Evaluation of Optical Coherence Tomography Retinal Thickness Parameters for Use in Clinical Trials for Neovascular Age-Related Macular Degeneration

Pearse A. Keane,1 Sandra Liakopoulos,1,2 Renu V. Jivrajka,1 Karen T. Chang,1 Tarek Alasil,1 Alexander C. Walsh,1 and Srinivas R. Sadda1

PURPOSE. To investigate the relationship between automated and manually derived measurements of central retinal thickness from optical coherence tomography (OCT) and to determine the relationship between the foveal center point (FCP) and the foveal central subfield (FCS) in neovascular age-related macular degeneration (AMD).

METHODS. Data were collected from 216 patients with newly diagnosed neovascular AMD, who underwent StratusOCT imaging at diagnosis. Raw StratusOCT images for each patient were analyzed with the publicly available custom software OCTOR, which allows accurate manual grading of OCT B-scans. Manually derived central retinal thickness measurements were compared with measurements obtained from automated StratusOCT analysis. Manually obtained measurements of FCP and FCS were also compared.

RESULTS. The mean (±SD) difference in thickness between automated and manually derived FCP thickness was 7.9 μm (±90.8), but the maximum difference was 455 μm. The limits of agreement (95% confidence interval), between automated and manually obtained FCP thicknesses, were −173.7 μm (lower limit) and 189.6 μm (upper limit), with a coefficient of determination (R2) of 0.49 (P < 0.001). In contrast, the R2 for manually derived FCP and manually derived FCS thickness was 0.95 (P < 0.001), with a smaller mean (±SD) difference in thickness of 13.8 μm (±29.8).

CONCLUSIONS. Manual correction of errors in automated OCT segmentation may be necessary for accurate interpretation of anatomic outcomes for clinical trials of neovascular AMD. In addition, although measurement of FCS remains preferable for assessment of central retinal thickness, accurate measurement of FCP may represent an adequate alternative when FCS is unavailable. (Invest Ophthalmol Vis Sci. 2009;50:3578–3585) DOI:10.1167/iovs.08-2728

Optical coherence tomography (OCT) plays an important role in the management of patients with age-related macular degeneration (AMD) by providing detailed cross-sectional images of the retina.1,2 StratusOCT (Carl Zeiss Meditec, Inc., Dublin, CA) includes image analysis software that automatically segments the inner and outer retinal boundaries and thus provides reproducible, quantitative information regarding neurosensory retinal thickness.3,4 Assessment of retinal morphology with other imaging modalities, such as color photography or fluorescein angiography, is associated with considerable variability, even among retinal specialists.5,6 Consequently, the measurement of central retinal thickness by OCT has been widely adopted as a secondary outcome parameter in clinical trials for the neovascular form of AMD.7,8

Despite the advantages offered by OCT over other, more well-established, imaging modalities, caution is still required in the interpretation of OCT outcomes, as significant errors are known to occur with the use of StratusOCT segmentation software.9,10 In addition, these errors occur more frequently in the context of neovascular AMD.11–13 In response, many clinical trials, for neovascular AMD and other disorders, use manual correction of retinal thickness measurements performed at dedicated image reading centers.8,14 Recent studies of diabetic macular edema (DME) suggest, however, that manual correction of errors in StratusOCT automated analysis may have no appreciable effect on the interpretation of anatomic outcomes for this condition.15 It remains unknown whether manual correction of StratusOCT errors is necessary for accurate interpretation of anatomic outcomes in clinical trials of neovascular AMD.

The question also arises as to the optimal measurement of central retinal thickness that should be used in clinical trials of neovascular AMD. Recent OCT studies have reported a variety of retinal thickness parameters, often under synonymous names, and there is no consensus regarding the relative value of these parameters.16,17 StratusOCT provides a measurement of average retinal thickness in the central 1-mm area of the fovea—the foveal central subfield (FCS)—that corresponds to Early Treatment Diabetic Retinopathy Study (ETDRS) subfield 9.18 StratusOCT also measures average retinal thickness at a single point at the center of the fovea—foveal center point (FCP).18 Studies for other disorders, such as DME, have shown a strong correlation between FCS and FCP, suggesting that FCP may be an adequate substitute for FCS—a potentially important finding, as manual correction of FCP is more straightforward than that of FCS.14,19 Unfortunately however, this relationship may not be valid in the context of neovascular AMD, where abrupt changes in the topography of the neurosensory retina may uncouple the association.

We have developed a custom software program, OCTOR, that enables accurate manual segmentation of OCT B-scans.18,20 In this study, we examined the relationship between automated and manually graded OCT outcomes in neovascular AMD and evaluated the role of central retinal thickness measurements for use in clinical trials of neovascular AMD.
**Materials and Methods**

**Data Collection**

OCT images from all patients \((n = 216)\) with newly diagnosed neovascular AMD at the Doheny Eye Institute, between January 2004 and December 2007, were collected and reviewed. Longitudinal data were also collected for the subset of the study population \((n = 47)\) that went on to receive treatment with intravitreal ranibizumab (Lucentis; Genentech, Inc., South San Francisco, CA) and that had StratusOCT imaging performed 3 months after treatment. Approval for data collection and analysis was obtained from the institutional review board of the University of Southern California. The research adhered to the tenets set forth in the Declaration of Helsinki.

StratusOCT images, made with the Radial Lines protocol of six high-resolution B-scans (512 A-scans per 6 mm B-scan) on a single StratusOCT machine, were collected. The Fast Macular Thickness protocol was used only when photographers were unable to obtain adequate high-resolution images, most commonly in patients with unstable fixation or poor cooperation.

**Automated StratusOCT Analysis Software**

OCT images from all patients were analyzed with StratusOCT Software Version 4.0. Measurement of retinal thickness by this software depends on identification of the internal limiting membrane and the inner hyperreflective band (now believed to correspond to the photoreceptor inner segment–outer segment [IS–OS] junction). The distance between the inner and outer retinal boundaries is then calculated across all sampled points, with interpolation of values in unsampled areas. The presence of morphologic features, such as fluid between the retina and the RPE, may confuse the segmentation algorithms and lead to inaccurate identification of the retinal boundaries. Retinal thickness maps generated by the StratusOCT software include average retinal thickness measurements for both FCS and FCP. In the StratusOCT automated output, FCP is expressed as the mean \(\pm SD\), obtained by averaging the center point thickness measurement from each of the six radial line B-scans. FCS is expressed as a single value, obtained by averaging all retinal thickness measurements in the central 1 mm of the fovea.

**Manual OCT Grading Software**

The OCTOR software used for OCT analysis was written by Doheny Image Reading Center software engineers to facilitate viewing and manual grading. OCTOR is publicly accessible at http://www.diesel.la and has been described and validated in previous reports. Univariate linear regression was also used to test for the correlation between automated and manually derived measurements of central retinal thickness (e.g., FCP). Mean differences between automated and manually derived measurements were calculated by using absolute values (as opposed to the simple mean, which could mask or minimize apparent differences). Relative (percentage) differences were calculated by dividing the value of the difference between the two measurements by the mean of the two measurements, and multiplying by 100. Bland-Altman plots were generated to facilitate comparisons between each method of central retinal thickness measurement. Similar calculations were performed for comparisons between automated measurements and the manually derived, combined retinal+subretinal fluid parameters.

After the grader draws the required layers in each of the 6 B-scans, the software calculates the distance in pixels between the manually drawn boundary lines for each of the various defined spaces. Using the dimensions of the B-scan image, the calculated pixels are converted into micrometers to yield a thickness measurement at each location. The thickness at all unsampled locations between the radial lines is then interpolated based on a polar approximation to yield a thickness map analogous to the StratusOCT output data. The interpolation algorithm, intergrader reliability, and intragrader reproducibility have been validated. Analogous to the StratusOCT software, OCTOR provides a report showing the calculated thicknesses for the nine Early Treatment of Diabetic Retinopathy Study macular subfields. The mean \(\pm SD\) of the FCP thickness is also calculated.

**Manual Grading Procedure**

OCT scans were analyzed by certified OCT graders at the Doheny Image Reading Center (PAK, SL, and KTC); all graders were masked to the results of automated image analysis. The inner retinal boundary (internal limiting membrane) and the outer retinal boundary (outer border of the photoreceptors) were drawn in each of the six OCT B-scans obtained for each eye. Both boundaries were drawn in accordance with the standard OCT grading protocol of the Doheny Image Reading Center. All OCT scans included in the study met reading center criteria for sufficient image quality, including the absence of significant artifactitious variations in signal intensity or generalized reductions in signal strength. No minimum value for signal strength was set, as manual grading with OCTOR often allows quantitative information to be accurately derived from images with low signal strength. To facilitate direct comparison with automated measurements and, in part, because of the complex morphology of neovascular AMD, we did not correct manual measurements for scan decentration. After completion of grading, OCTOR was used to calculate output parameters for the neurosensory retina: FCP thickness and FCS thickness. In addition, a combined thickness parameter (retina+subretinal fluid) was calculated for both the FCP and FCS. This combined parameter was calculated, as the automated StratusOCT algorithms often include areas of subretinal fluid in retinal thickness measurements.

**Statistical Methods**

Univariate linear regression was used to investigate the correlation between automated and manually derived measurements of central retinal thickness (e.g., FCP). Mean differences between automated and manually derived measurements were calculated by using absolute values (as opposed to the simple mean, which could mask or minimize apparent differences). Relative (percentage) differences were calculated by dividing the value of the difference between the two measurements by the mean of the two measurements, and multiplying by 100. Bland-Altman plots were generated to facilitate comparisons between each method of central retinal thickness measurement. Similar calculations were performed for comparisons between automated measurements and the manually derived, combined retinal+subretinal fluid parameters.

Linear regression was also used to test for the correlation between FCP thickness and FCS thickness by using values derived from manual grading with OCTOR software. Using this analysis, a regression equa-
tion was derived allowing a value for FCS to be imputed from manually graded FCP thickness. Mean differences between manually derived FCP and FCS values were calculated with absolute values, as previously described.

\[ P < 0.05 \] was considered statistically significant. Statistical analysis and graph generation was performed by using commercially available software (Intercooled Stata for Windows, ver. 9; Statacorp LP, College Station, TX). Data are expressed as the mean (±SD).

**RESULTS**

**Patient Enrollment and Baseline Characteristics**

Two hundred sixteen patients (216 eyes), presenting to the Doheny Eye Institute between January 2004 and December 2007, received a new diagnosis of CNV secondary to AMD and had StratusOCT imaging performed at the time of initial diagnosis. Seven patients were excluded, because although StratusOCT had been performed at the time of diagnosis, the raw StratusOCT data could not be located or exported. Three patients were excluded from the study, as their OCT images were of insufficient quality to permit grading. Therefore, in total, raw exported StratusOCT images were obtained and graded in 206 (of 216) patients. Of the 206 patients included in our analysis, 122 (59%) were women, and 84 (41%) were men. The mean age of the patients was 81 years (±7); the median age was also 81 years (range, 58–96).

**Automated versus Manually Derived OCT Outcomes for FCP**

The coefficient of determination (\( R^2 \)) between automated (StratusOCT) and manually derived (ORCTOR) FCP thickness was 0.49 (Fig. 1). Figure 2A shows the distribution of the differences between automated and manually derived FCP thickness. The mean absolute difference in thickness between automated and manually derived FCP thickness was 59.2 \( \mu \)m (±69.2). The maximum difference in thickness between automated and manually derived FCP thickness was 455 \( \mu \)m. Figure 2B shows the distribution of the relative differences between...
automated and manually derived FCP thickness. The mean relative difference between automated and manually derived FCP thickness was 19.6% (±21.3%). The maximum difference between automated and manually derived FCP thickness was 124.1%. In 118 (57.3%) cases, the difference between automated and manually derived FCP thickness was ≤10%, in 66 (32.0%) the difference was >10% to ≤20%, and in 39 (18.9%), the difference was >20%. As with FCP thickness, the differences between automated and manually derived FCS thickness are illustrated with histograms and Bland-Altman plots (Fig. 3).

**Comparison of Automated Outcomes with Combined Retinal and Subretinal Fluid Parameters**

For the FCP, the mean absolute difference in thickness between automated StratusOCT measurement and manually derived, combined retinal+subretinal fluid measurement was 46.7 μm (±65.6). The mean relative difference between automated and manually derived retinal+subretinal fluid FCP thickness was 14.6% (±18.1%).

For the FCS, the mean absolute difference in thickness between automated StratusOCT measurement and manually derived, combined retinal+subretinal fluid measurement was 41.1 mm (±55.0). The mean relative difference between automated and manually derived retinal+subretinal fluid FCS thickness was 12.2% (±14.7%).
Effect of StratusOCT Scanning Protocol on Automated and Manually Derived Central Retinal Thickness Measurements

One hundred forty-four (70%) eyes were imaged on StratusOCT with the Radial Lines scanning protocol, and 62 eyes (30%) were imaged on StratusOCT using the Fast Macular Thickness scan protocol. The mean absolute difference in thickness between automated and manually derived FCP thickness was 75.1 mm (±66.5) for eyes imaged with the Fast Macular Thickness scanning protocol, and 52.4 mm (±69.4) for eyes imaged with the Radial Lines protocol (P = 0.0025, Mann-Whitney rank sum test). The mean absolute difference in thickness between automated and manually derived FCS thickness was 63.1 mm (±56.8) for eyes imaged with the Fast Macular Thickness scanning protocol, and 46.7 mm (±61.7) for eyes imaged with the Radial Lines protocol (P = 0.0038, Mann-Whitney rank sum test).

Evaluation of Changes in Automated and Manually Derived Central Retinal Thickness Measurements over Time

Forty-seven patients, receiving treatment with intravitreous ranibizumab, had StratusOCT images available for analysis after 3 months of follow-up. An example of segmentation error at both baseline and follow-up in a single patient is shown in Figure 4. The mean change in manually derived FCP thickness between baseline and month 3 was −45.70 mm (±99.77). The mean change in automated FCP thickness between baseline and month 3 was −68.07 mm (±107.48) (P = 0.0935, Wilcoxon signed rank test). The mean change in manually derived FCP thickness between baseline and month 3 was −41.72 mm (±82.19) mm. The mean change in automated FCP thickness between baseline and month 3 was −63.36 mm (±95.11) (P = 0.0154, Wilcoxon signed rank test). Changes in automated and manually derived central retinal thickness measurements over time are depicted in Bland-Altman plots (Fig. 5).

Manually Derived OCT Outcomes for Central Retinal Thickness

The coefficient of determination ($R^2$) between manually derived FCP and FCS thicknesses was 0.94, and the following regression equation was obtained: $FCS = 0.86 \cdot FCP + 57.2$ (Fig. 6). Figure 7A shows the distribution of the absolute differences between FCP thickness and FCS thickness. The mean absolute difference in thickness between FCP and FCS was 13.8 μm (±29.8). The maximum difference in thickness between manually derived FCS and FCP was 110 μm. Figure 7B shows the distribution of the relative differences between FCP thickness and FCS thickness. The mean relative difference between FCS and FCP thicknesses was 6.9% (±12.6%). The maximum difference between FCS and FCP thicknesses was 67.1%.

**DISCUSSION**

In this retrospective study, we performed manual grading of OCT images to examine the association between automated and manually derived measurements of central retinal thickness and to elucidate the relationship between FCP and FCS, for use in clinical trials of neovascular AMD.

In this report, manually derived measurements of central retinal thickness differed from corresponding automated measurements in a manner likely to be of clinical significance. Bland-Altman plots comparing automated and manually derived measurements of central retinal thickness demonstrated wide upper and lower limits of agreement: −173.7 to 189.6 μm in the case of FCP (Fig. 2C). Moreover, the relative difference in thickness between methods was greater than 10% in a large number of cases (Fig. 2D). Similar discrepancies between automated and manually derived measurements were seen when changes in central retinal thickness were evaluated over time (Fig. 5). These differences may explain, at least in part, why clinical trials have failed to find a consistent correlation between changes in central retinal thickness and improvements in visual acuity.7,21,22 The magnitude of errors in automated OCT segmentation may also have implications for the conduct of clinical trials: Many of these trials use absolute changes in central retinal thickness on OCT as a means of guiding retreatment protocols in variable dosage regimens (Meyer CH, et al. IOVS 2008;49:ARVO E-Abstract 273).7 Furthermore, many clinical trials have minimum central retinal thickness levels as eligibility requirements for inclusion, and these assessments are made using automated OCT measurements. It is clear that failure to account for the discrepancy between automated and manually derived measurements of central retinal thickness could undermine our ability to accurately interpret anatomic outcomes in neovascular AMD. The discrepancy between automated and manually derived mea-
measurements was also significantly greater for images obtained with the Fast Macular Thickness scanning protocol, suggesting that use of the higher resolution Radial Lines protocol may be preferable for the provision of quantitative data in clinical trials.

Several factors account for the differences between automated and manually derived central retinal thickness measurements in neovascular AMD. In normal eyes, StratusOCT measures the thickness of the neurosensory retina from the internal limiting membrane to the photoreceptor IS-OS junction, thus underestimating retinal thickness by approximately 30 to 35 μm. The situation is more complex for neovascular AMD, however, as the IS–OS junction is often not easily seen, either due to disruption of the outer retina by the choroidal neovascular process or due to inadequacies of the StratusOCT axial resolution. StratusOCT automated software also consistently overestimates central retinal thickness when subretinal fluid is present by using the inner border of the RPE as the outer retinal boundary. To address this problem and to allow the fairest possible comparison between manual grading and automated StratusOCT segmentation, we repeated our analyses employing a combined retinal and subretinal fluid thickness parameter. With this correction, large overestimations of central retinal thickness were less commonly seen. However, significant differences between automated and manually derived central retinal thicknesses continued to exist. As well as overestimation, StratusOCT automated software commonly underestimates central retinal thickness in the context of intraretinal fluid collection, where the automated software often has trouble in accurate segmentation of the inner retinal boundary.

In this report, we also examined the relationship between FCP and FCS in patients with neovascular AMD. We hypothesized that, due to the complex retinal topography seen in neovascular AMD, FCS would correlate closely with FCP. Contrary to our hypothesis, FCS, in fact, correlated highly with FCP in our study, suggesting that FCP may be an adequate substitute for FCS in clinical trials for neovascular AMD—a potential advantage for these trials, as accurate measurement of FCP is more easily performed than that of FCS. Manual grading of FCS requires use of specialized programs such as OCTOR, or the recently released StratusOCT software version 5.0, which is not widely available at present. Caution must be used in making this assumption in studies of neovascular AMD with small sample sizes; however, as we have shown that large differences between FCP and FCS still commonly occur, these differences...
may confound the interpretation of results when FCP alone is used.

In this study, we derived a regression equation describing the relationship between FCS and FCP (FCS = 0.86 \cdot FCP + 57.2). When available, we believe the use of FCS is still preferable to that of FCP. FCS may be less prone to decentration than FCP, as it covers a 1-mm-diameter area at the center of the fovea, rather than a single point. In line with this difference, studies of neovascular AMD and other disorders have shown greater reproducibility for FCS than FCP. Furthermore, as FCS includes contributions from the parfoveal region, it may be a better indicator of near visual function than is FCP (e.g., reading speed). Use of a regression equation in clinical trials allows FCS values to be imputed from manually obtained FCP values in cases in which accurate measurement of the FCS parameter is unavailable. For example, in a recent large clinical trial of treatments for diabetic macular edema, manually derived FCP thickness measurements were used to impute data for FCS in 18% of scans.28

Our study has several strengths: in particular, the large sample size (in comparison to previous studies of OCT imaging in neovascular AMD) and the use of manual grading with specialized software, performed in a dedicated OCT image reading center. We have previously shown that analysis of OCT images with OCTOR software allows for highly reproducible measurements of FCP thickness (intraclass correlation coefficient, 0.99; weighted κ, 0.88).20 From these results, and our accumulated experience in the Doheny Image Reading Center, we believe that the accuracy of manual segmentation by trained human graders is such that it could serve as a gold standard against which to compare the results of automated analysis.

Our study also has several limitations. First, the conclusions drawn from this study may be technology-dependent. Although StratusOCT provides excellent axial resolution, it is limited by a low scanning speed and is thus dependent on interpolating algorithms to provide retinal thickness maps. In the future, clinical trials for neovascular AMD are likely to involve higher speed Fourier Domain OCT (FDOCT) that confers greater resistance to fixation errors and permits a higher sampling density of the scanned region, thus requiring less interpolation between sampled locations.29,30 Unfortunately, automated segmentation algorithms in FDOCT devices have yet to be validated for complex diseases such as neovascular AMD, and grading every B-scan in a dense FDOCT volume acquisition for a large number of patients is not feasible. Manual correction of a subset of scans appears to be a reasonable compromise until image analysis software associated this technology matures. Ideally, a set of easily derived image parameters could be identified that preferentially select cases or scans with large segmentation errors for reading-center manual correction. We have not been able to identify such parameters thus far, but this should be an important goal for future studies. Second, our analysis was restricted to the evaluation of central retinal thickness parameters. Even in cases with well-centered foveal B-scans, and accurate boundary detection, the use of these parameters may be erroneous, because of the presence of eccentrically positioned neovascular lesions. Paracentral subfields and total retinal volume can be obtained directly from the OCT Retinal Map Analysis and may be preferable to the calculation of thickness at a single point or subfield for juxta- or extrafoveal lesions.

In conclusion, despite the results of recent clinical trials of DME, it appears that manual correction of errors in automated OCT segmentation should still be a requirement for clinical trials of neovascular AMD. In addition, although measurement of FCS remains preferable for assessment of central retinal thickness, accurate measurement of FCP may represent an adequate alternative when FCS is unavailable.

References


