Quality of the Threshold Algorithm in Age-Related Macular Degeneration: Stratus versus Cirrus OCT

Ilse Krebs,1,2 Christiane Falkner-Radler,1,2 Stefan Hagen,1 Paulina Haas,1,2 Werner Brannath,3 Shilla Lie,1 Siamak Ansari-Shahrezaei,1,2 and Susanne Binder1,2

PURPOSE. Automatically generated measurements of the retinal volume or the central retinal thickness are based on correctly set threshold lines on the retinal surface and the retinal pigment epithelium. The purpose of this study was to compare the accuracy of threshold algorithm lines of Stratus OCT (macular cube 512 × 128) with those of Cirrus OCT.

METHODS. A consecutive series of patients at least 50 years of age with exudative age-related macular degeneration was included. Stratus OCT (retinal thickness program) and Cirrus OCT examinations (macular cube 512 × 128) were performed by the same examiner, the sequence of the examinations was randomized. Two independent examiners evaluated the positioning of the threshold algorithm lines and performed a grading of the failures. Logistic regression analysis was applied for evaluation of the failure rate.

RESULTS. One hundred four patients were included. For the entire OCT examination (6 scans Stratus OCT, 128 scans Cirrus OCT) algorithm line failures were detected in 69.2% of the Stratus OCT and in 25% of the Cirrus OCT examinations, with the difference reaching statistical significance (P < 0.001). The median failure grade was 1 (0–6) for Stratus and 0 (0–5.15) for Cirrus OCT. Age, measurement sequence, and investigator did not influence the error rates.

CONCLUSIONS. With Cirrus OCT automatically performed and therefore objective measurements of central retinal thickness and retinal volume were provided correctly in 69.2% of the scans. Furthermore, this latest software version offers the possibility of manual correction of false positioned algorithm lines (ClinicalTrials.gov number, NCT00568191). (Invest Ophtalmol Vis Sci. 2009;50:995–1000) DOI:10.1167/iovs.08-2617

Modern therapy for age-related macular degeneration (AMD), including antivasoproliferative substances, is increasingly based on the results of optical coherence tomography (OCT). Morphologic changes assessed by OCT, particularly measurements of retinal thickness, contribute to the evaluation of the activity of a neovascular lesion and are closely connected to the decision process for retreatment.1,2 Retinal thickness corresponds to the amount of fluid exudation of a neovascular lesion and consequently to the activity of a lesion,1,2 providing valuable data not only in clinical practice but also in clinical studies.2–4 Retinal thickness measurement is merely based on automatically set threshold algorithm lines on the retinal surface and the most outer high-reflective band, which is thought to correspond to the RPE–Bruch’s membrane complex. The accuracy of automatic measurements providing objective data is based on the correctness of the position of these border lines. Prior publications have reported on a considerable amount of threshold algorithm failures in Stratus OCT, the most widespread use OCT system in the recent past.5,6 A new OCT technology is now becoming available, the spectral domain OCT, providing considerable advantages concerning resolution and acquisition speed compared with the time-domain technology of the Stratus OCT.7

The purpose of this study was to compare the accuracy of the positioning of the threshold algorithm lines of the time domain (Stratus OCT) with that of the spectral domain OCT (Cirrus OCT). The influence of the higher resolution and different acquisition modality of the Cirrus OCT on the correctness of the border lines was examined. A prospective randomized examiner-blinded comparative crossover design was used for this study.

METHODS

Patients aged at least 50 years with neovascular AMD verified by fluorescein angiography (FA), regardless of whether newly diagnosed or pretreated, were recruited into this prospective study from January to April 2008. According to the results of the pilot study (Krebs I, et al. IOVS 2008;49:ARVO E-Abstract 1884), a sample of 104 participants was necessary. The sample size calculation was geared to the sign test at a two-sided level of 0.05 and a power of 80%, assuming an absolute difference of 20% in the failure rates between Cirrus and Stratus OCT and a 50% proportion of discordant pairs. Only one eye of each patient was included, and the eye currently referred for treatment, under treatment, or under observation after treatment was selected. If both eyes met the inclusion criteria the eye with the better distance acuity was selected. According to biomicroscopy with or without fluorescein angiography a staging of the lesions was performed: stage 1, active lesion without fibrotic parts; and stage 2, lesion containing fibrosis. After distance acuity testing was performed with ETDRS charts at a 2-m distance, the pupils were dilated and Stratus and Cirrus OCT were performed.

The time domain OCT was performed with the Stratus OCT model 3000, software version 4.0 (Carl Zeiss Meditec, Inc., Dublin, CA), using...
the retinal thickness program with six radial lines through the center of the foveal region.

Spectral domain OCT was performed with the Cirrus OCT (Carl Zeiss Meditec, Inc.) software version 3.0. The macular cube 512 × 128 program was applied consisting of 128 horizontal lines of 512 A-scans. For both examinations an internal fixation beam was used or an external fixation light for the fellow eye when internal fixation was not possible. Fixation was controlled by the examiner.

To eliminate a possible influence of the sequence of examination, of different examination regimen (6 lines in Stratus OCT, 128 lines in Cirrus OCT), and of the examiners evaluating the examinations, we implemented a series of precautions:

1. Each of the two examiners (SH, PH) performed Stratus and Cirrus OCT in 52 of the 104 patients included.
2. Stratus and Cirrus tests were performed immediately, one after the other. The sequence of the OCT examinations was block randomized so that 52 patients first underwent Cirrus and then Stratus and the other 52 patients first underwent Stratus and then Cirrus examinations. The randomization was stratified by examiner (SH, PH).
3. The images of the central horizontal scan of both OCT systems were printed and anonymized by the study investigator (IK).
4. The only identical scan, the central horizontal scan (scan 64 of the cube 512 × 128 program of Cirrus OCT and the horizontal scan of Stratus OCT) was chosen to provide identical conditions for both examinations.
5. The Stratus and Cirrus central horizontal scan prints were evaluated by two independent examiners (SA, CF), who were randomly assigned to evaluate Stratus or Cirrus printouts of a patient’s results. This random allocation was stratified according to those who performed the OCT examinations (SH, PH) and sequence of OCT examinations.

For a correct measurement of central retinal thickness (CRT; mean thickness in an area of a 1 mm diameter) and retinal volume (RV; volume between the two threshold lines in a 6-mm area) the threshold algorithm lines of all scans of the examinations (6 in Stratus OCT, 128 in Cirrus OCT) have to be placed correctly. IK and SH also evaluated the overall Stratus or Cirrus OCT examination for correctness and performed a grading of the severity of threshold algorithm line errors. The Stratus or Cirrus OCT examination was randomly assigned to either IK or SH, so that each of them evaluated only one examination per patient. This random assignment was also stratified according to those performing the OCT examination (SH, PH) and sequence of OCT examinations. The grading system Sadda et al. have introduced for Stratus OCT was used and applied also to Cirrus OCT. Points were given for each of the following errors in each of the 6 Stratus scans and in each of the 128 Cirrus scans. Central location was defined as the part of the examination used for central retinal thickness measurement (central 1 mm of each Stratus scan; central area with a diameter of 1 mm in Cirrus OCT): (1) Generally, every failure was given 1 point, (2) plus 1 point if the extent of the failure was more than 1 mm and less than 3 mm; (3) plus 2 points if the extent was more than 3 mm; (4) plus 1 point if the failure was more than one third but less than two thirds of the retinal thickness; (5) plus 2 points if the failure was more than two thirds of the retinal thickness; (6) plus 1 point if the failure was in the central location.

The overall scores of the examination were averaged (divided by 6 for Stratus OCT and 128 for Cirrus OCT) to generate an average error score in each case with Stratus and with Cirrus OCT. CRT and RV were determined and the results compared between Stratus and Cirrus OCT.

The Austrian ethics committee, Government of the City of Vienna, approved the study, all the patients signed a written consent, and the data collection was in compliance with the tenets of the Declaration of Helsinki.

### Statistical Methods

Quantitative variables were summarized by minimum (min), maximum (max) first, second, and third quartiles (Q1, Med, and Q3), qualitative variables by frequency tables. The threshold algorithm line failure rates at the central horizontal scan and the whole examination error grading were compared between Stratus and Cirrus by pair-wise sign tests. In addition, supplementary logistic regression analyses with robust variance estimates were performed to estimate the difference in the failure rates between Stratus and Cirrus as well as the influence of baseline characteristics such as age, distance acuity, and stage of the lesion on the corresponding failure rate. Additional covariates of these logistic regression analyses were the measurement sequence (first Cirrus or first Stratus OCT) and investigator (SH or PH). In the logistic regression analysis for the whole-examination failure rate, failure was defined to occur when the average whole-examination grading was at least 1/6. This corresponds to a whole-examination grade of at least 1 for Stratus and of at least 128/6 = 21.3 for Cirrus. The relationship between the two OCT measurements of CRT and RV was investigated by Pearson’s correlation coefficients, as well as their patient-wise differences by box plots, means, and 95% confidence intervals for the mean. $P < 0.05$ was considered statistically significant.

### RESULTS

One hundred four patients with a median age of 78 years (range, 52–95 years) were included: 65.4% women, 34.6% men. Of the eyes included, 39.8% were untreated, and 70.2% were in treatment. Of the lesions identified, 68.7% lesions were occult and minimal classic lesions without signs of fibrosis (stage 1), and 31.3% contained fibrotic parts (stage 2). Distance acuity and OCT parameters were similar in both sequence groups (Table 1).

The failure rate in the central horizontal scan with Stratus OCT was 40.4% (42/104), whereas it was 9.6% (10/104) with Cirrus OCT (Table 2). The sign test indicated a statistically significant difference in the failure rate for Stratus and Cirrus OCT.

#### Table 1. Summary of OCT Parameters

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Minimum</th>
<th>Maximum</th>
<th>25% Quantile</th>
<th>Median</th>
<th>75% Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRT Stratus, all</td>
<td>104</td>
<td>0.02</td>
<td>557</td>
<td>209.8</td>
<td>244</td>
<td>301.5</td>
</tr>
<tr>
<td>CRT Cirrus, all</td>
<td>104</td>
<td>131</td>
<td>542</td>
<td>270.8</td>
<td>305</td>
<td>349.3</td>
</tr>
<tr>
<td>CRT Stratus, correct</td>
<td>42</td>
<td>131</td>
<td>557</td>
<td>198.5</td>
<td>221.5</td>
<td>266</td>
</tr>
<tr>
<td>CRT Cirrus, correct</td>
<td>85</td>
<td>166</td>
<td>542</td>
<td>275</td>
<td>307</td>
<td>349</td>
</tr>
<tr>
<td>RV Stratus, all</td>
<td>104</td>
<td>4.6</td>
<td>10.34</td>
<td>6.53</td>
<td>6.98</td>
<td>7.52</td>
</tr>
<tr>
<td>RV Cirrus, all</td>
<td>104</td>
<td>3.3</td>
<td>14.8</td>
<td>9.5</td>
<td>10</td>
<td>10.83</td>
</tr>
<tr>
<td>RV Stratus, correct</td>
<td>32</td>
<td>4.6</td>
<td>8.31</td>
<td>6.53</td>
<td>6.77</td>
<td>7.17</td>
</tr>
<tr>
<td>RV Cirrus, correct</td>
<td>73</td>
<td>8.2</td>
<td>14.8</td>
<td>9.7</td>
<td>10.1</td>
<td>10.8</td>
</tr>
</tbody>
</table>

All, all 104 measurements; correct, only measurements without border line failures.
TABLE 3. Logistic Regression Analysis of Failure Assessments in the Central Horizontal Scan

<table>
<thead>
<tr>
<th>Type of Failure</th>
<th>Stratus OCT</th>
<th>Cirrus OCT</th>
<th>Odds Ratio</th>
<th>Lower CI</th>
<th>Upper CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure rates for central horizontal line</td>
<td>Stratus OCT</td>
<td>Cirrus OCT</td>
<td>0.147</td>
<td>0.067</td>
<td>0.323</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Failure rates in entire examination (average grading ≥ 1/6)</td>
<td>Stratus OCT</td>
<td>Cirrus OCT</td>
<td>0.207</td>
<td>0.053</td>
<td>0.803</td>
<td>0.02</td>
</tr>
<tr>
<td>Age</td>
<td>1.002</td>
<td>0.963</td>
<td>1.043</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence</td>
<td>1.195</td>
<td>0.599</td>
<td>2.385</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investigator</td>
<td>1.080</td>
<td>0.540</td>
<td>2.157</td>
<td>0.83</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n = 104.

The box plots in Figure 1 compare the distance acuity of patients with correctly and incorrectly set lines determined by Cirrus OCT. Failure rates were also found to be higher in patients with stage 2 than in those with stage 1 lesions. (Cirrus OCT failures were 7% in stage 1 and 15.2% in stage 2; Stratus OCT failures 35.2% in stage 1 and 51.5% in stage 2). The influence of stage barely reached statistical significance when adding stage as factor to the logistic regression analysis while the covariates did not have an influence on the failure rate in the overall examination (6 scans of Stratus OCT and 128 scans of Cirrus OCT) increased the grading between Cirrus and Stratus OCT was −0.83 in median (min, −6; Q1, −1.9; Q3, 0; max, 4.82) and the sign test indicated that overall examination failures were of a lower grade with Cirrus than with Stratus (P < 0.001).

A considerable difference between Stratus and Cirrus OCT in the distribution of failures was noted concerning misidentification of the anterior threshold algorithm line (57.7% in Stratus OCT versus 12.5% in Cirrus OCT). However, the number of scan artifacts was similar in Stratus and Cirrus OCT. More detailed information is presented in Table 5.

Examples of correct and incorrect positioning of algorithm lines are presented in Figure 2. A comparison of the Stratus and Cirrus images in the figure shows the alignment of the Stratus OCT images (detachment of the RPE is considerably flattened).8

The median CRT was 244 μm in Stratus OCT and 305 μm in Cirrus OCT when all 104 measurements were accounted for, including those with artifacts in the central millimeter. Measurements without artifacts in the central millimeter led to a median CRT of 221.5 μm in Stratus and 307 μm in Cirrus OCT. More detailed information is presented in Table 1. Thirty-six patients had no failures in the central millimeter in both OCT Cirrus measurements. The correlation of these Cirrus and Stratus CRT measurements was 0.91 and their mean difference 55.5 μm (95% CI: 46.3–64.7 μm).

The median of the RV was 6.98 mm³ in Stratus OCT and 10.0 mm³ in Cirrus OCT in all 104 patients. Concerning measurements without any failures (24 eyes) the median RV was 6.77 mm³ in Stratus and 10.1 mm³ in Cirrus OCT. More details...
are found in Table 1. The correlation of the two OCT volume measurements in the eyes was 0.91. The mean difference between the Cirrus and Stratus volume measurements was 3.12 mm$^3$ (95% CI: 3.02–3.21 mm$^3$).

**DISCUSSION**

Automatically performed correct measurements of the retinal thickness in OCT examinations are based on correctly set threshold algorithm lines at the retinal surface and the hyperreflective band of the retinal pigment epithelium (RPE)-choriocapillaris layer. In published articles, a considerable number of algorithm line failures have been reported in the Stratus OCT, especially in cases of AMD compared with eyes with other diseases.\(^9\),\(^10\) However, the interreader agreement concerning the detection of algorithm line failures was very good.\(^11\),\(^12\) We found in a retrospective study of 233 participants, other diseases.\(^9\),\(^10\) However, the interreader agreement concerning the detection of algorithm line failures was very good.\(^11\),\(^12\) We found in a retrospective study of 233 participants, scores were significantly correlated to the diagnosis AMD. Ray et al.\(^13\) presented results of a study with 171 participants, found an incidence of algorithm failures in 92%, with 13.5% of them having severe failures. Higher average failure scores were significantly correlated to the diagnosis AMD. Ray et al.\(^13\) presented results of a study with 171 participants, 15.8% of them with neovascular AMD. Algorithm failures were frequent, in 70% of eyes with neovascular AMD and 86% of eyes treated with photodynamic therapy. As the Cirrus OCT has recently become commercially available, only a few data regarding segmentation failures have been published. Ahlers et al.\(^14\) reported on a small series (22 eyes) of selected patients (eyes with fibrovascular RPE detachment) and found significant errors in 27.3% and limited quality in 9.89% evaluated with a prototype of Cirrus OCT. In the present study the overall failure rate for Stratus OCT (69.2%) reconfirmed the results of prior studies. In Cirrus OCT threshold algorithm failures were significantly less frequent (29.8%). Reasons for failures were mostly disease related; scan artifacts were rare in Stratus and Cirrus OCT. In concordance with prior examinations based on Stratus OCT, we found a higher rate of algorithm failures in lesions containing fibrosis in all OCT examinations and in eyes with reduced distance acuity.\(^6\) Although the quality of the scans may be influenced by difficulties with fixation in decreased distance acuity, the overall quality of the scans was very good, even in eyes with reduced distance acuity. The pathologic features of the lesion (most of them containing fibrotic parts) were the major cause of algorithm line failures in eyes with low distance acuity. Misinterpretation of the anterior and posterior threshold algorithm line were less frequent in Cirrus OCT; however, improved algorithm line

**TABLE 5. Distribution of Algorithm Line Failures**

<table>
<thead>
<tr>
<th></th>
<th>Stratus OCT</th>
<th>Cirrus OCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior border misidentified</td>
<td>22/50.6</td>
<td>7/21.9%</td>
</tr>
<tr>
<td>Posterior border misidentified</td>
<td>4/5.6</td>
<td>12/57.5%</td>
</tr>
<tr>
<td>Anterior and posterior border misidentified</td>
<td>38/52.8</td>
<td>6/18.8%</td>
</tr>
<tr>
<td>Scan artefacts (blinking, scan off center)</td>
<td>8/11.1</td>
<td>7/21.9%</td>
</tr>
<tr>
<td>Central localization of failures (within a central circle of 1 mm)</td>
<td>62/59.6</td>
<td>19/18.3%</td>
</tr>
</tbody>
</table>

Data are the number/percentage of patients with failure in the respective regions.

![Image](http://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/932958/ on 02/24/2018)
The software version 3.0 of Cirrus OCT contains a built-in module to manually correct algorithm line failures. Although software options for the correction of Stratus OCT algorithm failures are available, they are not built in. The data have to be exported, and they are not supported by the manufacturer of the Stratus OCT. The error rates of this part of the study confirmed a significantly less frequent occurrence of threshold algorithm line failures in Cirrus OCT.

The significantly better detection of the border lines is attributable to the new spectral domain technology, which offers a series of advantages to the time domain technology of Stratus OCT: higher resolution, faster scan acquisition, and raster scanning of an area of 6 × 6 mm by 200 horizontal aligned B-scans each containing 200 A-scans in the cube 200 × 200 program, or by 128 horizontal aligned B-scans of higher resolution (containing 512 A-scans) in the cube 512 × 128 program. In contrast, Stratus OCT acquires 6 scans in a radial pattern. There is considerable “normalization” before setting the threshold algorithm lines leading to failures of measurement in Stratus OCT. Although there is vertical movement correction between the B-scans, there is no alignment normalization of the RPE inside each B-scan in Cirrus OCT (the differences of alignment are shown in the first line of Fig. 2). These additional failures of measurement were not part of this study.

CRT and retinal volume were calculated in both OCT systems, although different values were expected, because the positioning of the posterior (RPE) line is different in both OCT systems. In Stratus OCT, the posterior algorithm line is positioned on the surface of the RPE–choriocapillaris complex, probably in the area of the outer segments of the photoreceptors, whereas in Cirrus OCT the posterior line is positioned in the area of the RPE itself. The differences of the median values of CRT were 55.5 μm and of RV 3.12 mm³, which exceeded the values that Leung et al. found in the normal retina (difference of mean CRT 20.8). However, they compared Stratus OCT with a different spectral domain OCT, the Topcon 3D OCT (Topcon, Tokyo, Japan). In both studies the calculation of differences are based on a small number of examinations. Larger studies including eyes with different diseases are needed to arrive at a recommendation for the conversion of Stratus values.

In OCT we evaluated the amount of reflected light, and from our knowledge of the behavior of different kinds of tissue (reflectivity), we can determine from the reflectivity the anatomy or pathology of the retina. In exudative AMD the RPE, the neovascular tissue, and the fibrotic tissue form a highly reflective complex. Frequently, even experienced examiners cannot identify the RPE in this hyperreflective complex without any doubt. Therefore, even with improved technology, algorithm failures cannot be excluded completely. The experienced examiners evaluating the OCT scans in the present study were instructed to judge a scan to be correct, when they could not identify a more likely position of the RPE than the automatically set lines, if the position of the RPE could not be definitely determined. Only histologic examinations could provide more accuracy in detecting the RPE.

References


