyses, the significance level was set to 0.05.

Sample Size Justification

The sample size calculation is for the sign test at two-sided level 0.05 and a power of 80% assuming that, for 20% of the patients, Spectralis will be incorrect, whereas Cirrus will be correct, and in 5% of the cases Cirrus will be incorrect and Spectralis correct. These assumptions are based on pilot data,[10] and they give a total sample size of 92. Assuming a dropout rate of approximately 15%, 112 patients (i.e., 14 patients per group) would be sufficient to meet the power requirement.

The study was confirmed by the national review board. All patients signed a written consent, and data collection complied with the Declaration of Helsinki.

Results

Demographic Data

Enrolled were 112 eyes of 112 patients with a mean age of 76.5 ± 7.9 years (range, 51–89); 37.5% were men and 62.5% were women. Fourteen patients were randomized in each of the eight groups. In groups 1 to 4, the same examiner (repeatability) and, in groups 5 to 8, different examiners (reproducibility) performed the repeated measurements. Active occult lesions without RPE detachment amounted to 34.8%, 41.1% were occult lesions with fibrovascular RPE detachment, 1.8% were predominantly classic lesions, and 22.3% were fibrotic lesions. The mean distance acuity was 64 ± 31 letters (range, 1–130) for all patients: 68 ± 29 for groups 1 to 4 and 60 ± 31 for groups 5 to 8. In only 8% was the distance acuity of the fellow eye at least 1 line better than that in the study eye. In 2.7% of the Cirrus scans and in 7.1% of the Spectralis scans, an external fixation light was useful.

Correlations

For automatic measurements, the ICC for Cirrus was 0.61 for groups 1 to 4 and 0.65 for groups 5 to 8. For Spectralis, an ICC of 0.93 for groups 1 to 4 and 0.86 for groups 5 to 8 was calculated. Cirrus created a map in all patients and all groups, whereas Spectralis did not create a map in some cases, with the missing-line failure at 7.1% in groups 1 to 4 and 19.6% in groups 5 to 8. Therefore, in these Spectralis OCT examinations, no CRT value was available, and reproducibility analysis could not be performed. Of the examinations affected, 46.7% were the first examination, 40% the second, and 13.3% both. Analyses of repeatability and reproducibility are presented in Table 1. To compensate for these software-related differences, an analysis based on similar objective criteria for both systems was added to Table 1 (scans of minimum signal strength).

The mixed model revealed a significant effect of OCT system ($P < 0.0001$) but not of measurement ($P = 0.8$ and 0.1), examiner ($P = 0.2$ and 0.9), or interaction term OCT system*examiner ($P = 0.1$ and 0.8) for both sets of groups. Bland-Altman plots of the automatic measurements of CRT are presented in Figure 1.

Signal Strength

The mean signal strengths were 5.7 ± 1.3 and 5.6 ± 1.4 (range, 2–9) for the Cirrus and 18.3 ± 4.5 and 18.6 ± 5.2 (range, 7–28) for the Spectralis. The correlation of the signal strength between the first and second examinations was significantly positive for the Cirrus (Spearman $\rho = 0.47$, $P < 0.0001$) and the Spectralis (Spearman $\rho = 0.34$, $P = 0.0002$). The correlation between the signal strength of Cirrus and Spectralis was not significant ($P = 0.8$). The incidence of scans with low signal strength was 18.8% and 15.2% for Cirrus and 14.3% and 18.8% for Spectralis, for the first and second examinations, respectively, including 6.3% of eyes exhibiting low-quality scans in all the examinations. For Cirrus, a significantly positive correlation between the absolute difference of measurements and the signal strength was observed ($P = 0.0002$); for Spectralis, the correlation did not reach significance ($P = 0.06$).

Table 1. Repeatability and Reproducibility of Central Retinal Thickness Measurement

Random Group OCT Device	Mean Difference (95% CI)	Within SD	Between SD	ICC (95% CI)	*n*
Automatic measurements					
1–4					
Cirrus	6.089 (−11.900 to 24.078)	47.269	60.123	0.618 (0.427 to 0.756)	56
Spectralis	−11.038 (−20.642 to −1.435)	25.404	97.844	0.934 (0.888 to 0.961)	52
5–8					
Cirrus	5.839 (−13.816 to 25.494)	51.598	70.780	0.653 (0.474 to 0.78)	56
Spectralis	9.778 (−7.754 to 27.310)	41.206	97.436	0.861 (0.762 to 0.921)	45
Segmentation error correction					
1–4					
Cirrus	4.357 (−2.813 to 11.527)	19.014	87.283	0.955 (0.924 to 0.973)	56
Spectralis P	0.075 (−1.522 to 1.673)	4.061	93.285	0.998 (0.997 to 0.999)	53
Spectralis BM	−5.245 (−13.536 to 3.046)	21.429	107.307	0.953 (0.921 to 0.973)	52
5–8					
Cirrus	3.518 (−3.977 to 11.013)	19.769	99.006	0.962 (0.936 to 0.977)	56
Spectralis P	−0.036 (−2.487 to 2.414)	6.352	92.680	0.995 (0.992 to 0.997)	55
Spectralis BM	6.784 (−6.496 to 20.065)	33.414	120.786	0.928 (0.878 to 0.958)	53
Additional correction of centralization					
1–4					
Cirrus	0.607 (−1.498 to 2.713)	5.526	93.056	0.996 (0.994 to 0.998)	56
Spectralis	−0.389 (−2.018 to 1.241)	4.191	87.468	0.998 (0.996 to 0.999)	54
5–8					
Cirrus	3.5 (−0.184 to 7.184)	9.952	100.914	0.99 (0.984 to 0.994)	56
Spectralis	1.589 (−0.644 to 3.822)	5.950	105.293	0.997 (0.995 to 0.998)	56
Low-quality scans (signal strength <5 in Cirrus, <15 dB in Spectralis) excluded					
1–4					
Cirrus	−6.15 (−19.72 to 7.42)	32.26	66.72	0.81 (0.683 to 0.89)	46
Spectralis P	−7.13 (−15.33 to 1.08)	18.38	104.63	0.968 (0.94 to 0.983)	39
5–8					
Cirrus	2.79 (−8.69 to 14.28)	24.81	70.59	0.89 (0.801 to 0.941)	39
Spectralis P	11.05 (−10.20 to 32.30)	45.04	89.84	0.806 (0.656 to 0.895)	37

Data are the mean difference, within and between standard deviation and ICC of Cirrus and Spectralis for random groups 1–4 and 5–8, respectively. The posterior line was positioned at the RPE with Spectralis P and at Bruch's membrane with Spectralis BM.

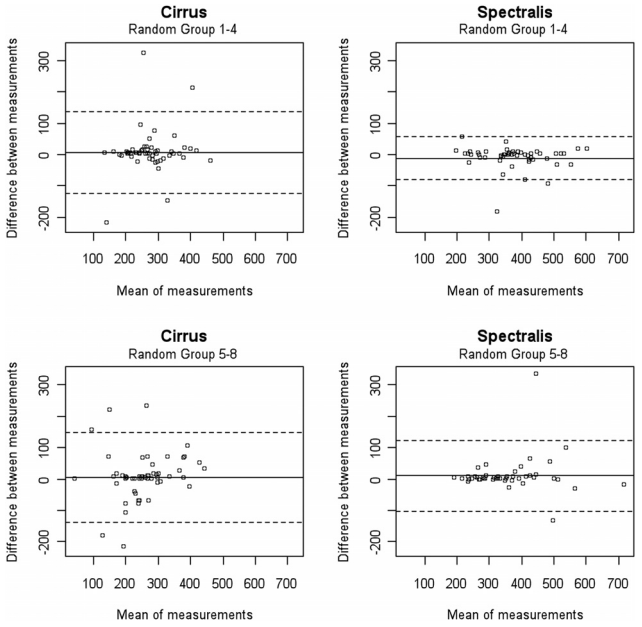

FIGURE 1. Bland-Altman plots of the automatic measurements of CRT with the OCT systems Cirrus and Spectralis and for groups 1 to 4 and groups 5 to 8 are presented separately.

Segmentation Errors

In the Cirrus scans, segmentation errors were found by the first examiner (IK) in 55.4% and by the second examiner (ES) in 50.9%. In the Spectralis scans, errors were detected significantly more frequently: by the first examiner in 75% and by the second examiner in 72.3% ($P < 0.0001$). There were clinically nonsignificant discrepancies in the edges of the scan square or the area near the optic disc. The mean error scores per line with Cirrus were 0.53 and 0.52 and with Spectralis were 0.83 and 0.98 in the first and second examinations, respectively. Besides the OCT system, distance acuity ($P = 0.002$) and age ($P = 0.05$) also revealed significance in the logistic regression, with error score as the dependent and the other parameter as the independent variable. The examiner performing the evaluation had no significant influence.

The central 1-mm-diameter area used for CRT calculation was free of segmentation errors in 67.9% and 74.1% of Cirrus scans and in 42.0% and 33.0% of Spectralis scans in the first and second examinations, respectively. Stage 1 failures were seen in 10.7% and 8.0% of Cirrus scans and in 25.9% and 31.3% of Spectralis scans and stage 2 failures in 6.3% in both Cirrus examinations and in 11.6% and 9.8% of Spectralis scans in the first and second examinations, respectively. In Cirrus scans, stage 3 failures occurred in 15.2% and 11.6% (6.3% in both examinations) and in Spectralis scans in 20.5% and 25.9% (14.3% in both examinations) in the first and second examinations, respectively. A mixed model with patient as the random effect and with the logarithm of the absolute differences between the two measurements of Cirrus and Spectralis as the dependent variable (which quantifies the variability of the two measurements) revealed a significant effect of the central error (estimate [95% CI]: 0.785 [0.572–0.999]; $P < 0.0001$), of the OCT system*central error interaction (estimate [95% CI]: −0.404 [−0.732 to −0.077]; $P = 0.02$), but not of the OCT system ($P = 0.4$). This result indicates that the variability of measurements of CRT increases with central error score with both systems, with a significantly more pronounced increase for Cirrus than for Spectralis (Fig. 2). The incidence of different types of failures for both OCT machines is listed in Table 2. Box plots of the logarithm of the absolute difference between measurements for Cirrus and Spectralis are shown in Figure 2. Examples of different errors are presented in Figure 3.

Error Correction

Correction of the erroneous threshold algorithm lines in both examinations by the examiner (IK), using the built-in software, improved the ICC of Cirrus to 0.96 in both groups and additional correction of the location of decentralized scans improved the ICC to 1.0 and 0.99 in groups 1–4 and 5–8. For Spectralis, the ICC after error correction (posterior line at the RPE) was 1.0 in both groups before and after additional correction of the centralization (Table 1). All Cirrus examinations were correctable. Fifty-three Spectralis examinations in groups 1 to 4 and 55 in groups 5 to 8 were correctable. The differences between automated and corrected values in the two sets of groups for Cirrus were −0.27 ± 0.91 μm (range, −756 to 66) and −30.9 ± 100.2 μm (range, −763 to 100 μm), respectively, and for Spectralis were 35.8 ± 80.6 mm (range, −124 to 377) and 36.4 ± 84.1 μm (−179 to 531), respectively,. The mixed model revealed a significant influence of only the OCT system on the corrected measurements ($P < 0.0001$), similar to the automatic measurements. The reproducibility of error correction is presented in Table 3.

The quantiles of the absolute difference between the corrected measurements of CRT revealed improved results and are

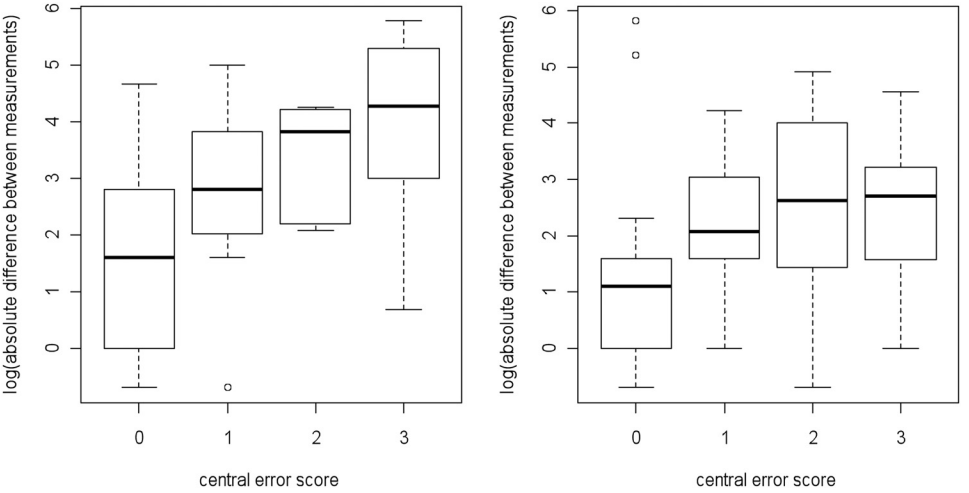

FIGURE 2. Boxplots of the logarithm of the absolute difference between measurements for Cirrus (*left*) and Spectralis (*right*) are presented separately. The *x*-axis shows the maximum central error of the two measurements of Cirrus or Spectralis, respectively (0, no error; 1, one third or less of CRT; 2, more than one third but less than or equal to two thirds of CRT; and stage 3, more than two thirds of CRT).

TABLE 2. Incidence of Failures

	Cirrus		Spectralis	
	Measurement1	**Measurement 2**	**Measurement**	**Measurement 2**
Segmentation Failures				
Inner border-line failure	33 (29.5)	27 (24.1)	6 (5.4)	6 (5.4)
Outer border-line failure	22 (19.6)	20 (17.9)	51 (45.5)	58 (51.8)
Degraded scan	7 (6.2)	8 (7.1)	2 (1.8)	0 (0)
Scan off screen	5 (4.5)	3 (2.7)	15 (13.4)	13 (11.6)
Scan edges off screen	1 (0.9)	2 (1.8)	2 (1.8)	3 (2.7)
No line draw	10 (8.9)	10 (8.9)	40 (35.7)	45 (40.2)
Line incompletely	13 (11.6)	4 (3.6)	20 (17.9)	20 (17.9)
Scan inverted	0 (0)	0 (0)	1 (0.9)	7 (6.2)
Sum	91	74	137	152
Motion Artifacts (with or without Segmentation Errors)				
Blinking	13 (11.6)	12 (10.7)	0 (0)	0 (0)
Saccades	48 (42.9)	56 (50.0)	0 (0)	0 (0)
Flipped scans	0 (0)	1 (0.9)	4 (3.6)	1 (0.9)

Data are the absolute frequency (percentage) of different segmentation errors and motion artifacts (regardless of whether they caused segmentation errors) for the two measurements of Cirrus and Spectralis, respectively. Multiple entries are possible.

presented in Table 4. Bland-Altman plots of the corrected measurements of CRT are presented in Figure 4.

Cirrus versus Spectralis CRT

The mean values of Cirrus automatic CRT were 271 ± 87.3 μm (range, 34–514) and 265.1 ± 77.39 μm (range, 20–474), in the first and second examinations, respectively. The mean Spectralis values were 370.9 ± 105.4 μm (range, 195–708) and 369.3 ± 102.2 μm (range, 193–729), respectively. In 84.4% of the automatically recorded Spectralis examinations with RPE detachment, the posterior line was placed at Bruch's membrane, in 44.4% of those there were line failures. When the signal of Bruch's was weak, the posterior line followed the RPE (15.6%), but always with failures. In 92.8% of the scans, the foveal finder in Cirrus failed to detect the fovea; in 6.3%, the fovea was identified correctly; and in 0.9%, the peak of an RPE detachment was misinterpreted as the fovea. Therefore, manual correction of centralization was required. Corrected segmentation and location failures resulted in 296.4 ± 98.5 and 292.5 ± 93.7 μm, respectively, with the Cirrus. For Spectralis, 394.3 ± 115.4 and 394.0 ± 114.0 μm, respectively, were calculated when the posterior line was corrected at Bruch's membrane and 345.2 ± 92.3 and 344.7 ± 94.6 μm, when corrected at the RPE. The mean difference of the corrected values between Spectralis (posterior line at the RPE) and Cirrus were 36.4 ± 27.2 and 38.0 ± 27.7 μm, for the first and second examinations, respectively. The linear model revealed a significantly positive correlation between the mean of the two repeated measurements of the corrected CRT of the Cirrus and Spectralis ($P < 0.0001$); the following conversion formula was used [95% CI]: CRT Cirrus = −25.760 [−43.942 to −7.578] + 0.966 [0.913–1.018] × CRT Spectralis ($df = 108$).

DISCUSSION

SD-OCT technology has resulted in a series of differences (improvements) in comparison to TD-OCT technology. A higher resolution as well as a faster acquisition regimen has been implemented. A different positioning of the posterior line in a more posterior position within the hyperreflectivity of the outer retina (different threshold algorithms) has also been implemented. With Stratus, good reproducibility was also found in neovascular AMD cases, in which reproducibility might be influenced by the pathology itself. For CRT, an ICC of 0.84 acquired by the fast macular thickness program and of 0.91 by the retinal thickness program for reproducibility and repeatability of 0.72 and 0.90 was found, respectively.[13] However, with respect to SD technology, new studies are needed to evaluate accuracy and reproducibility of retinal thickness measurements.

For normal eyes Leung et al.[8] reported an ICC of 0.92 for CRT with the 3D OCT by Topcon (Tokyo, Japan) compared to 0.88 for CRT with Stratus. Wolf-Schnurrbusch et al.[9] and Pierro et al.[14] compared the reproducibility of different SD machines in normal eyes, both presenting the best results for Spectralis followed by Cirrus (ICC: 0.97 and 0.93, respectively). For eyes with neovascular AMD in the present study, we found, for automatically measured CRT and unadjusted scans, an ICC of 0.61 for repeatability and 0.65 for reproducibility for Cirrus and 0.93 and 0.83 for Spectralis, respectively. Differences between repeatability and reproducibility were caused by a larger number of advanced cases in the reproducibility groups and were expressed also by a better distance acuity in the repeatability groups. The variability of the two repeated CRT measurements correlated with the quality of segmentation in the central 1-mm area required for CRT, whereas there was no influence of the sequence of the examinations or the examiners. A main cause of the differences in reproducibility between Cirrus and Spectralis is that the Spectralis software does not create maps in cases of low scan quality. Therefore (mainly patient related) low scan quality influenced reproducibility only in Cirrus. Establishing the same conditions by excluding low-quality scans in both systems equalized repeatability and reproducibility to 0.81 and 0.89 for Cirrus and 0.97 and 0.81 for Spectralis. Similar results were reported by Parravano et al.,[15] who found a decrease in the coefficient of variation from 5.37 for CRT in Cirrus to 4.15 when examinations with severe failures were excluded. For manually corrected scans, both systems offer software for manual error correction, which could be applied with a high degree of reproducibility. Ho et al.[16] reported an ICC of 0.92 for manually corrected values in Cirrus concerning reproducibility in eyes with various macular diseases, which is

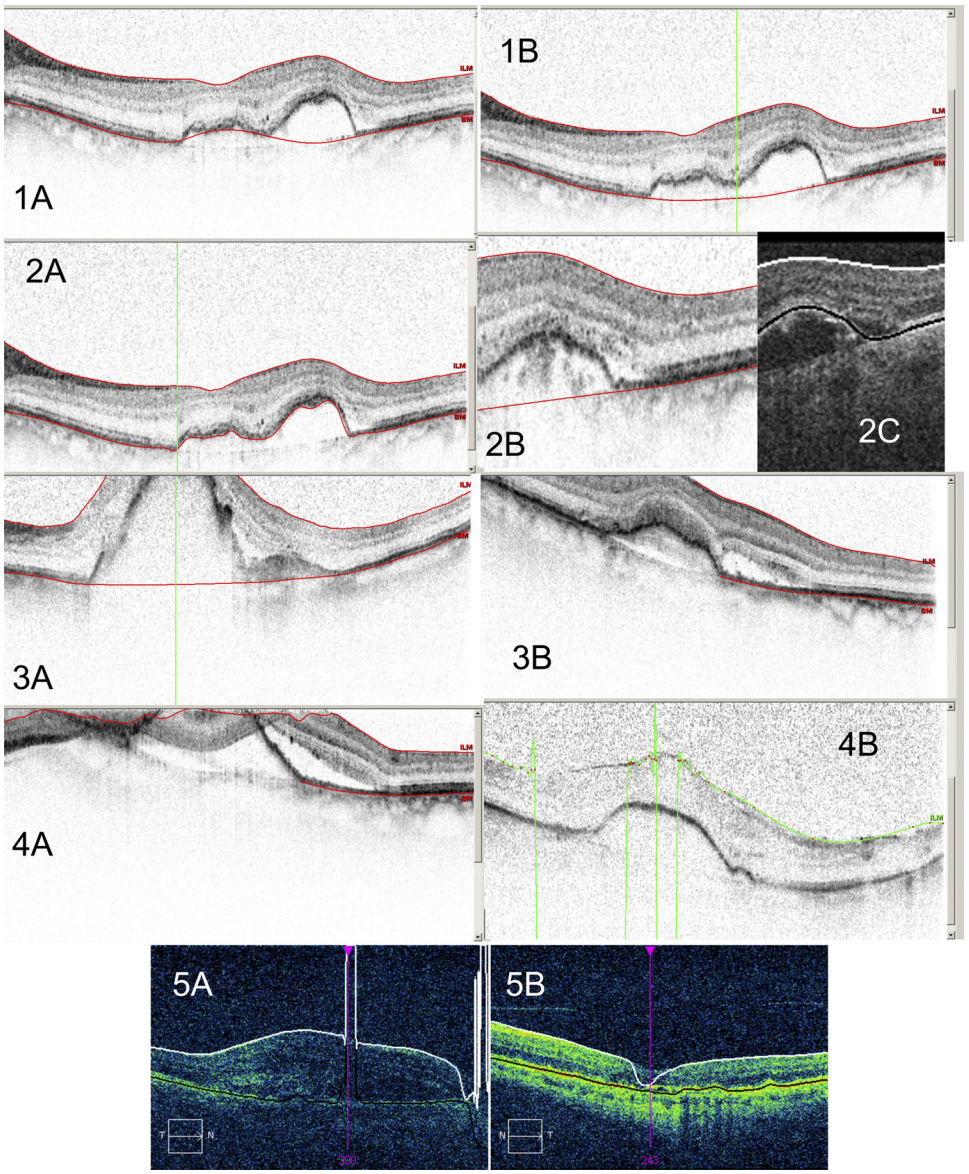

FIGURE 3. (**1A–2C**) Scans of patient 31, a woman 67 years of age with detachment of the retinal pigment epithelium (RPE). (**1A**) Automatic segmentation with outer borderline failure, posterior line partly at the RPE and partly at Bruch's membrane; (**1B**) manual correction of the posterior line to coincide with Bruch's membrane; (**2A**) manual correction to the RPE. (**2B**) Spectralis and (**2C**) Cirrus scans are magnifications of the beginning of the RPE detachment, showing the more posterior position of the posterior line in the Spectralis scan. (**3A**) Off-screen error; (**3B**) the cutoff edge, (**4A**) an inverted scan; and (**4B**) an incomplete line. (The *red* points that had to be moved to correct the error were off-screen, and the error was thus not correctable.) (**3A–4B**) Spectralis images. (**5A**, **5B**) Cirrus images showing (**5A**) an error-degraded scan (i.e., low signal strength) and (**5B**) an inner border line failure.

comparable to our results (0.96 for Cirrus and 1.0 for Spectralis). In conducting studies on AMD, it is of interest to know which changes in CRT are due to the course of the disease and which are due only to the inaccuracy of the measurement (for Stratus, a value of 100 μm was estimated). Calculation of the quantiles of the absolute differences exhibited a 95% chance that differences of more than 20 μm are not related to test–retest variability. Also Parravano et al.[15] concluded that it is safe to assume that changes of more than 26 μm in Cirrus are pathologic.

Although Cirrus and Spectralis both use SD technology, there are notable differences that should be pointed out, affecting control of scan localization, number of scans, position of the posterior reference line, and number and nature of errors.

TABLE 3. Reproducibility of Corrected Measurements

Random Group	Mean Difference (95% CI)	Within SD	Between SD	ICC	100% Agreement	*n*
Cirrus						
1–4	−0.23 (−0.81 to 0.35)	1.53	92.998	1.00	67.86	56
5–8	1 (0.24 to 1.76)	2.12	101.59	1.00	64.29	56
Spectralis						
1–4	−0.426 (−2.02 to 1.16)	4.09	105.14	0.998	57.41	54
5–8	−0.15 (−1.16 to −0.87)	2.63	125.62	1.00	52.73	55

Data are the mean difference between the two repeated measurements of the first examination, the within and between SD, and the ICC. As the difference between the two measurements was 0 in many cases, the ICC may not be valid and thus, the percentage of identical measurements is also shown.

TABLE 4. Quantiles of the Absolute Difference between the CRT Measurements

Random Group OCT	0.75	0.80	0.85	0.9	0.95
Automatic measurements					
1–4					
Cirrus	23.5	25	48	66	163.75
Spectralis	18	23.8	34.7	41.7	74.75
5–8					
Cirrus	69.25	70	77.75	108	192.25
Spectralis	21	30.4	34.8	48.4	89.4
1–8					
Cirrus	48	66.8	70	92.7	196.15
Spectralis	18	26.4	35.2	46.6	85.4
Corrected measurements					
1–4					
Cirrus	7	8	9.75	12.5	17.75
Spectralis	6	6.6	7.2	11.4	31.4
5–8					
Cirrus	6.25	8	8.75	11.5	22
Spectralis	6.5	7	7.5	11	17.5
1–8					
Cirrus	7	8	9.35	12	18.35
Spectralis	6	7	7.55	11	18.70

The 0.75, 0.80, 0.85, 0.9, and 0.95 quantiles of the absolute difference of measurements of CRT in groups 1–4 and 5–8 and for all patients, for automatic and corrected measurements, respectively.

Control of Scan Localization

In the Spectralis system, a second light beam (the eye tracker) records an image of the retina and, based on maps of more than 1000 points, controls the correct position of the scan despite eye movements. In repeated examinations, the follow-up scan is recorded only when positioned at the same location. Whereas with the Cirrus, 128 raster scans are recorded within seconds and the examiner cannot influence the process after the scan acquisition is started, the operator of the Spectralis has to control and correct the position and quality of the scan with the joy stick throughout the examination. Particularly in elderly and low-vision patients who cannot avoid movements of their eyes and the head, it takes a long time before a successful scan can be recorded. Therefore, it was not possible to perform examinations with a comparable density of scans before the eye tracker timed out at 5 minutes. In the original study protocol, volume scans containing 121 lines were planned. Because the volume scans could be completed in only 4 of 10 patients, the randomization was stopped and restarted with a lower scan density of 32 scans 240 μm apart. This scan density is the most frequently applied in clinical practice and was justified by the results of a study by Sadda et al.[17] on the Cirrus system. They reported that a density of 32 horizontal B-scans (spaced 188 μm apart) causes only minimal change in calculated CRT. However, for CRT measurement, there were only 5 lines used in the Spectralis (in the Stratus, 6 radial lines), whereas in the Cirrus there were 22 lines. Therefore, more interpolation is necessary in Spectralis and Stratus scans than in Cirrus scans. The comparison of reproducibility data, the error scores, and the CRT values could be influenced by this difference in line density.

The eye tracker in Spectralis not only controls the scan position within one volume scan but also verifies that, in follow-up examinations, the scans are positioned at identical locations, providing a high degree of reproducibility. In Cirrus, different strategies are applied to control centralization. The foveal finder automatically sets the center of the ETDRS scheme to the fovea. Unfortunately, in exudative AMD, the fovea cannot be detected by the system. A more important tool is the postprocessing correction of the localization. Landmarks can be set on the video image at anatomically distinct points in two examinations on different days; thereafter, the images and the ETDRS schemes are moved to an identical position. This very useful tool could not be tested in this study because it is not available for examinations conducted on the same day. Therefore, the ETDRS scheme was moved manually to the fovea by the examiner in both examinations. This adjustment improved the Cirrus ICC to 1.0. Manually moving the ETDRS scheme to the foveal region is also possible with the Spectralis (simultaneously in both examinations and therefore without effect on reproducibility) and was applied to obtain CRTs comparable to those of the Cirrus. After correction of segmentation errors and centralization, a mean difference between Cirrus and Spectralis of 36.4 and 38.0 μm for both examinations was found, respectively, and a conversion formula was calculated, similar to the formula we have calculated to convert Stratus to Cirrus CRT.[18]

Position of the Posterior Reference Line

Both machines position the anterior line at the ILM, the posterior boundary line in the area of the third hyperreflective band of the outer retina, resulting in higher CRT values than Stratus TD and other SD systems. In the Spectralis system, this line is positioned even more posteriorly at the posterior surface of the line, corresponding to Bruch's membrane, than in Cirrus, where the posterior line is positioned at the anterior surface of this line. Therefore, Spectralis delivers higher values in normal eyes (278 μm Cirrus vs. 288 μm Spectralis, by Wolf-Schnurrbusch et al.,[9]; 253.9 vs. 273.2 by Pierro et al.[14]). More impressive are the differences in neovascular AMD, where the third hyperreflective line is frequently thickened (Mylonas et al.,[19] 327 μm, vs. Han et al.[12], 383 μm). A consequence of these different algorithms is that Cirrus takes into account the RPE detachment and Spectralis does not, as was described in a pilot study.[10] Although in most of the eyes, the localization of the posterior line by Spectralis was at Bruch's membrane, segmentation errors were frequent in cases of RPE detachment, because the automatically set line was positioned somewhere between the RPE and Bruch's membrane. Automatically set lines at the detached RPE were seen only in cases

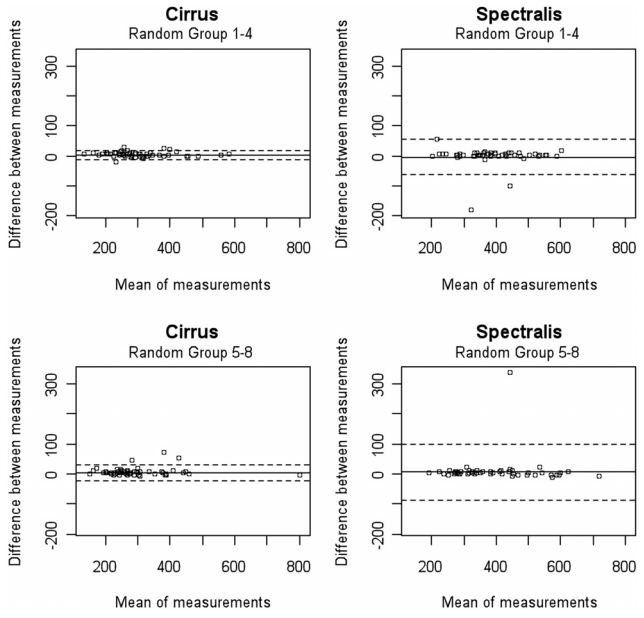

FIGURE 4. Bland-Altman plots of the corrected measurements of CRT, for the Cirrus and Spectralis OCT systems and for groups 1 to 4 and groups 5 to 8, are shown separately.

of very high detachments and a resulting weak signal of Bruch's membrane. Commonly, RPE detachments are not taken into account in clinical practice when the activity of a lesion is determined or the effect of a treatment is documented. Therefore, the posterior line was set manually to the posterior surface of the RPE line in these cases. Furthermore, the position of the Spectralis posterior line was corrected to a more anterior position in fibrotic lesions, where discrepancies were also seen between Cirrus and Spectralis.

Segmentation Errors

The number of threshold algorithm line errors was significantly higher with Spectralis. Overall, segmentation errors were frequent (in 55.4% of Cirrus and 75% of Spectralis scans) and exceeded the value found in Cirrus in a prior study (32.7%),[20] where only eyes with exudative AMD were also included. When counting only failures involving the central 1 mm needed for CRT calculation, we found 32.1% and 25.1% of Cirrus and 58% and 67% of Spectralis scans contained failures in the first and second examinations, respectively; 21.4% and 17.9% for Cirrus and 32.1 and 35.7% for Spectralis accounted for at least one third of retinal thickness and were therefore of clinical relevance. The results of this grading could have been influenced by the higher values of CRT and the lower scan density of Spectralis, and therefore the grading system introduced by Sadda et al.[11] was also applied. This grading calculates error scores/line. However, similar results were found (0.53 and 0.52 for Cirrus and 0.83 and 0.98 for Spectralis, in the first and second examinations). These values exceeded those of Mylonas et al.[19] (errors >50 μm in length and >0.5 mm in width were clinically relevant), who found clinical relevant failures in 6% of Cirrus and 27% of Spectralis scans of 26 AMD patients. Although the failures were more frequent and severe with Spectralis, they did not influence reproducibility to the same extent. In 13.4% of Spectralis scans, no map and therefore no CRT value influencing reproducibility was available in at least one examination. Furthermore, the eye tracker line positioning in the same disease at the same location led to the same segmentation error in both examinations.

There were also differences concerning the distribution of errors between the two SD-OCT systems in the present study. Because of the necessary manual adjustments by the examiner throughout the longer lasting Spectralis examination, it is by far more difficult to keep a scan within the screen. This results in failures such as inverted scans, cut off edges, and scans off-screen,[21] failures not frequently seen with the Cirrus. Furthermore, there are scans of lower signal strength within the Spectralis volume scans, resulting in incomplete lines or missing lines. In these cases, frequently no maps are provided by the Spectralis software.

In contrast, motion artifacts are by far less frequent in Spectralis scans because of the eye tracker. These artifacts were frequent in Cirrus scans, but did not influence CRT measurements in most of the cases (saccades). There was no case of blinking (meanwhile no scans are acquired) involving the central region and only one case with larger eye movements (resulting in repeated acquisition of the same section), which would be relevant to measurement. Failures of the localization of the posterior reference line at Bruch's membrane, which makes it more difficult to detect the outer boundary line, were more frequent in Spectralis scans, whereas the inner line was involved more frequently in Cirrus scans. This result was in accordance with findings in another study.[12]

The limiting factor of the study design was that repetitions of low-quality scans were not allowed, although such repeat testing is done in clinical practice with Cirrus. These repetitions are easily made with Cirrus, within seconds, but are very strenuous for elderly patients tested with Spectralis. Therefore, we feared a possible bias caused by the fatigue of the patients. The sequence of examinations was randomized, to exclude a possible bias caused by decreasing concentration after repeated examination. However, according to the results of a prior study with Stratus, we had to expect patient-related low-quality scans in cases of cataract, narrow pupil, or corneal or vitreal opacities in approximately 7%.[22] Including a consecutive series of patients not selected resulted in 6% of the patients with low-quality scans in each of the four examinations. Furthermore, the signal strength of the repeated examinations correlated significantly in both systems, which may indicate a relationship between signal strength and patient-related conditions.

In summary, repeatability and reproducibility were influenced in both OCT systems by algorithm line errors and, in the Cirrus system, by variability of centralization, as well. Manual correction of these failures, excluding scans with severe central failures or those with low signal strength, improved reproducibility and repeatability.

References

1. Fung AE, Lalwani GA, Rosenfeld PJ, et al. An optical coherence tomography-guided, variable dosing regimen with intravitreal ranibizumab (Lucentis) for neovascular age-related macular degeneration. *Am J Ophthalmol.* 2007;143:566–583.
2. Krebs I, Ansari-Shahrezaei S, Goll A, Binder S. Activity of neovascular lesions treated with bevacizumab: comparison between optical coherence tomography and fluorescein angiography. *Graefes Arch Clin Exp Ophthalmol.* 2008;246:811–815.
3. Malamos P, Sacu S, Georgopoulos M, et al. Correlation of high-definition optical coherence tomography and fluorescein angiography imaging in neovascular macular degeneration. *Invest Ophthalmol Vis Sci.* 2009;50:4926–4933.
4. Chakravarthy U, Adamis AP, Cunningham ET, et al. VEGF Inhibition Study in Ocular Neovascularization (V.I.S.I.O.N.) Clinical Trial Group, Year 2 efficacy results of 2 randomized controlled clinical trials of pegaptanib for neovascular age-related macular degeneration. *Ophthalmology*. 2006;113:1508.e1–25.
5. Rosenfeld PJ, Brown DM, Heier JS, Boyer, et al.; MARINA Study Group. Ranibizumab for neovascular age-related macular degeneration. *N Engl J Med*. 2006;355:1419–1431.
6. Kaiser PK, Brown DM, Zhang K, et al. Ranibizumab for predominantly classic neovascular age-related macular degeneration: subgroup analysis of first-year ANCHOR results. *Am J Ophthalmol.* 2007;144:850–857.
7. Hagen S, Krebs I, Haas P, et al. Reproducibility and comparison of retinal thickness and volume measurements in normal eyes determined with two different Cirrus OCT scanning protocols. *Retina*. 2011;31:41–47.
8. Leung CK, Cheung CY, Weinreb RN, et al. Comparison of macular thickness measurements between time domain and spectral domain optical coherence tomography. *Invest Ophthalmol Vis Sci.* 2008;49:4893–4897.
9. Wolf-Schnurrbusch UE, Ceklic L, Brinkmann CK, et al. Macular thickness measurements in healthy eyes using six different optical coherence tomography instruments. *Invest Ophthalmol Vis Sci.* 2009;50:3432–3437.
10. Smretschnig E, Krebs I, Moussa S, Ansari-Shahrezaei S, Binder S. Cirrus OCT versus Spectralis OCT: differences in segmentation in fibrovascular pigment epithelial detachment. *Graefes Arch Clin Exp Ophthalmol.* 2010;248:1693–1698.
11. Sadda SR, Wu Z, Walsh AC, et al. Errors in retinal thickness measurements obtained by optical coherence tomography. *Ophthalmology.* 2006;113:285–293.
12. Han IC, Jaffe GJ. Evaluation of artifacts associated with macular spectral-domain optical coherence tomography. *Ophthalmology.* 2010;117(6):1177–1189.

13. Krebs I, Hagen S, Brannath W, et al. Repeatability and reproducibility of retinal thickness measurements by optical coherence tomography in age-related macular degeneration. *Ophthalmology.* 2010;117(8):1577–1584.
14. Pierro L, Giatsidis SM, Mantovani E, Gagliardi M. Macular thickness interoperator and intraoperator reproducibility in healthy eyes using 7 optical coherence tomography instruments. *Am J Ophthalmol.* 2010;150:199–204.
15. Parravano M, Oddone F, Boccassini B, et al. Reproducibility of macular thickness measurements using Cirrus spectral domain optical coherence tomography in neovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci.* 2010;51(9):4788–4791.
16. Ho J, Sull AC, Vuong LN, et al. Assessment of artifacts and reproducibility across spectral- and time-domain optical coherence tomography systems. *Ophthalmology.* 2009;116:1960–1970.
17. Sadda SR, Keane PA, Ouyang Y, Updike JF, Walsh AC. Impact of scanning density on measurements from spectral domain optical coherence tomography. *Invest Ophthalmol Vis Sci.* 2010;51:1071–1078.
18. Krebs I, Falkner-Radler C, Hagen S, et al. Quality of the threshold algorithm in age-related macular degeneration: Stratus versus Cirrus OCT. *Invest Ophthalmol Vis Sci.* 2009;50:995–1000.
19. Mylonas G, Ahlers C, Malamos P, et al. Comparison of retinal thickness measurements and segmentation performance of four different spectral and time domain OCT systems in neovascular age-related macular degeneration. *Br J Ophthalmol.* 2009;93(11):1453–1460.
20. Krebs I, Haas P, Zeiler F, Binder S. Optical coherence tomography: limits of the retinal-mapping program in age-related macular degeneration. *Br J Ophthalmol.* 2008;92(7):933–935.
21. Ho J, Castro DP, Castro LC, et al. Clinical assessment of mirror artifacts in spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci.* 2010;51:3714–3720.
22. Krebs I, Hagen S, Smretschnig E, Womastek I, Brannath W, Binder S. Conversion of Stratus optical coherence tomography (OCT) retinal thickness to Cirrus OCT values in age-related macular degeneration. *Br J Ophthalmol.* Published online February 24, 2011.