RESEARCH ARTICLE

Reliability and reproducibility of spectral and time domain optical coherence tomography images before and after correction for patients with age-related macular degeneration [version 1; peer review: 1 approved, 1 approved with reservations]

Aymen Rashid¹*, Yasir J Sepah¹,²*, Roomasa Channa¹, Elham Hatef¹, Matthew Shulman³, Diana V Do², Quan Dong Nguyen²

¹The Wilmer Eye Institute, Johns Hopkins University, Baltimore, Maryland, 21287, USA
²Stanley M. Truhlsen Eye Institute, University of Nebraska Medical Center, Omaha NE, 68105, USA
³University of Maryland School of Medicine, Baltimore ND, 21205, USA

* Equal contributors

First published: 23 May 2013, v1
https://doi.org/10.12688/f1000research.2-131.v1
Latest published: 05 Mar 2015, v2
https://doi.org/10.12688/f1000research.2-131.v2

Abstract

Purpose: To evaluate the reproducibility and reliability of Optical Coherence Tomography scans (OCT) obtained using the Time Domain (TD-OCT) Stratus™ OCT, and the Spectral Domain (SD-OCT) Spectralis™ and Cirrus™ OCT devices before and after manual correction in eyes with either Neovascular (NV-AMD) or Non-Neovascular (NNV-AMD) Age-related Macular Degeneration.

Methods: We conducted a prospective observational study of 36 patients (50 eyes) with NV-AMD or NNV-AMD at a university-based retina practice. OCT scans were taken simultaneously using one TD-OCT and two SD-OCT devices. Macular thickness measurements were assessed before and after correction of the OCT algorithm by constructing Bland-Altman plots for agreement and calculating intra-class correlation coefficients (ICCs) and coefficients of repeatability (COR) to evaluate intra-class repeatability.

Results: The Spectralis device had the highest number of images needing manual correction. All machines had high ICCs, with Spectralis having the highest. Bland-Altman plots indicated that there was low agreement between both Cirrus™ and Stratus™ and Spectralis™ and Stratus™, while there was good agreement between the Cirrus™ and Spectralis™ devices. The CORs were lowest for Spectralis™ and similar with each other and had higher values for Cirrus™ and Stratus™. Agreement, CORs, and ICCs generally
improved after manual correction, but only minimally.

**Conclusion:** Agreement is low between devices, except between both SD-OCT machines. Manual correction tends to improve results.

**Keywords**
age related macular degeneration, neovascularization, optical coherence tomography, spectral domain, time domain

This article is included in the Eye Health gateway.

**Corresponding author:** Quan Dong Nguyen (qnguyen4@jhmi.edu)

**Competing interests:** No competing interests were disclosed.

**Grant information:** The author(s) declared that no grants were involved in supporting this work.

**Copyright:** © 2013 Rashid A et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

**How to cite this article:** Rashid A, Sepah YJ, Channa R et al. Reliability and reproducibility of spectral and time domain optical coherence tomography images before and after correction for patients with age-related macular degeneration [version 1; peer review: 1 approved, 1 approved with reservations] F1000Research 2013, 2:131 https://doi.org/10.12688/f1000research.2-131.v1

**First published:** 23 May 2013, 2:131 https://doi.org/10.12688/f1000research.2-131.v1
**Introduction**

Optical Coherence Tomography (OCT) is a non-invasive imaging modality that allows acquisition of cross-sectional images of the retina. OCT is useful in monitoring and evaluating retinal thickness in many retinal disorders. One example is Age-related Macular Degeneration (AMD), a progressive, blinding disease that is mostly Non-Neovascular (NNV-AMD) but can be associated with choroidal Neovascularization (NV-AMD). Currently, OCT is also being employed as an outcome measure in many multicenter clinical trials of AMD with Time Domain OCT (TD-OCT) devices being the most common. Spectral Domain OCT (SD-OCT) is a newer technology that obtains high-resolution scans of the retina.

As this technology is increasingly being utilized by many ophthalmologists to evaluate and monitor patients and guide treatment decisions, it is important to understand the reliability and accuracy of thickness measurements obtained with the various devices currently available. Recently, studies have shown that in patients with AMD, there is a high frequency of errors in automated retinal thickness measurements due to incorrect segmentation of the retina in the TD-OCT machine, specifically in NV-AMD. Using a Spectral Domain OCT (SD-OCT) device Menke et al. found that retinal thickness measurements in NNV-AMD cases had fewer errors than in NV-AMD cases, mostly due to the pathology of the former disease resulting in retinal pigment epithelial (RPE) layer changes.

Manual correction of the OCT algorithm is an option in newer generations of the OCT review software and as more devices are coming to the market, it is important to understand the reliability and accuracy of thickness measurements from different devices before and after correction. To date, no other study has examined the effects of manual correction of the thickness algorithm in SD-OCT and TD-OCT machines in eyes with AMD. In our study, we evaluated the intra-session repeatability and agreement in retinal thickness measurements for patients with NV-AMD and NNV-AMD before and after manual correction using three different OCT devices: Stratus™-TD-OCT and two SD-OCTs, Spectralis™ and Cirrus™.

**Methods**

Institutional Review Board (IRB)/Ethics Committee approval was obtained and HIPAA guidelines were followed for the study. Written Informed consent was obtained from study subjects.

**Patients and scanning**

Patients with a confirmed diagnosis of AMD were enrolled in the study. Two senior retina specialists (QDN and DVD) made the diagnosis of AMD. Patients under treatment with intra-vitreal injections of anti-Vascular Endothelial Growth Factor (VEGF) agents were also allowed to participate in the study. Patients were scanned twice by certified OCT operators on a TD-OCT device (Stratus™ OCT) and two SD-OCT devices (Spectralis™ and Cirrus™ OCT) machines in random order and with 5–10 minutes between each device. The same operator performed all the scans on any given patient. Scans on a single device were performed consecutively and 5 minutes apart from each other.

**Optical coherence tomography**

One TD-OCT machine, Stratus™ (software version 4), and two SD-OCT machines, Spectralis™ (software version 5.01) and Cirrus™ (software version 5.0.0.326) were used. Stratus™ is a TD-OCT machine that uses a super luminescent diode with a wavelength of 820 nm. It provides an axial resolution of 10 µm and image acquisition speed of 400 A-scans/second. Using the Spectralis™, two Fast Macular Thickness Maps (FMTM) were acquired from each eye. The FMTM is created through acquiring six radial B-scans, each consisting of 512 A-scans, and at an angle of 30° from each other with the point of intersection centered on the fovea.

Spectralis™ uses a super luminescent diode with a wavelength of 870 nm. It provides axial resolution of 4 µm and image acquisition speeds of up to 40,000 A-scans/second. Two volume scans were acquired from each eye using a raster scan of 19 lines covering 20x15° of the fundus. Using the TruTrack™ functionality of the Spectralis™ OCT, each line was averaged 15 times or more. Cirrus™ HD-OCT also uses a super luminescent diode with a wavelength of 840 nm. It provides images with an axial resolution of 5 µm and acquisition speeds of 27,000 A-scans per second. We acquired two 512x128 macular cube scans (128 B-scans and 512 A-scans, covering a retinal area of 6.0x6.0 mm) from each eye.

**Error determination, manual correction, and exclusion of scans**

Scans from each of the three devices were reviewed at the Retinal Imaging Research and Reading Center at the Wilmer Eye Institute by independent graders. Incorrect identification of inner and outer retinal boundaries by automated algorithms in Spectralis™ and Cirrus™ devices was manually corrected. Stratus™ images could not be corrected due to the lack of editing capabilities in the operating system provided with the machine at the time the study was conducted. Only five patients required corrections and were excluded from the analysis. The proprietary software identifies retinal boundaries for measurement of retinal thickness that are specific to each device. Whereas each device identifies the inner limiting membrane (ILM) as the inner boundary of retina, identification of the outer boundary is different for each device. Stratus™ identifies the junction between the inner and outer segments of photoreceptors (IS/OS) as the outer boundary, Spectralis™ identifies the posterior border of the retinal pigment epithelium (RPE),
and Cirrus™ identifies the inner border of the RPE as the outer retinal boundary.

Whenever the foveal center could be identified, grids were repositioned for scans with off-center positioning of the Early Treatment Diabetic Retinopathy Study (ETDRS) grid. However, in some cases, morphological changes associated with the advanced disease made identification of the foveal center unreliable. Adjustment of grid position was not possible for Stratus™ OCT. Scans were excluded from analysis only if identification of retinal layers and determination of the retinal thickness was not possible. OCT scans from which extraction of thickness data for the central 1 mm sub-field was not reliable, due to missing data in the image or the scan being out of range, were also excluded from analysis.

The retinal thickness measurements of the nine standard ETDRS subfields (Figure S1 illustrates the nine-subfield abbreviations) were recorded from each device before and after correcting the errors in the scans’ algorithm.

**Statistical analysis**

Bland-Altman plots were constructed to determine agreement between devices; both 95% confidence intervals and limits of agreements were calculated. Reproducibility of measurements was determined by calculating the coefficients of repeatability (COR) for each machine. Intraclass correlation coefficients (ICCs) were used to determine the reproducibility for each device. The statistical significance of difference in thickness before and after correction of images across devices was determined via the student’s t-test with α = 0.05 with Bonferroni correction for multiple comparisons. STATA version 10 and Microsoft Excel 2007 were used for data management and analysis. The statistical analysis was performed before and after any manual corrections were made to the algorithm errors described above.

**Results**

Fifty eyes from 36 patients were included in the study; 29 eyes had NV-AMD and 21 eyes had NNV-AMD. The mean age of the study subjects was 76.6 years. Males had a mean of 76.3 with a range of 61 to 90 years. Females had a mean of 76.83 with a range of 53 to 90 years.

**Exclusion and corrections**

**Stratus™**

Scans from four eyes could not be recovered from the database and scans from three eyes had algorithm errors with incorrect identification of retinal boundaries and were excluded from analysis. Scans were not corrected for off-center positioning of the scan as moving the ETDRS grid was not possible with the available software version.

**Cirrus™**

Scans in six eyes scanned first and eight eyes scanned second were corrected either for off-center fixation of the eye or for incorrect automated identification of retinal boundaries. The thickness measurements before and after correction were not statistically significant (p<0.05) for any of the subfields and also when stratified by diagnosis.

**Spectralis™**

Thirty-three scans among the first set and 32 among the second set were corrected. The inner inferior subfield for NV-AMD was the only subfield that was statistically significant before and after correction. Figure 1 plots the frequency of the differences before and after correction for the central subfield for all scans. 77% of the differences were less than 48 µm and 50% were less than 10 µm.

**OCT characteristics**

The mean (+SD) of the macular thickness of all of the subfields, including the central 1 mm subfield (foveal thickness; FTH) for Stratus™, Cirrus™, and Spectralis™ devices before and after manual correction of scans, stratified by diagnosis of NV-AMD and NNV-AMD, is shown in Table 1. For NV-AMD, the FTH values for the central 1 mm were 375 µm (+129 µm), 253 µm (+74 µm), 312 µm (+110 µm) for Spectralis™, Stratus™, and Cirrus™ respectively. After correction, the values were 335 µm (+106 µm) for Spectralis™ and 318 µm (+110 µm) for Cirrus™. On the other hand, the FTH values for NNV-AND in the central 1 mm before correction were 298 µm (+87 µm), 193 µm (+32 µm), and 229 µm (+30 µm) for Spectralis™, Stratus™, and Cirrus™ respectively. Spectralis™ was the only device to have a different FTH value (248 µm +56 µm) after correction. Overall, Spectralis™ had...
A comparison of thickness measurements between three machines demonstrated that most values were significantly different \((p<.05)\). Spectralis™ vs. Cirrus™ before correction: for Neovascular Age-related Macular Degeneration (NV-AMD), T1, S1, and I2 \((p<.05)\); and for Non-Neovascular (NNV-AMD), C1, T1, N1, and I2. Spectralis™ vs. Cirrus™ after correction \((p<.05)\): for NV-AMD, every field except S2, and I2 were not significant; and for NNV-AMD, the inner subfields were not significant \((p>.05)\). Spectralis™ vs. Stratus™ after correction: for NV-AMD, C1 \((p<.05)\).

### Table 1

<table>
<thead>
<tr>
<th>Subfield</th>
<th>Mean ± standard deviation (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All eyes</td>
</tr>
<tr>
<td></td>
<td>NV-AMD</td>
</tr>
<tr>
<td></td>
<td>NNV-AMD</td>
</tr>
<tr>
<td></td>
<td>Spectralis™ Before</td>
</tr>
<tr>
<td></td>
<td>Stratus™ Before</td>
</tr>
<tr>
<td></td>
<td>Cirrus™ Before</td>
</tr>
<tr>
<td></td>
<td>Spectralis™ Before</td>
</tr>
<tr>
<td></td>
<td>Stratus™ Before</td>
</tr>
<tr>
<td></td>
<td>Cirrus™ Before</td>
</tr>
<tr>
<td>C1</td>
<td>343 ± 119 301 ± 98 229 ± 67</td>
</tr>
<tr>
<td>N1</td>
<td>348 ± 74 329 ± 70 267 ± 50</td>
</tr>
<tr>
<td>S1</td>
<td>346 ± 74 327 ± 66 260 ± 40</td>
</tr>
<tr>
<td>T1</td>
<td>345 ± 74 321 ± 59 250 ± 46</td>
</tr>
<tr>
<td>I1</td>
<td>347 ± 72 325 ± 70 256 ± 59</td>
</tr>
<tr>
<td>N2</td>
<td>307 ± 42 302 ± 44 250 ± 49</td>
</tr>
<tr>
<td>S2</td>
<td>297 ± 42 292 ± 42 223 ± 28</td>
</tr>
<tr>
<td>T2</td>
<td>284 ± 49 278 ± 46 220 ± 45</td>
</tr>
<tr>
<td>I2</td>
<td>292 ± 61 286 ± 63 233 ± 55</td>
</tr>
</tbody>
</table>

### Table 2

Intraclass correlation coefficient percentages before and after correction.

<table>
<thead>
<tr>
<th>Subfield</th>
<th>ICC values (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(All eyes)</td>
</tr>
<tr>
<td></td>
<td>NV-AMD</td>
</tr>
<tr>
<td></td>
<td>NNV-AMD</td>
</tr>
<tr>
<td></td>
<td>Spectralis™</td>
</tr>
<tr>
<td></td>
<td>Stratus™</td>
</tr>
<tr>
<td></td>
<td>Cirrus™</td>
</tr>
<tr>
<td></td>
<td>Spectralis™</td>
</tr>
<tr>
<td></td>
<td>Stratus™</td>
</tr>
<tr>
<td></td>
<td>Cirrus™</td>
</tr>
<tr>
<td></td>
<td>Spectralis™</td>
</tr>
<tr>
<td></td>
<td>Stratus™</td>
</tr>
<tr>
<td></td>
<td>Cirrus™</td>
</tr>
<tr>
<td>C1</td>
<td>99.6 99.4 97.2</td>
</tr>
<tr>
<td>N1</td>
<td>99 99.5 87.6*</td>
</tr>
<tr>
<td>S1</td>
<td>99.3 99.5 82.8*</td>
</tr>
<tr>
<td>T1</td>
<td>98.2 99 96</td>
</tr>
<tr>
<td>I1</td>
<td>98.9 99 95.9</td>
</tr>
<tr>
<td>N2</td>
<td>99.1 99.6 98.4</td>
</tr>
<tr>
<td>S2</td>
<td>99.6 99.5 71.2*</td>
</tr>
<tr>
<td>T2</td>
<td>99.2 99.4 92.7</td>
</tr>
<tr>
<td>I2</td>
<td>98.3 99 94.9</td>
</tr>
</tbody>
</table>
the highest retinal thickness values (range: 280 to 372 µm),
dependent on the subfield. The retinal thickness measure-
ments obtained via Cirrus™ were slightly less (range: 230 to
320 µm), while Stratus™ had the lowest values, ranging from
190 to 270 µm. There were no significant (p<0.05) differences
between the mean FTH of the first and second scans for each
of the three devices.

The central subfield ICC values for all three machines were
very high at 99.6%, 97.2% and 96.4% before correction for
Spectralis™, Stratus™, and Cirrus™ respectively, and 99.4%,
and 97.4% after correction for Spectralis™ and Cirrus™. The
ICC values were greater than 95% for all subfields and both
diagnoses except the outer inferior field for NNV-AMD for
Spectralis™. Stratus™ values ranged from 78.9% to 99.2% for
NV-AMD and 94.7% to 99% for NNV-AMD, before and af-
after correction, respectively. Cirrus™ values ranged from 88.5%
to 99.9% and 99.1% to 99.8% for NV-AMD before and after
correction, respectively. The values for NNV-AMD for Cir-
rus™ ranged from 99.3% to 99.9% and 71.4% to 99.7% be-
fore and after correction, respectively. Table 2 shows the ICC
values between images for all three machines before and after
correction, both combined and stratified by diagnosis. It should
be noted that all of the machines had ICC values >90% for the
central subfield while the Spectralis™ had no subfields less than
99% after correction. In the central subfield, Spectralis™ had a
COR of 20 µm NV-AMD which increased to 23 µm; both Cir-
rus™ and Stratus™ had relatively larger CORs of 64 µm (re-
duced to 49 µm after correction) and 35 µm, respectively. For
NNV-AMD, the COR for the central subfield was 15 µm for
both Cirrus™ and Spectralis™, and was 24 µm for Stratus™.
After correction, the value decreased for Spectralis™ to 12 µm
and increased to 36 µm for Cirrus™. The COR of all subfields
for each device before and after correction of algorithms and
stratification by disease are given in Table 3.

Overall Spectralis™ had the lowest COR, with values ranging
from 5–30 µm. Cirrus™ and Stratus™ had similar values rang-
ing from 5–70 µm, even after correction. The COR for Cir-
rus™ increased by 15–40 µm after correction for NNV-AMD.
Also, Cirrus™ COR values were 10-30 µm higher than Stra-
tus™ values for both NV-AMD and NNV-AMD. Agreement
between machines was poor, except between Spectralis™ and
Cirrus™ after correction. Tables 4–5 show 95% confidence
intervals and limits of agreement of the Bland-Altman plots
between devices before and after manual correction.

Figure 2A–F show Bland-Altman plots with 95% confidence
intervals for the FTH comparison of the machines before and
after correction. Before correction, the mean difference
between the machines was 32 µm for Spectralis™ vs. Cirrus™,
52 µm for Cirrus™ vs. Stratus™, and 84 µm for Spectralis™
vs. Stratus™. Manual correction reduced the differences, with
it being 15 µm for Spectralis™ vs. Cirrus™, 51 µm for Cir-
rus™ vs. Stratus™, and 67 µm for Spectralis™ vs. Stratus™.
When stratified by diagnoses, the values were 34 µm and
29 µm for Spectralis™ vs. Cirrus™, 53 µm and 47 µm for
Cirrus™ vs. Stratus™, and 88 µm and 79 µm for Spectralis™
vs. Stratus™ for NV-AMD and NNV-AMD respectively, be-
fore correction. After manual correction, the values reduced to
17 µm and 14 µm Spectralis™ vs. Cirrus™ and 70 µm and
61 µm Spectralis™ vs. Stratus™ for NV-AMD and NNV-
AMD, respectively. The confidence interval widths, on aver-
age, were 5–10 µm smaller than when comparing between an
SD-OCT and a TD-OCT machine. The average interval width
decreased between 5–10 µm after correction for any disease and
comparison, except for the Cirrus™ vs. Stratus™ comparison.

Discussion
The advent of OCT has revolutionized the way patients with
retinal disorders are evaluated and monitored. However, like
every new device, the current devices employing time- or spec-
tral domain technology have certain limitations. One such
common and clinically relevant issue is the presence of random
errors in the identification of the inner and outer boundaries of
the retina by the OCT algorithm. With respect to AMD,
studies have shown that lesions such as fibrotic scars, choroi-
dal neovascularization disrupting the RPE, and subretinal fluid
would produce errors in the automated segmentation algo-
rithms because the software would not correctly delineate the
outer retinal boundary. In our study, we found that 66% of
the Spectralis™, 14% of the Cirrus™ and 6.5% of the Stra-
tus™ scans had algorithm errors. Giani et al. reported similar
results; for Cirrus™, they reported 25% and 16% algorithm
error rates for NNV-AMD and NV-AMD, respectively. How-
ever, for Spectralis™, they reported 16.67% and 57.6% algo-

\[ http://dx.doi.org/10.6084/m9.figshare.679840 \]

Reasons for differences in our error rates compared to previous
studies include the lack of standard definition of an algorithm
error. Rather than having an exact definition of an algorithm
error, which may not be clinically significant, in our study, the
decision was made by two masked observers who determined
if the correction would be important. In addition, even though
Spectralis™ segments the outer border of the RPE, a study by
Jaffe et al. reported that it may also be including the Bruch's
membrane in its calculation, thus including sub-RPE pathology such as drusen when segmenting the outer border of the retina\(^7\). These differences may be due to the fact that our study was prospective and, while acquiring scans, the operators tried their best to ensure no errors occurred during scan acquisition. Lastly, we did not exclude scans if the signal strength was low or if the machine gave a low analysis confidence message, as other studies have done\(^8\)-\(^10\).

After correction the thickness measurements for the Spectralis\(^\text{TM}\) and Cirrus\(^\text{TM}\) scans were not significantly different. This may be due to the fact that the majority of the scans required minor corrections. For example, most of the Spectralis\(^\text{TM}\) scans resulted in a 10 µm or less change in the central subfield thickness. Retinal thickness measurements were similar in both SD-OCT machines and were greater than Stratus\(^\text{TM}\). Correction reduced the difference of the thickness measurements between the two SD-OCT devices to less than 20 µm; in some cases as noted above, the difference was no longer statistically significant. Other studies in normal and pathologic eyes including Diabetic Macular Edema (DME) and macular degeneration have also demonstrated that the difference in retinal thickness between the SD machines can be attributed to the differences in segmentation of the automated algorithms\(^7\),\(^10\),\(^11\).

Despite the large numbers of scans with algorithm errors, the COR of Spectralis\(^\text{TM}\) was lower for every subfield than that of Stratus\(^\text{TM}\) or Cirrus\(^\text{TM}\). The COR of Cirrus\(^\text{TM}\) was equal to or larger than Stratus\(^\text{TM}\) for both forms of the disease. In all three SD-OCT, the COR was generally better for NNV-AMD when compared to NV-AMD, especially after correction. This difference between diseases can be attributed to the pathology of NV-AMD disrupting the outer border, which makes it difficult for the automated algorithm to accurately segment the retinal layers\(^12\),\(^13\). At this point, we are not aware of any previous study looking at the repeatability of Spectralis\(^\text{TM}\) images in AMD. Previous studies on normal eyes have reported a high repeatability of measurements with Spectralis\(^\text{TM}\), with differences between repeated measurements being within 1 µm\(^11\),\(^14\). For Stratus\(^\text{TM}\) OCT images, other studies have found central subfield repeatability values in patients with NV-AMD to be 50 µm and 32–35 µm for NNV-AMD patients after correction/exclusion of scans with errors\(^8\),\(^15\); our study confirms this finding. There has been one other published study looking at the repeatability of Cirrus\(^\text{TM}\) OCT in NV-AMD, which found a central subfield repeatability value of 42 µm before correction and 27 µm after exclusion\(^16\). The difference between this study and our measurements may be associated with our smaller sample size. In addition, we chose not to exclude any poor quality scans, which may cause larger differences.

In addition to a lower COR, Spectralis\(^\text{TM}\) also had the highest ICC values for both NV-AMD and NNV-AMD, before and after correction. For NV-AMD, Cirrus\(^\text{TM}\) had higher coefficients after correction, and for NNV-AMD, Cirrus\(^\text{TM}\) had lower coefficients compared to Stratus\(^\text{TM}\). While no previous studies have reported ICC values for AMD patients, Pierro et al. found comparable results in normal eyes, with Cirrus\(^\text{TM}\) ICC values ranging from 83–97% and Stratus\(^\text{TM}\) ICC values from 72–95%\(^17\). The most likely reason for the low repeatability and high ICC values for Spectralis\(^\text{TM}\) is the eye-tracking capability, which ensures that artifacts due to eye movement are minimized and the machine scans only when the tracking software identifies the same position on the fundus\(^14\).

Bland-Altman plots indicate that there is agreement between SD-OCT machines. Correcting images also influenced agreement between machines. We found that 95% confidence intervals were narrower compared to an SD-OCT and TD-OCT and correcting the algorithm errors further narrowed the intervals.
Table 4 95% confidence intervals for Bland-Altman Plots. A: Before correction. B: After correction.

### A. Bland-Altman 95% confidence intervals before correction (µm)

<table>
<thead>
<tr>
<th>Subfield</th>
<th>All eyes</th>
<th>NV-AMD</th>
<th>NNV-AMD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spectralis™ vs. Cirrus™ vs. Stratus™</td>
<td>Spectralis™ vs. Cirrus™ vs. Stratus™</td>
<td>Spectralis™ vs. Cirrus™ vs. Stratus™</td>
</tr>
<tr>
<td>Central 1 mm</td>
<td>42, 89</td>
<td>37, 65</td>
<td>91, 149</td>
</tr>
<tr>
<td>N1</td>
<td>29, 63</td>
<td>44, 65</td>
<td>81, 118</td>
</tr>
<tr>
<td>S1</td>
<td>21, 49</td>
<td>44, 71</td>
<td>74, 109</td>
</tr>
<tr>
<td>T1</td>
<td>21, 43</td>
<td>41, 59</td>
<td>73, 99</td>
</tr>
<tr>
<td>I1</td>
<td>23, 43</td>
<td>41, 74</td>
<td>79, 111</td>
</tr>
<tr>
<td>N2</td>
<td>13, 24</td>
<td>40, 61</td>
<td>57, 81</td>
</tr>
<tr>
<td>S2</td>
<td>16, 27</td>
<td>42, 73</td>
<td>59, 93</td>
</tr>
<tr>
<td>T2</td>
<td>9, 17</td>
<td>35, 49</td>
<td>52, 65</td>
</tr>
<tr>
<td>I2</td>
<td>9, 33</td>
<td>28, 51</td>
<td>54, 67</td>
</tr>
<tr>
<td>Average Width</td>
<td>21.5</td>
<td>24</td>
<td>30.2</td>
</tr>
</tbody>
</table>

### B. Bland-Altman 95% confidence intervals after correction (µm)

<table>
<thead>
<tr>
<th>Subfield</th>
<th>All eyes</th>
<th>NV-AMD</th>
<th>NNV-AMD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spectralis™ vs. Cirrus™ vs. Stratus™</td>
<td>Spectralis™ vs. Cirrus™ vs. Stratus™</td>
<td>Spectralis™ vs. Cirrus™ vs. Stratus™</td>
</tr>
<tr>
<td>Central 1 mm</td>
<td>8, 30</td>
<td>40, 70</td>
<td>55, 91</td>
</tr>
<tr>
<td>N1</td>
<td>11, 34</td>
<td>45, 64</td>
<td>63, 88</td>
</tr>
<tr>
<td>S1</td>
<td>5, 29</td>
<td>43, 69</td>
<td>56, 87</td>
</tr>
<tr>
<td>T1</td>
<td>3, 21</td>
<td>44, 58</td>
<td>53, 79</td>
</tr>
<tr>
<td>I1</td>
<td>4, 17</td>
<td>41, 74</td>
<td>60, 86</td>
</tr>
<tr>
<td>N2</td>
<td>8, 20</td>
<td>39, 60</td>
<td>53, 74</td>
</tr>
<tr>
<td>S2</td>
<td>10, 26</td>
<td>41, 70</td>
<td>54, 90</td>
</tr>
<tr>
<td>T2</td>
<td>5, 15</td>
<td>34, 47</td>
<td>45, 60</td>
</tr>
<tr>
<td>I2</td>
<td>3, 29</td>
<td>29, 50</td>
<td>48, 60</td>
</tr>
<tr>
<td>Average Width</td>
<td>18.3</td>
<td>22.8</td>
<td>25.3</td>
</tr>
</tbody>
</table>

Average Width
<table>
<thead>
<tr>
<th>Subfield</th>
<th>All eyes</th>
<th>NV-AMD</th>
<th>NNV-AMD</th>
<th>Spectrals™ vs. Cirrus™</th>
<th>Spectrals™ vs. Stratus™</th>
<th>Spectrals™ vs. Cirrus™</th>
<th>Spectrals™ vs. Stratus™</th>
<th>Spectrals™ vs. Cirrus™</th>
<th>Spectrals™ vs. Stratus™</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central 1 mm</td>
<td>63 (228, -101)</td>
<td>52 (142, -38)</td>
<td>120 (305, -64)</td>
<td>64 (220, -92)</td>
<td>64 (165, -37)</td>
<td>127 (319, -64)</td>
<td>69 (244, -150)</td>
<td>32 (85, -21)</td>
<td>110 (287, -68)</td>
</tr>
<tr>
<td>N1</td>
<td>32 (112, -47)</td>
<td>50 (106, -6)</td>
<td>87 (168, 4)</td>
<td>35 (119, -49)</td>
<td>51 (115, -14)</td>
<td>92 (175, 8)</td>
<td>28 (85, -29)</td>
<td>49 (93, 7)</td>
<td>79 (158, 1)</td>
</tr>
<tr>
<td>N2</td>
<td>34 (108, -40)</td>
<td>58 (142, -26)</td>
<td>92 (203, -20)</td>
<td>40 (160, -80)</td>
<td>65 (164, -35)</td>
<td>103 (235, -28)</td>
<td>28 (68, -13)</td>
<td>47 (92, 1)</td>
<td>75 (134, 15)</td>
</tr>
<tr>
<td>S1</td>
<td>41 (143, -60)</td>
<td>55 (120, -10)</td>
<td>100 (217, -17)</td>
<td>52 (185, 82)</td>
<td>60 (138, 17)</td>
<td>107 (243, -27)</td>
<td>39 (129, -51)</td>
<td>47 (87, 7)</td>
<td>89 (172, 5)</td>
</tr>
<tr>
<td>I1</td>
<td>36 (100, -29)</td>
<td>58 (163, -47)</td>
<td>96 (195, -4)</td>
<td>32 (103, -38)</td>
<td>55 (182, -72)</td>
<td>94 (198, -10)</td>
<td>34 (100, -31)</td>
<td>62 (123, 1)</td>
<td>96 (194, 2)</td>
</tr>
<tr>
<td>N2</td>
<td>16 (40, -9)</td>
<td>42 (86, -2)</td>
<td>59 (98, 19)</td>
<td>14 (49, -19)</td>
<td>41 (95, -15)</td>
<td>59 (107, 11)</td>
<td>12 (34, -9)</td>
<td>44 (55, 23)</td>
<td>58 (81, 35)</td>
</tr>
<tr>
<td>S2</td>
<td>25 (74, -24)</td>
<td>58 (154, -38)</td>
<td>77 (175, -21)</td>
<td>23 (66, -20)</td>
<td>67 (186, -52)</td>
<td>88 (214, 40)</td>
<td>21 (38, 4)</td>
<td>44 (62, 25)</td>
<td>63 (93, 34)</td>
</tr>
<tr>
<td>T1</td>
<td>18 (50, -13)</td>
<td>51 (117, 15)</td>
<td>69 (145, 7)</td>
<td>20 (62, -22)</td>
<td>54 (132, -28)</td>
<td>70 (152, 13)</td>
<td>17 (41, 7)</td>
<td>46 (76, 15)</td>
<td>68 (138, 1)</td>
</tr>
<tr>
<td>I2</td>
<td>22 (95, -50)</td>
<td>40 (106, -26)</td>
<td>54 (147, -40)</td>
<td>26 (116, 42)</td>
<td>56 (92, 19)</td>
<td>15 (58, -29)</td>
<td>53 (99, 7)</td>
<td>69 (105, -33)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5 Mean difference (limits of agreement). A: Before correction. B: After correction.**
The mean difference between machines indicates that the lowest differences were between Spectralis™ and Cirrus™, especially after correction. This is mostly likely due to the effects of manually correcting the Spectralis™ images and the fact that both machines have similar scanning technologies. The limits of agreement were similarly very wide for all three machines, and were narrower after correction of images, especially for the two SD-OCT machines. Jaffe et al. reported similar results looking at NV-AMD, with limits of agreements being approximately 225 µm between a SD-OCT and TD-OCT². The poor agreement suggests that clinicians should exercise caution when trying to use the data from different machines interchangeably.

Our study is not without its limitations. First, the software version for the Stratus™ images would not allow correction of images. However, very few images needed to be corrected. In addition, two people independently manually corrected the images, resulting in inaccuracies in segmentation line correction. Finally, the images were only taken at one imaging center, which could have resulted in bias.

In summary, we found that although Spectralis™ had the highest frequency of errors in AMD patients, correction of images did not result in significant changes in retinal thickness due to the errors being very small. Spectralis™ had the lowest coefficient of repeatability values. Thus Spectralis™ may be the best suited for examining minute morphological and thickness changes. Also, because of the wide Bland-Altman 95% intervals, there is not much agreement between the SD-OCT and TD-OCT machines. Based on our findings, we recommend...
that scans be carefully analyzed at reading centers before the thickness values are accepted as reliable.

**Author contributions**

AR and YJS and RC conceived the study. AR, RC, and EH and MS carried out the research. AR, YJS, RC and MS prepared the first draft of the manuscript. DVD and QDN supervised the project. All authors were involved in the revision of the draft manuscript and have agreed to the final content.

**Competing interests**

No competing interests were disclosed.

**Grant information**

The author(s) declare that no grants were involved in supporting this work.

**Acknowledgments**

The current manuscript was partially presented at the ARVO annual meeting, Fort Lauderdale, Florida in 2010.

### Supplementary figure

![Figure S1 The Early Treatment Diabetic Retinopathy Study grid. The following Early Treatment Diabetic Retinopathy Study grid depicts the abbreviations for the nine subfields. C1–Central 1 mm, N1–Inner nasal, S1–Inner superior, T1–Inner temporal, I1–Inner inferior, N2–Outer nasal, S2–Outer superior, T2–Outer temporal, I2–Outer inferior.](image)

**References**


Igor Kozak
Vitreoretinal Division, King Khaled Eye Specialist Hospital, Riyadh, Saudi Arabia

This study reports on reliability and reproducibility of optical coherence tomography (OCT) scans before and after manual correction in eyes with age-related macular degeneration (AMD). It concludes that manual correction improves automated segmentation and that the agreement, as evidenced by intraclass correlation coefficients and coefficients of repeatability, is better between spectral-domain OCT instruments. The study brings important numerical comparisons of measurements of central foveal thickness using different instruments. Such comparisons are crucial especially with transitioning from time-domain to spectral-domain OCT technology in ongoing and future clinical trials. The paper is well written. Minor comments for authors:

1. The same operator acquired OCT scans of the same eyes on the same instruments. It would be useful to know how many graders and how independently performed manual correction at the Reading Center.

2. The Spectralis system via TruTrack provides excellent ability to perform follow-up scans from the exact retinal areas. How was this dealt with using other two systems in order to avoid sampling error?

3. Misidentification of inner retinal layer is a common artifact and has been found in a large number of scans including the instruments used in this study. The authors are encouraged to cite some of those studies such as Ho et al. (2009).

4. Another useful study to mention with respect to comparing time-domain and spectral-domain OCT instruments: Mylonas et al. (2009).

5. Based on a study of reproducibility in Stratus OCT, any artifact resulting in an error that is more than 50μm is clinically significant, suggesting 50 μm as a cutoff for retreatment of neovascular AMD patients (Patel et al., 2009). In another study any artifacts resulting in automated segmentation errors of more than 10% of the actual (manually measured) ETDRS center subfield thickness were considered clinically significant (Browning et al., 2008). In this study, some of the variations after manual correction surpassed these margins.
Maybe some comment in Discussion regarding this issue.

**Competing Interests:** No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
significant reservations, as outlined above.

Author Response ( ) 28 Feb 2014

Yasir Sepah, Stanford University, USA

Dear Dr. Krebs,

Thank you for your valuable comments. We agree that our study is not the only study that has tried to deal with this topic. We will make changes to the manuscript to clarify this statement.

We will add the results and analysis of the peripheral areas to the discussion in the revision.

No sample size calculation was performed before the conduct of the study. Two eyes of the same patient were included because of difference in the severity of the disease between the two eyes of the same patient. The control group patients contributed only one eye to the analysis.

**Competing Interests:** None