Article

Reading Center Characterization of Central Retinal Vein Occlusion Using Optical Coherence Tomography During the COPERNICUS Trial

Francis Char DeCroos^{1,2}, Sandra S. Stinnett¹, Cynthia S. Heydary¹, Russell E. Burns¹, and Glenn J. Jaffe¹ ✉

¹ Duke University Eye Center, Duke University, Durham, NC

² Wills Eye Institute/Mid Atlantic Retina, Philadelphia, PA

Correspondence: Glenn J. Jaffe, MD, Duke Eye Center, Director, Duke Reading Center, DUMC Box 3802, Durham, NC 27710, USA. e-mail: jaffe001@mc.duke.edu

Received: 1 July 2013

Accepted: 3 November 2013

Published: 20 December 2013

Keywords: optical coherence tomography; central vein occlusion; clinical trial; aflibercept

Citation: DeCroos FC, Stinnett SS, Heydary CS, Burns RE, Jaffe GJ. Reading center characterization of central retinal vein occlusion using optical coherence tomography during the COPERNICUS trial. *Trans Vis Sci Tech.* 2013;2(7):7, <http://tvstjournal.org/doi/full/10.1167/tvst.2.7.7>, doi:10.1167/tvst.2.7.7

Purpose: To determine the impact of segmentation error correction and precision of standardized grading of time domain optical coherence tomography (OCT) scans obtained during an interventional study for macular edema secondary to central retinal vein occlusion (CRVO).

Methods: A reading center team of two readers and a senior reader evaluated 1199 OCT scans. Manual segmentation error correction (SEC) was performed. The frequency of SEC, resulting change in central retinal thickness after SEC, and reproducibility of SEC were quantified. Optical coherence tomography characteristics associated with the need for SECs were determined. Reading center teams graded all scans, and the reproducibility of this evaluation for scan quality at the fovea and cystoid macular edema was determined on 97 scans.

Results: Segmentation errors were observed in 360 (30.0%) scans, of which 312 were interpretable. On these 312 scans, the mean machine-generated central subfield thickness (CST) was $507.4 \pm 208.5 \mu\text{m}$ compared to $583.0 \pm 266.2 \mu\text{m}$ after SEC. Segmentation error correction resulted in a mean absolute CST correction of $81.3 \pm 162.0 \mu\text{m}$ from baseline uncorrected CST. Segmentation error correction was highly reproducible (intraclass correlation coefficient [ICC] = 0.99–1.00). Epiretinal membrane (odds ratio [OR] = 2.3, $P < 0.0001$), subretinal fluid (OR = 2.1, $P = 0.0005$), and increasing CST (OR = 1.6 per 100- μm increase, $P < 0.001$) were associated with need for SEC. Reading center teams reproducibly graded scan quality at the fovea (87% agreement, kappa = 0.64, 95% confidence interval [CI] 0.45–0.82) and cystoid macular edema (92% agreement, kappa = 0.84, 95% CI 0.74–0.94).

Conclusions: Optical coherence tomography images obtained during an interventional CRVO treatment trial can be reproducibly graded. Segmentation errors can cause clinically meaningful deviation in central retinal thickness measurements; however, these errors can be corrected reproducibly in a reading center setting.

Translational Relevance: Segmentation errors are common on these images, can cause clinically meaningful errors in central retinal thickness measurement, and can be corrected reproducibly in a reading center setting.

Introduction

Optical coherence tomography (OCT) has proven instrumental to objectively characterize cystoid macular edema (CME) from central retinal vein occlusion (CRVO).^{1,2} This imaging modality can quantify retinal thickness changes in eyes with CME,¹ and is

superior to contact lens-assisted biomicroscopy to identify foveal edema.³ Macular edema is a major contributor to vision loss in patients with CRVO, and OCT has been especially useful to monitor macular edema treatment.^{4–8} In addition to characterizing CME, OCT can facilitate identification of associated visually significant pathologies such as subretinal fluid (SRF),⁹ epiretinal membrane (ERM),¹⁰ and

vitreomacular adhesion (VMA).¹¹ For these reasons, OCT has been used to monitor eyes of patients who participated in recent interventional trials for CRVO. For example, OCT has been used for this purpose in the Central Retinal Vein Occlusion Study (CRUISE), Global Evaluation of Implantable Dexamethasone in Retinal Vein Occlusion with Macular Edema study (OZURDEX GENEVA), and Standard Care vs Corticosteroid for Retinal Vein Occlusion (SCORE) study.⁴⁻⁶ Though OCT has been used in these and other clinical trials for CRVO, the OCT grading methodologies vary across studies. For example, in the SCORE trial, OCT images were typically evaluated by a single reader in contrast to the team-based approach used at our reading center as described below.

To accurately measure macular edema secondary to CRVO, it is important that central retinal thickness measurements be reproducible. Though this goal is straightforward, it is not always so simple to obtain these measurements. For example, during the SCORE trial analysis of central retinal thickness, it was noted that 28.9% of OCT scans required remeasurement with handheld calipers. The most frequent indication for these manual corrections was automated segmentation error.⁵ Other groups have reported that inaccurate automated retinal thickness measurements from OCT software segmentation algorithms are common.¹²⁻¹⁵ In one study that reviewed manual remeasurement in over 2000 OCT scans,¹⁴ incorrect automated segmentation placement was noted as the most common indication for segmentation error correction.

Segmentation errors are frequent, and the magnitude of central retinal thickness measurement error caused by misplaced segmentation lines in eyes with CRVO can be substantial. A smaller study of 28 eyes with retinal vein occlusion reported a 129- μ m error in automated mean retinal center point measurement.¹⁴ This study, however, did not examine the precision of measurement corrections performed. Also, particular CRVO-associated morphological characteristics on OCT may be associated with higher rates of segmentation error corrections (SEC). If these OCT factors were known and identified in an eye with CRVO, then reading center readers and managing clinicians could better predict when segmentation errors would occur on automated measurements. Finally, reproducible measurement of change in central retinal thickness is particularly important since this parameter is a key secondary endpoint of

interventional trials for CRVO such as the COPENICUS trial described in the present report.

In the current study, we describe for the first time, to the best of our knowledge, the reproducibility of a team-based approach to grade OCT images generated during an interventional, multicenter trial of CRVO-associated macular edema and the reproducibility of SEC. The impact of segmentation errors on automated measurement of central retinal thickness and OCT morphological findings associated with the need for SEC were also determined.

Materials and Methods

Trial Overview

The COPENICUS study was a randomized, multicenter, phase 3 trial that investigated the safety and efficacy of aflibercept (Regeneron, Tarrytown, NY) compared to placebo as a treatment for visual acuity loss from CRVO-associated macular edema (ClinicalTrials.gov, ID NCT00943072).¹⁶ This trial includes eyes with CRVO complicated by center-involving macular edema greater than 250 μ m in thickness and visual acuity between 20/40 and 20/320. For the duration of the trial, eyes were followed with serial Stratus OCT (Carl Zeiss Meditech, Dublin, CA) scans at each study visit. Spectral domain OCT was not universally available at study sites during the COPENICUS trial. All experimental procedures adhered to the tenets of the Declaration of Helsinki; appropriate institutional review board approval was obtained, and all participants engaged in an informed consent process and signed a written consent document prior to enrollment in the COPENICUS trial.

Standardized OCT Acquisition

All study scans were acquired by certified OCT technicians using Stratus OCT machines with software version 4.0 or greater. Prior to granting certification, the OCT technician training process required submission of 16 OCT scans, at least 4 of which had macular pathology. The Duke Reading Center provided feedback to OCT technicians on these images regarding scan placement and avoidable artifacts. The training protocol also emphasized appropriate focus, scan saturation, line length, and line placement while acquiring OCT images. Feedback was provided to all OCT technicians for each graded scan through an automated system that reported scan placement and quality and also

Figure 1. (a) Stratus OCT radial image showing the 125- μ m on-screen horizontal distance between *blue* caliper tips that defines minimum threshold for horizontal segmentation error. (b) OCT radial image showing the 125- μ m on-screen vertical distance between *blue* caliper tips that defines minimum threshold for vertical segmentation error. OCT images were acquired using FMTM protocol, and automated segmentations are *white*. All segmentation errors were measured with on-screen calipers built into the Stratus software.

identified individual scans of concern for the duration of the trial.

All patients were imaged with both the Stratus fast macular thickness map (FMTM) and macular thickness map (MTM) protocols. The FMTM was used to determine central subfield thickness (CST), as scan centration was facilitated by the rapid scan acquisition time. The MTM was used to assess morphological features. In cases in which the FMTM quality was inadequate, the MTM was substituted to determine CST. Each of these protocols included six radial lines of 6-mm length placed across the fovea center at 30°

rotational increments. Scans were deidentified, labeled with a unique code, and then submitted to the reading center. Scans were submitted for the following study visits: screening, baseline, week 4, week 8, week 12, week 16, week 20, week 24, week 36, and week 52 after baseline.

OCT Grading

Certified readers at the Duke Reading Center reviewed all 1199 study-eye OCT scans. No scans from fellow eyes were included. For this study, the fovea on a 6-mm radial OCT image was defined as the

Figure 2. Stratus OCT surface map demonstrating retinal thickness and total macular volume before (a) and after (b) manual segmentation corrections were performed for the 125- μ m errors in segmentation line placement shown in Figure 1.

Figure 3. (a) OCT retinal thickness map demonstrating inner segmentation line error (*white arrow*). (b) Same OCT retinal thickness map after manual segmentation line correction has been performed, resulting in 122- μ m change in central subfield thickness (*red arrows*). OCT images were acquired using FMTM protocol, and segmentation lines are *white*.

horizontal region 1 mm in width with midpoint at the foveal center. First, OCT scan quality was verified on each scan at the foveal center point, the fovea, and then the remainder of each radial image. Scan quality at the fovea was classified as “interpretable” if the appropriate scan region was available, correctly positioned, and adequately saturated. The designation “not interpretable” was applied if one or more radial line images were missing, incorrectly positioned, or very poorly saturated. Any observed segmentation error required correction as described below.

Segmentation errors, defined as any vertical or horizontal deviation 125 μ m or more from the expected position at either the internal limiting membrane (ILM) or the photoreceptor inner segment–outer segment (IS-OS) junction, were identified on each interpretable scan on-screen ($n = 984$; Figs. 1, 2). When the OCT IS-OS junction could not be clearly differentiated from the RPE, automated segmentation of the outer retina was performed at the first hyperreflective border adjacent to the RPE. All segmentation errors were measured using on-screen calipers built into the Stratus software. An on-screen

Figure 4. OCT surface map demonstrating nine ETDRS subfields for automated calculation of retinal thickness and volume for the right (a) and left (b) eye. TIM, temporal inner macula; SIM, superior inner macula; NIM, nasal inner macula; IIM, inferior inner macula; TOM, temporal outer macula; SOM, superior outer macula; NOM, nasal outer macula; IOM, inferior outer macula.

deviation of 125 μm was chosen empirically as a minimum threshold that was a large enough value that readers could detect consistently during grading, yet was minimal enough to allow corrections of relatively small measurement deviations. Segmentation errors were manually corrected on individual radial line images using the retinal map function. Figure 3 demonstrates a representative segmentation error before and after correction. Segmentation error corrections were not performed if scan quality at the fovea was graded as “not interpretable” (48 of 360, 13.3%). Since readers could not accurately perform SEC with confidence on scans of “not interpretable” quality, automated retinal thickness measurements from these scans were likewise not determined. Rather, for the 48 scans of “not interpretable” quality, if possible, the foveal center point thickness (CPT) was measured manually with software-based calipers for all individual radial line images depicting the fovea.

Several automated retinal thickness measurements, including CPT, CST, and total macular volume (TMV), were automatically generated from FMTM scans using commercial Stratus software after any erroneous segmentation lines had been corrected. If the FMTM scan quality at the fovea was not interpretable, automated retinal measurements from an interpretable MTM were substituted, if available. Software-generated CPT was the distance between the

ILM and the photoreceptor IS-OS junction at the intersection of the six radial lines at the foveal center point. The circular map was divided into nine subfields bounded by concentric rings of 1-, 3.45-, and 6-mm radius as specified by the Early Treatment Diabetic Retinopathy Study (ETDRS).¹⁷ Mean subfield retinal thickness was automatically generated via software from the previously obtained six radial scans. Thickness for the circular central subfield was termed “central subfield thickness.” Similarly, the surrounding eight retinal thickness measurements were designated as inner and outer subfields of each quadrant (Fig. 4). Volume for each subfield equaled the mean retinal thickness multiplied by the surface area of the segment. Total macular volume equaled the sum of these nine subfield volumes. Although many retinal characteristics were quantified from OCT for this trial, the endpoint of interest for OCT analysis was change in CST.

Each MTM was analyzed for the following morphological features: CME, SRF, ERM, and vitreomacular attachment (VMA) (Fig. 5). The grading variables were included in a model to predict the need for SEC as described below. For CME, SRF, ERM, and VMA, one of the following grading criteria was assigned: feature present, feature absent, or not interpretable (due to scan quality, incorrect placement, or absent scan). If CME was present in the fovea, the vertical (width) and horizontal dimensions

Figure 5. Representative morphologic features noted on OCT images derived from the MTM protocol: (a) cystoid macular edema (*white arrow*), (b) epiretinal membrane (*white arrow*) and subretinal fluid (*red arrow*), (c) vitreomacular adhesion (*white arrow*) and subretinal fluid (*red arrow*).

(thickness) of the largest cyst were measured from a single radial line scan using software-based calipers. Standardized photographs were assigned to reference particular morphological characteristics. Hyporeflective spaces that are formed between the ERM and ILM are not located within the retina, and were not considered CME. During scan appraisal, readers

evaluated each of the six MTM images prior to assigning any grading values.

Two masked readers first independently graded all scans. A separate data transcriptionist then identified discrepant values between the paired readers for any continuous or categorical or grading variables. In particular, caliper-based measurements were consid-

Figure 6. OCT scan work flow for temporal reproducibility analysis of hypothetical individuals, senior reader A and senior reader B, demonstrating both intra-arbitrator repeatability and interarbitrator agreement studies. *Solid lines*, initial grading; *dashed lines*, reproducibility analysis.

ered discrepant if the readers' on-screen measurements differed by more than 25 μm from one another. All graded scan pairs with discrepant data were then presented to a senior reader who reconciled discrepancies between the initial pair of graded scans and recorded his or her final grade. Any finding or value that remained controversial after arbitration was forwarded to the director of grading or the reading center director for a final decision.

A group of two primary readers and one senior reader composed a reading center team. To create each team, two readers were chosen at random from a pool of five readers and were matched with one of two senior readers. To ensure consistency across reading center teams, ongoing monthly meetings were held to make sure readers adhered to grading protocols, to address common grading discrepancies, and to review

particularly challenging scans. Readers, senior readers, the director of grading, and the reading center director typically attended each meeting.

Segmentation Error Correction Reproducibility

All reading center members participating in any reproducibility study were masked to the results of the initial grading. First, an analysis was performed to determine SEC reproducibility. From the first 1199 consecutive study scans transferred to the reading center, 360 scans had segmentation errors. Of these 360 scans, 312 were interpretable, and subsequent SEC was then performed on these 312 scans. On the remaining 48 "not interpretable" scans, one or more radial line images were missing,

Figure 7. OCT scan work flow for temporal reproducibility analysis of hypothetical reading center team C and team D demonstrating both interteam agreement studies. *Solid lines*, initial grading; *dashed lines*, reproducibility analysis. A reading team was composed of a senior reader and a pair of readers who reviewed a particular scan. A repeat grading team for agreement studies was defined as consisting of both a senior reader and at least one reader different from the team that performed the original grading.

incorrectly positioned, or very poorly saturated, so SEC was not performed. For the reproducibility analysis, 25 scans were randomly selected from the 312 scans for which segmentation errors had been corrected. These images encompassed 25 different scans sampled from 20 patients. The same reader who had performed the initial SEC performed a repeat SEC. This reader knew that all scans required repeat SEC, but was masked to prior placement of segmentation lines. After repeat SEC, automated retinal thickness and volume were again computed for each scan and compared to the values generated after initial SEC. Mean absolute retinal CPT measurement error was calculated as the mean of the absolute value of differences in retinal CPT prior to and following SEC. A similar calculation was

performed to determine mean absolute CST measurement error.

OCT Morphological Feature Grading Reproducibility

Intra-arbitrator, interarbitrator, and interteam analysis was performed on 97 scans randomly selected from the first 1199 consecutive scans transferred to the reading center. These scans encompassed study visits between screening and week 36. All reproducibility testing was performed 5 weeks or more after initial grading to avoid possible reader recall of the initial arbitration¹⁸ and to monitor temporal variability. For intra-arbitrator reproducibility analysis, a senior reader received 48 or 49 pairs of graded scans that he or she had previously arbitrated. Discrepant

Figure 8. Bland-Altman plot for 312 automated CPT measurements generated before and after manual segmentation error correction. The solid line and dashed lines indicate mean and 95% limits of agreement, respectively.

values within these scan pairs were arbitrated in a standard fashion, and the results for the repeat arbitration were compared to the initial arbitration (Fig. 6). To evaluate interarbitrator agreement, each senior reader arbitrated 48 or 49 pairs of graded scans that he or she had not previously arbitrated. Discrepant values within these scan pairs were arbitrated in a standard fashion, and the results for the second arbitration were compared to those from the first arbitration (Fig. 6). To measure interteam agreement, one team graded scans initially, and then a repeat grading team, defined as consisting of at least one different reader and a different arbitrator from the initial grading team, repeated the grading arbitration process (Fig. 7). All reproducibility analyses mentioned above were performed for CME and scan quality at the fovea (horizontal region 1 mm in width with midpoint at the foveal center). Cystoid macular edema and scan quality at the fovea were selected to facilitate comparison to the SCORE trial, in which these variables were used to quantitatively evaluate OCT grading reproducibility.

Statistical Analysis

Calculations were performed to determine the appropriate sample size to compare the difference in mean CST during the SEC reproducibility evaluation. A sample size of 13 would have 90% power to detect a difference in means of 25 μm , assuming a standard deviation of differences of 25 μm , using a paired *t*-test with a 0.05 two-sided significance level. Any scan pairs demonstrating a disparity of 25 μm or more in vertical dimension required arbitration per reading center grading protocol. For scans requiring SEC, absolute differences and paired differences for automated retinal thickness measurements were computed. The significance of the paired difference was assessed with the Wilcoxon signed rank test of median difference equal to zero. A two-sided *P* value less than 0.05 was considered significant. In addition, the intraclass correlation with 95% confidence intervals was used to assess agreement. Bland-Altman plots were constructed to investigate any trend for differing

Automated Central Subfield Thickness

Figure 9. Bland-Altman plot for 312 automated CST measurements generated before and after manual segmentation error correction. The solid line and dashed lines indicate mean and 95% limits of agreement, respectively.

CPT and CST variability after correction across a range of macular thicknesses.¹⁹

The percent agreement and the kappa statistic²⁰ were calculated as measures of agreement for categorical variables (quality of scan and presence of CME). Percent agreement was computed as the number of concordant grading pairs divided by the total number of grading pairs multiplied by 100.

Multiple logistic regression analysis was performed on covariates to determine association with SEC for all interpretable scans ($n = 984$). The morphological features on OCT tested as potential covariates were CME, CME at the fovea, maximal CME cyst width, SRF, SRF at the fovea, VMA, VMA at the fovea, ERM, ERM at the fovea, and CST. To avoid multicollinearity, CPT and maximal CME cyst thickness were not included in the model. The stepwise selection method was performed to select covariates. Assumptions for logistic regression were met. The profile likelihood ratio method was used with a two-tailed P value of 0.05. Interaction and goodness-of-fit diagnostics were also performed. No cases with missing data were identified. All analyses were

performed with SAS 9.2 software (SAS Institute, Inc., Cary, NC).

Results

Segmentation Error Correction Reproducibility

Segmentation errors were corrected to determine the impact of segmentation errors on clinically pertinent central retinal thickness measurements. Prior to SEC, the mean machine-generated values for CPT and CST were 499.0 ± 228.3 and 507.4 ± 208.5 μm , respectively. After SEC, corrected values for mean CPT and mean CST were 592.0 ± 294.0 and 583.0 ± 266.2 μm , respectively. For the 312 scans that required SEC, the mean absolute central retinal thickness errors were 102.3 ± 186.0 and 81.3 ± 162.0 μm for CPT and CST, respectively. Figures 8 and 9 show Bland-Altman plots for the CPT and CST, respectively. These plots demonstrate that the magnitude of the segmentation error-induced CPT and CST errors increase with greater retinal thickness.

Table 1. Intraclass Correlations for Automated Retinal Thickness and Volume Measurements From Repeat Segmentation Error Correction (SEC) Performed on 25 OCT Scans That Initially Required SEC

Variable	N	Intraclass Correlation	Lower Confidence Limit	Upper Confidence Limit
Central foveal thickness	25	1.00	1.00	1.00
Central subfield thickness	25	1.00	1.00	1.00
Nasal inner macula	25	1.00	0.99	1.00
Nasal outer macula	25	0.99	0.99	1.00
Superior inner macula	25	1.00	1.00	1.00
Superior outer macula	25	1.00	1.00	1.00
Temporal inner macula	25	1.00	1.00	1.00
Temporal outer macula	25	1.00	1.00	1.00
Total macular volume	25	1.00	1.00	1.00

Central foveal thickness, thickness at foveal center point; CST, mean thickness across entire ETDRS central subfield.

Automated retinal thickness and volume calculations were highly reproducible after repeat SEC. For CPT, the interclass correlation was 1.00 (95% confidence interval [CI] 1.00–1.00, Table 1), and the mean absolute thickness difference was $3.7 \pm 5.7 \mu\text{m}$ (Table 2). For CST, the interclass correlation was 1.00 (95% CI 1.00–1.00), and the mean absolute thickness difference was $2.2 \pm 2.0 \mu\text{m}$. For TMV, the interclass correlation was 1.00 (95% CI 1.00–1.00), and the mean absolute volume difference was $0.1 \pm 0.1 \text{ mm}^3$. The interclass correlations, mean paired differences, and mean absolute differences for the eight peripheral ETDRS subfields after repeat SEC are shown in Tables 1 and 2.

OCT Morphological Features Associated With Need for Segmentation Error Correction

After stepwise selection of covariates, on adjusted analysis, ERM (odds ratio [OR] = 2.3, 95% CI 1.7–3.3, $P < 0.0001$), SRF (OR = 2.1, 95% CI 1.4–3.2, $P = 0.0005$), and increased CST (OR = 1.6 per 100- μm increase, 95% CI 1.5–1.8, $P < 0.0001$) were significantly associated with need for SEC (Table 3). The overall model was statistically significant ($P < 0.0001$), and the C statistic was 0.851.

OCT Morphological Feature Grading Reproducibility

Reproducibility rates were high for the categorical variables scan quality at the fovea and CME. For the intra-arbitrator repeatability analysis, agreement rates were 97% and 95% for scan quality at the fovea and CME, respectively. Similarly, the kappa statistic was 0.92 (95% CI 0.82–1.00) and 0.90 (95% CI 0.82–0.98)

for scan quality at the fovea and CME, respectively. For the interarbitrator agreement analysis, agreement rates were 93% and 98% for scan quality at the fovea and CME, respectively. Likewise, kappa statistics were 0.79 (95% CI 0.64–0.94) and 0.96 (95% CI 0.91–1.00) for scan quality at the fovea and CME, respectively. For the interteam agreement analysis, agreement rates were 87% and 92% for scan quality at the fovea and CME, respectively, and corresponding kappa statistics were 0.64 (95% CI 0.45–0.82) and 0.84 (95% CI 0.74–0.94) (Table 4).

Discussion

In the present study, we found that OCT images obtained during an interventional CRVO treatment trial can be graded in a reproducible manner. Segmentation errors are common on these images, can cause clinically meaningful errors in central retinal thickness measurement, and can be reproducibly corrected in a reading center setting. Furthermore, the morphological features ERM, SRF, and increased CST were most associated with segmentation errors.

We and others have previously reported a significant error rate with the use of Stratus OCT software to measure retinal thickness in an automated manner. In the first such report, we showed that 43.2% of automated OCT thickness maps from a sample of 171 scans had artifacts that led to incorrect retinal thickness measurements in 62.2% of the scans. That study examined automated measurements calculated from scans acquired solely with the FMTM protocol, and scans underwent further analysis only if the foveal contour was abnormal on initial review of a

Table 2. Paired Differences for Automated Retinal Thickness and Volume Measurements From Repeat SEC Performed on 25 OCT Scans That Initially Required SEC

Variable	Statistic	Initial Grading	Repeat Grading	Absolute Difference	Paired* Difference
Central foveal thickness	<i>N</i>	25	25	25	25
	Mean (SD)	566.5 (242.9)	565 (243.3)	3.7 (5.7)	1.6 (6.6)
	Min, median, max	186, 631, 893	187, 631, 889	0, 2, 27	−11, 0, 27
Central subfield thickness	<i>N</i>	25	25	25	25
	Mean (SD)	554.7 (212.6)	555 (211.9)	2.2 (2)	−0.3 (3.1)
	Min, median, max	220, 608, 853	221, 606, 854	0, 2, 10	−10, 0, 4
Inferior inner macula	<i>N</i>	25	25	25	25
	Mean (SD)	452.5 (119.4)	453.6 (120.6)	3.5 (5.6)	−1.1 (6.5)
	Min, median, max	271, 426, 676	275, 425, 676	0, 1, 20	−20, 0, 14
Inferior outer macula	<i>N</i>	25	25	25	25
	Mean (SD)	323.7 (92.5)	324.6 (93)	1.8 (3)	−0.9 (3.4)
	Min, median, max	225, 293, 609	225, 290, 603	0, 0, 10	−10, 0, 6
Nasal inner macula	<i>N</i>	25	25	25	25
	Mean (SD)	488.6 (158.8)	491.3 (159.7)	4.4 (9.9)	−2.8 (10.5)
	Min, median, max	292, 453, 783	294, 450, 778	0, 1, 49	−49, 0, 6
Nasal outer macula	<i>N</i>	25	25	25	25
	Mean (SD)	385.2 (117.5)	389.6 (119.7)	4.7 (12.7)	−4.3 (12.8)
	Min, median, max	244, 338, 665	244, 352, 675	0, 0, 53	−53, 0, 2
Superior inner macula	<i>N</i>	25	25	25	25
	Mean (SD)	450.1 (117.6)	449.6 (117.5)	1.4 (2.2)	0.5 (2.6)
	Min, median, max	285, 429, 656	285, 437, 656	0, 0, 8	−8, 0, 6
Superior outer macula	<i>N</i>	25	25	25	25
	Mean (SD)	325.1 (86.2)	325.6 (88.3)	1.7 (2.9)	−0.4 (3.3)
	Min, median, max	239, 301, 561	236, 301, 572	0, 0, 11	−11, 0, 5
Temporal inner macula	<i>N</i>	25	25	25	25
	Mean (SD)	434.2 (110.6)	434.4 (111.1)	1.7 (2.3)	−0.3 (2.9)
	Min, median, max	253, 429, 645	257, 432, 645	0, 1, 7	−6, 0, 7
Temporal outer macula	<i>N</i>	25	25	25	25
	Mean (SD)	306.4 (73)	305.5 (71.3)	1.6 (4.1)	0.9 (4.3)
	Min, median, max	222, 297, 495	222, 297, 475	0, 0, 20	−3, 0, 20
Total macular volume	<i>N</i>	25	25	25	25
	Mean (SD)	10.4 (2.6)	10.5 (2.6)	0.1 (0.1)	0 (0.1)
	Min, median, max	7.1, 10.3, 16.5	7.1, 10.3, 16.5	0, 0, 0.3	−0.3, 0, 0.1

Paired difference, value of repeat grading subtracted from initial grading; absolute difference, absolute value of difference between initial grading and repeat grading values. All measurements are in micrometers except for total macular volume (mm³).

* *P* value based on Wilcoxon signed rank test of median difference equal to zero and *P* value not significant for any variable (*P* > 0.09).

surface map.¹⁴ These factors suggest that automated thickness measurement errors may be even more common in practice than reported in that early study. Supporting this assertion is a later series by Sadda and associates,¹² who reported that 92% of automated thickness maps were erroneous, though only 13.5% of

these 200 scans had severe errors. That study included images acquired using both the FMTM and MTM protocol, and every radial line image was reviewed in its entirety for the presence of measurement error.

Neither of these studies focused specifically on retinal venous occlusive diseases; however, other

Table 3. Associations Between Morphological Features and Need for SEC From 984 OCT Scans

Feature	Odds Ratio	95% Lower Confidence Interval	95% Upper Confidence Interval	P Value
Subretinal fluid	2.1	1.4	3.2	0.0005
Epiretinal membrane	2.3	1.7	3.3	<0.0001
Central subfield thickness per 100- μ m increase	1.6	1.5	1.8	<0.0001

Multiple logistic regression model included the following covariates: CME, CME at the fovea, maximal CME cyst width, SRF, SRF at the fovea, VMA, VMA at the fovea, ERM at the fovea, and CST. No missing values were in the dataset, and the C statistic for the model was 0.851.

investigators who did so similarly reported frequent detection of measurement errors. In the SCORE trial, 28.9% of OCT scans required manual measurement, most commonly due to segmentation errors.⁵ In a series of 11 eyes with venous occlusion, 19% of initial automated CPT measurements varied more than 25 μ m from a manually corrected value.¹³ Similarly, 23.9% of OCT scans required manual grading in a series of over 2800 scans of eyes with venous occlusive disease. The most common indication (67.6%) for manual measurement was segmentation error.¹⁴ The frequency of scans requiring SEC in these series is

roughly comparable to the 30.0% (360 of 1199) rate observed in our study. From our data and several previous investigations, it is clear that OCT automated segmentation errors are commonplace.

These automated segmentation errors not only are frequent, but also can lead to a substantial central retinal thickness measurement error. After SEC in 312 OCT scans, we observed a 102.3- μ m mean correction of absolute CPT error and 81.3- μ m mean correction of absolute CST error. The mean absolute center point error observed in our study was slightly less than the 129- μ m CPT error observed in a prior

Table 4. Temporal (More Than 5 Weeks After Initial Reading) Reproducibility Studies Performed for Scan Quality at the Fovea and CME

Variable	Sample Size	Percent Agreement	Kappa Statistic	Kappa Statistic 95% Lower Confidence Interval	Kappa Statistic 95% Upper Confidence Interval
Intra-arbitrator					
Scan quality at the fovea	97	97	0.92	0.82	1.00
Cystoid macular edema	97	95	0.90	0.82	0.98
Interarbitrator					
Scan quality at the fovea	97	93	0.79	0.64	0.94
Cystoid macular edema	97	98	0.96	0.91	1.00
Interteam					
Scan quality at the fovea	97	87	0.64	0.45	0.82
Cystoid macular edema	97	92	0.84	0.74	0.94

CME graded present versus absent; fovea on OCT, horizontal region 1 mm in width with midpoint at the foveal center on a 6-mm radial Stratus OCT image; not interpretable scan quality, missing, incorrectly positioned, or very poorly saturated; scan quality at the fovea graded interpretable versus not interpretable at the fovea (1-mm horizontal region around the foveal center point). Intra-arbitrator, repeatability analysis for 97 scans: An individual arbitrator received graded scans from a reader pair that he or she had previously arbitrated. Discrepant values within these scan pairs were arbitrated in a standard fashion, and the results for the repeat arbitration were compared to the initial arbitration. Interarbitrator, agreement analysis for 97 scans: An individual arbitrator received graded scans from a reader pair that he or she had not previously graded. Discrepant values within these scan pairs were arbitrated in a standard fashion, and the results for the second arbitration were compared to the first arbitration by the initial arbitrator. Interteam, agreement analysis for 97 scans: A repeat grading reading team (consisting of a new arbitrator and at least one new reader) received scans previously graded by the initial grading team. After grading by the repeat grading team, the results were compared to those of the initial team.

investigation of 28 eyes with retinal vein occlusion. In that study, the larger error originated from a smaller 422.4- μm uncorrected mean CPT measurement, compared to the smaller error in our series originating from a larger 499.0- μm mean uncorrected measurement.¹⁴ These disparities may be due to differences in sample size and method to determine scan quality.

In contrast to eyes with other pathologies such as diabetic macular edema (DME), segmentation errors seem to occur more frequently in eyes with CRVO (30% vs. <19%).²¹ Similarly, manual remeasurement in eyes with CRVO may result in larger central retinal thickness measurement correction compared to eyes with DME. For example, caliper remeasurement of central foveal thickness in a clinical trial for eyes with DME resulted in a corrected value for change in CPT of $34 \pm 18 \mu\text{m}$ compared to an uncorrected value of $23 \pm 17 \mu\text{m}$.²¹ This disparity in measurement errors between eyes with CRVO and eyes with DME is likely related, at least in part, to thinner baseline central retinal measurements in eyes with DME. For example, the average baseline thickness in a Diabetic Retinopathy Clinical Research Network trial was $340 \pm 123 \mu\text{m}$,²² a value that was almost 50% less than that observed in the present study.

The magnitude and frequency of central retinal thickness errors after automated segmentation indicate a need to formally address erroneous segmentation in trials for CRVO. Other automated segmentation algorithms have been reported that reduce segmentation error as much as 15%; however, these protocols are not in widespread use in clinical trials.²³ If segmentation errors are not corrected, clinically important measurement errors will occur, and they may decrease the validity of OCT-based endpoints. Likewise when caring for individual patients with CRVO, physicians should take caution to avoid basing treatment decisions on incorrect measurements. Undetected OCT segmentation errors resulting in misleading software-calculated retinal thickness may spuriously mimic a therapeutic effect or may conversely mask treatment efficacy. We observed that segmentation errors caused an underestimation of over 75 μm in both mean CPT and mean CST.

To verify the reproducibility of SECs, we compared automated retinal thickness and volume measurements before and after correction. We observed that automated retinal thickness and volume measurements after repeat SEC were extremely reproducible, with near-perfect intraclass correlations and very small mean absolute differences of less than 5 μm . Our 2.2- μm mean absolute difference in CST after SEC is comparable to the 1.2- μm intergrader differ-

ence reported by Sadda and colleagues²⁴ when they validated a computerized manual OCT grading program in 20 consecutive eyes. Those eyes were frequently the fellow healthy eye of patients with a diseased eye, and all segmentation lines were initially properly placed by the Stratus segmentation algorithm.

Particular morphological characteristics on OCT may be helpful in calling attention to the need to scrutinize automated thickness measurements in eyes with CRVO. We observed that ERM, SRF, and increased CST were each significantly associated with the need for SEC. Similarly, Domalpally and coworkers¹⁴ examined OCT scans from eyes with retinal vein occlusion, neovascular age-related macular degeneration, and DME. They reported that SRF and increased retinal thickness were most strongly associated with a need to manually remeasure retinal thickness. Our study differed from that report in that ERM was also associated with need for SEC. This difference may be due to dissimilarity in the frequency or severity of ERM (or both) in our cohort of eyes with CRVO when compared to the data aggregated from several different pathologies in the other study.¹⁴ On the other hand, Domalpally and coworkers did not list all morphological features tested for association with SEC. Thus their study may not have formally tested for a relationship between ERM and SEC.

Our reading center teams were able to reproducibly designate scan quality at the fovea. A consistent and universally applied standard of scan quality is the first step to obtain reproducible grading results. Our reading teams demonstrated high levels of agreement for scan acceptability, similar to the 89% to 91% rates of agreement reported for OCT scan quality during the SCORE trial.⁵ Similarly, reading center teams frequently noted CME on OCT scans. Reading center teams also identified this morphological feature with high levels of repeatability and agreement. These data were derived from a single reading center and thus may not be generalized to another reading center. However, our data are comparable to the 76% to 83% rates of agreement for cystoid spaces reported during the SCORE trial.⁵ Cystoid macular edema was chosen for reproducibility evaluation as this morphological feature can meaningfully impact the study endpoint of central retinal thickness in eyes with CRVO. Additionally, CME was selected since a prior OCT reproducibility study at our center demonstrated the lowest level of reproducibility when grading CME.²⁵

We implemented a team-based grading approach in which each OCT scan was evaluated by two

independent readers and arbitrated by a third senior reader.^{26,27} This method was chosen to maximize reading consistency during the trial, despite varying reader experience and reader turnover. This team-based grading process likewise allowed a senior reader to review a higher number of scans and to establish a close feedback loop with newer readers to maximize consistent grading over time. Importantly, senior readers demonstrated high rates of grading reproducibility during interarbitrator and intra-arbitrator testing. In contrast, prior series detailing OCT reading protocols have utilized individual examiners^{5,28} and multiple grader pairs in parallel.²⁵

The protocol used in the ETDRS study for fundus photograph grading has most similarity to our team protocol for reading OCT scans. In the ETDRS study, only baseline fundus photos were reviewed by a pair of readers, and a subset of discrepancies were arbitrated by a senior grader.¹⁷ Our protocol differed in that independent reading teams graded both baseline and follow-up OCT scans, all discrepancies were arbitrated by a senior reader, and reproducibility studies were performed by a full reading center team.

Higher-resolution spectral domain OCT technology (SD-OCT) offers many opportunities to further study macular edema associated with CRVO. Compared to conventional time domain OCT (TD-OCT), such as Stratus OCT used for this study, SD-OCT offers increased image resolution and more rapid data acquisition leading to subsequent decreased motion artifact.^{29,30} Nonetheless, we and others have shown that segmentation errors occur relatively frequently on SD-OCT images, and, if not corrected, will lead to clinically meaningful thickness measurement errors.^{28,31,32} For example, we found that more than 80% of Cirrus (Stratus OCT, Carl Zeiss Meditech) and Spectralis (Heidelberg Engineering, Carlsbad, CA) volume cubes had segmentation errors on at least one scan, and 37.5% had at least one scan artifact in the central 1-mm subfield.³¹ Accordingly, it is still necessary to correct segmentation errors in interventional clinical trials that use SD-OCT.

Our reading center protocol for systematic grading of OCT scans that depict macular edema secondary to CRVO is reproducible and can be used in a highly consistent fashion to correct the retinal thickness errors frequently generated by automated segmentation algorithms. Accurate retinal thickness measurements are critical to allow meaningful comparisons of treatment effects in interventional trials such as the COPERNICUS study described in the present report, as well as to optimally manage patients in a clinical setting.

Acknowledgments

Supported in part by the Heed Foundation and Ronald G. Michels Foundation.

Disclosure: **F.C. DeCroos**, None; **S.S. Stinnett**, None; **C.S. Heydary**, None; **R.E. Burns**, None; **G.J. Jaffe**, None

References

1. Hee MR, Puliafito CA, Wong C, et al. Quantitative assessment of macular edema with optical coherence tomography. *Arch Ophthalmol*. 1995; 113:1019–1029.
2. Lerche RC, Schaudig U, Scholz F, Walter A, Richard G. Structural changes of the retina in retinal vein occlusion—imaging and quantification with optical coherence tomography. *Ophthalmic Surg Lasers*. 2001;32:272–280.
3. Brown JC, Solomon SD, Bressler SB, et al. Detection of diabetic foveal edema: contact lens biomicroscopy compared with optical coherence tomography. *Arch Ophthalmol*. 2004;122:330–335.
4. Brown DM, Campochiaro PA, Singh RP, et al. Ranibizumab for macular edema following central retinal vein occlusion: six-month primary end point results of a phase III study. *Ophthalmology*. 2010;117:1124–1133.
5. Domalpally A, Blodi BA, Scott IU, et al. The Standard Care vs Corticosteroid for Retinal Vein Occlusion (SCORE) study system for evaluation of optical coherence tomograms: SCORE study report 4. *Arch Ophthalmol*. 2009;127:1461–1467.
6. Haller JA, Bandello F, Belfort R Jr, et al. Randomized, sham-controlled trial of dexamethasone intravitreal implant in patients with macular edema due to retinal vein occlusion. *Ophthalmology*. 2010;117:1134–1146, e1133.
7. Prager F, Michels S, Kriechbaum K, et al. Intravitreal bevacizumab (Avastin) for macular oedema secondary to retinal vein occlusion: 12-month results of a prospective clinical trial. *Br J Ophthalmol*. 2009;93:452–456.
8. Priglinger SG, Wolf AH, Kreutzer TC, et al. Intravitreal bevacizumab injections for treatment of central retinal vein occlusion: six-month results of a prospective trial. *Retina*. 2007;27:1004–1012.
9. Antcliff RJ, Stanford MR, Chauhan DS, et al. Comparison between optical coherence tomogra-

- phy and fundus fluorescein angiography for the detection of cystoid macular edema in patients with uveitis. *Ophthalmology*. 2000;107:593–599.
10. Wilkins JR, Puliafito CA, Hee MR, et al. Characterization of epiretinal membranes using optical coherence tomography. *Ophthalmology*. 1996;103:2142–2151.
11. Gallemore RP, Jumper JM, McCuen BW II, et al. Diagnosis of vitreoretinal adhesions in macular disease with optical coherence tomography. *Retina*. 2000;20:115–120.
12. Sadda SR, Wu Z, Walsh AC, et al. Errors in retinal thickness measurements obtained by optical coherence tomography. *Ophthalmology*. 2006;113:285–293.
13. Taban M, Sharma S, Williams DR, Waheed N, Kaiser PK. Comparing retinal thickness measurements using automated fast macular thickness map versus six-radial line scans with manual measurements. *Ophthalmology*. 2009;116:964–970.
14. Domalpally A, Danis RP, Zhang B, Myers D, Kruse CN. Quality issues in interpretation of optical coherence tomograms in macular diseases. *Retina*. 2009;29:775–781.
15. Ray R, Stinnett SS, Jaffe GJ. Evaluation of image artifact produced by optical coherence tomography of retinal pathology. *Am J Ophthalmol*. 2005;139:18–29.
16. Brown DM, Heier JS, Clark WL, et al. Intravitreal aflibercept injection for macular edema secondary to central retinal vein occlusion: 1-year results from the phase 3 COPERNICUS study. *Am J Ophthalmol*. 2013;155:429–437, e427.
17. Early Treatment Diabetic Retinopathy Study design and baseline patient characteristics. ETDRS report number 7. *Ophthalmology*. 1991;98:741–756.
18. Peto T. Reading centers of fundus imaging. *J Ophthalmic Photog*. 2010;32:47–49.
19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
20. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37–46.
21. Glassman AR, Beck RW, Browning DJ, Danis RP, Kollman C. Comparison of optical coherence tomography in diabetic macular edema, with and without reading center manual grading from a clinical trials perspective. *Invest Ophthalmol Vis Sci*. 2009;50:560–566.
22. Fong DS, Strauber SF, Aiello LP, et al. Comparison of the modified Early Treatment Diabetic Retinopathy Study and mild macular grid laser photocoagulation strategies for diabetic macular edema. *Arch Ophthalmol*. 2007;125:469–480.
23. Haeker M, Abramoff M, Kardon R, Sonka M. Segmentation of the surfaces of the retinal layer from OCT images. *Med Image Comput Comput Assist Interv*. 2006;9:800–807.
24. Sadda SR, Joeres S, Wu Z, et al. Error correction and quantitative subanalysis of optical coherence tomography data using computer-assisted grading. *Invest Ophthalmol Vis Sci*. 2007;48:839–848.
25. Zhang N, Hoffmeyer GC, Young ES, et al. Optical coherence tomography reader agreement in neovascular age-related macular degeneration. *Am J Ophthalmol*. 2007;144:37–44.
26. DeCroos FC, Toth CA, Stinnett SS, et al. Optical coherence tomography grading reproducibility during the Comparison of Age-related Macular Degeneration Treatments Trials. *Ophthalmology*. 2012;119:2549–2557.
27. DeCroos FC, Toth CA, Folgar FA, et al. Characterization of vitreoretinal interface disorders using OCT in the interventional phase 3 trials of ocriplasmin. *Invest Ophthalmol Vis Sci*. 2012;53:6504–6511.
28. Krebs I, Smretschne E, Moussa S, et al. Quality and reproducibility of retinal thickness measurements in two spectral-domain optical coherence tomography machines. *Invest Ophthalmol Vis Sci*. 2011;52:6925–6933.
29. Srinivasan VJ, Wojtkowski M, Witkin AJ, et al. High-definition and 3-dimensional imaging of macular pathologies with high-speed ultrahigh-resolution optical coherence tomography. *Ophthalmology*. 2006;113:2054.e1–2054.e14.
30. Wojtkowski M, Bajraszewski T, Gorczynska I, et al. Ophthalmic imaging by spectral optical coherence tomography. *Am J Ophthalmol*. 2004;138:412–419.
31. Han IC, Jaffe GJ. Evaluation of artifacts associated with macular spectral-domain optical coherence tomography. *Ophthalmology*. 2010;117:1177–1189, e1174.
32. Ho J, Sull AC, Vuong LN, et al. Assessment of artifacts and reproducibility across spectral- and time-domain optical coherence tomography devices. *Ophthalmology*. 2009;116:1960–1970.