The Standard Care vs Corticosteroid for Retinal Vein Occlusion (SCORE) Study System for Evaluation of Optical Coherence Tomograms

Amitha Domalpally, MD; Barbara A. Blodi, MD; Ingrid U. Scott, MD, MPH; Michael S. Ip, MD; Neal L. Oden, PhD; Andreas K. Lauer, MD; Paul C. VanVeldhuisen, PhD; for the SCORE Study Investigator Group

Objective: To describe grading procedures for optical coherence tomographic (OCT) images of participants in the Standard Care vs Corticosteroid for Retinal Vein Occlusion (SCORE) Study.

Methods: Optical coherence tomograms were taken at clinical sites with the Stratus OCT using fast macular and crosshair scan protocols. Paper prints of images were evaluated at a central reading center. Quality evaluation identified the accuracy of OCT-measured retinal thickness data and was categorized as good, fair, borderline, or ungradable. Manual measurement of center point thickness was performed on borderline images. Morphological evaluation identified cystoid spaces, subretinal fluid, and vitreoretinal interface abnormalities. Reproducibility of grading was assessed through formal quality control exercises.

Results: A randomly selected set of 106 images was identified for quality control. The first 2 annual regrades showed 91% and 89% intergrader agreement for OCT quality. Intraclass correlation for manually measured center point thickness was 0.99 per year. For morphological variables, intergrader agreement for cystoid spaces was 83% and 76%. Reproducibility for subretinal fluid and vitreoretinal interface abnormalities could not be interpreted owing to their limited presence in the sample.

Conclusion: Optical coherence tomogram evaluation procedures used in the SCORE Study are reproducible and can be used for multicenter longitudinal studies of retinal vein occlusion.


Optical coherence tomography (OCT) allows qualitative and quantitative evaluation of macular edema secondary to retinal diseases. The principles behind OCT imaging have been described previously. Optical coherence tomography–based central retinal thickness measurement is an important endpoint in clinical trials related to retinal vascular diseases, including retinal vein occlusions. Morphological characterization of macular edema using OCT provides additional information regarding the pathophysiology of retinal vein occlusions such as frequency of occurrence of subretinal fluid and subretinal hemorrhage.

The Standard Care vs Corticosteroid for Retinal Vein Occlusion (SCORE) Study consists of 2 multicenter randomized phase 3 trials comparing the efficacy and safety of standard care with that of intravitreal injection(s) of triamcinolone acetonide for the treatment of vision loss due to macular edema associated with central retinal vein occlusion and branch retinal vein occlusion. Change in retinal thickness at the center of the macula, as assessed by OCT, is one of the secondary efficacy outcomes of the study. Certified imaging technicians at each clinical site obtained OCT images using a standardized imaging protocol. The images were evaluated at the Fundus Photograph Reading Center, University of Wisconsin, Madison (reading center) by trained and certified ocular disease evaluators (graders). In this article, we describe the evaluation procedures for OCT images and provide data on the reproducibility of the grading system. The original imaging protocol used for the study is available from the United States National Technical Information Service.

METHODS

IMAGING PROTOCOL

The OCT images were obtained using the Stratus OCT3 (Carl Zeiss Meditec, Dublin, Califor...
nias) by imaging technicians who were certified by the reading center (2 images were from OCT2 [Carl Zeiss Meditec]). The OCT images were initially evaluated for quality by the imaging technician at the clinic site before submission to the reading center, and poor quality images were retaken if possible.

As directed by the SCORE Study protocol, the study eye of the subject was scanned at baseline and every 4 months thereafter for up to 36 months. The fellow eye was scanned at baseline, month 4, month 12, month 24, and month 36. The scanning protocol included the fast macular scan (128 A scans per B scan) and the higher-resolution (512 A scans per B scan) crosshair scan, both using 6.0-mm scan length and 6.0-mm display diameter. The characteristics of the fast macular scanning protocol are summarized elsewhere.12,13 Each submission to the reading center required paper printouts of the fast macular thickness map analysis report (6 individual retinal thickness reports consisting of 6 individual radial B scan images) and the horizontal and vertical crosshair scans. All images were deidentified for subject identifiers in compliance with Health Insurance Portability and Accountability Act regulations.

**Grading Procedure**

Graders evaluated the OCT images independently (ie, no reference made to previous visits or other image types for the same subject). Each image was graded by a single grader.

For each visit, the following information was derived from an OCT image: (1) the quality of the scan, (2) the center point thickness and the retinal thickness measurements of the 9 Early Treatment Diabetic Retinopathy Study subfields, (3) total macular volume, and (4) retinal morphology. Retinal thickness measurements were taken directly from the fast macular thickness map analysis report unless the center point measurement was determined to be inaccurate by quality assessment. If the center point was measured inaccurately by the software, manual measurement was performed (see below).

**Assessment of Overall Quality**

The goal of the overall quality assessment was to determine the accuracy of the numeric output of the fast macular thickness report. The graders reviewed this report to assess the presence of artifacts, mainly boundary line errors and decen- tration. The categories for quality assessment were good, fair, borderline, and ungradable. A grade of good indicated that the OCT image was free of artifacts and the software-generated center point thickness available on the paper prints was recorded in the grading form. A grade of fair indicated that boundary line errors were present but do not involve the center point. In such cases, the center point thickness was accurate but the subfield grid had inaccurate values. The evaluation form allowed the grader to document the reliability of each subfield value within the grid. Figure 2 shows examples of good and fair categories.

Borderline quality indicated that the center point thickness was inaccurate (Figure 3) and manual measurement must be performed. Boundary line errors affecting the center point and/or decen- tration could lead to a grade of borderline quality. All images with a standard deviation of more than 10% of center point thickness were measured manually.14

---

**Figure 1.** Crosshair scan overlaid with a lateral scale for assessing lateral extent of morphological abnormalities. Dashed lines represent the limits of the central and the inner subfields on the B scan.

---

**Figure 2.** Examples of good and fair categories.

---

**Figure 3.** Examples of borderline and ungradable categories.
Ungradable quality indicated that the entire image had such severe inaccuracies that a manual measurement could not be performed. Examples of ungradable quality include severe scan alignment artifact, in which the OCT image is cut off from the display window and the inner and/or outer retinal layers are not visible for determination of boundaries. Very low signal strength could also result in ungradable images (Figure 2).

**TYPES OF ERRORS ON OCT IMAGES**

**Boundary Line Errors**

Boundary line errors are artifacts in which the segmentation algorithm fails to identify the inner retina and/or the outer retina correctly. Boundary line errors are identified by inaccurate tracing of the white lines on the OCT image and by the presence of wedge or bow tie artifacts in the false color map (Figures 2 and 3). Boundary line errors are the most common cause of inaccurate center point thickness.15

**Decentration**

Decentration of the OCT image occurs when the intersection of the 6 radial line scans of the fast macular scanning report does not coincide with the center of the macula. Decentration was assessed using both the map report and the individual B scan images (Figure 4). All 6 B scans were examined to identify the location of the fovea with respect to the center point of

---

**Figure 2.** Quality categories for optical coherence tomograms.
A shift of the fovea by more than 10 A scans (500 µm) on either side of the center point (A scan number 64) was considered decentration. The false color map helped identify the shift of the fovea (blue area in eyes with foveal depression) from the central subfield. The gray scale fundus image (which depicts the location of the 6 radial lines with respect to the anatomical fovea) was also used to assess centration, although it was taken after the scan was complete.

Other indicators such as low signal strength (<3%) and low analysis confidence message (a tool found in newer software versions) alerted the graders to the presence of an artifact.

**MANUAL MEASUREMENT OF CENTER POINT THICKNESS**

Manual measurement of the center point thickness included 2 basic steps: identification of the fovea and the measurement itself. In decentered scans, identification of the fovea is important to determine the point for manual measurement. From the 6 radial B scans, the image that best represented the foveal depression was chosen for manual measurement. If the foveal depression was absent, the crosshair scans were compared with the fast macular scans for identification of the fovea. Other features used to identify the fovea included the attenuation (or absence) of the ganglion cell layer and nerve fiber layer at the fovea. In eyes with cystoid edema, the location of the largest cyst was assumed to be the fovea. In borderline quality images with boundary line errors, the white line denoting the automatically detected internal limiting membrane and/or retinal pigment epithelium (RPE) was ignored and the correct layers identified.

Handheld digital calipers were used for manual measurement of center point thickness on paper prints. After ensuring that the digital readout was calibrated, the caliper tips were opened until the points for measurement just obscured the internal limiting membrane and RPE lines at the fovea. A scale factor was used to convert the caliper readout from millimeters to micrometers. **Figure 5** describes an example of the scale factor calculation. In the SCORE Study, 28.9% of the baseline images were manually measured.
MORPHOLOGICAL EVALUATION

There were 3 main codes for assessing the presence of a morphological abnormality: absent, questionable (meaning probable), and definitely present. The definitions of these 3 codes were similar to the definitions used in Early Treatment Diabetic Retinopathy Study Report 1016 and Age-Related Eye Disease Study Report 6.17

The higher-resolution crosshair scans were used to assess 3 distinct retinal morphologies: intraretinal cystoid spaces, subretinal fluid, and vitreoretinal interface abnormalities. The measurement procedure for quantifying these morphologies using the calipers was similar to that of center point thickness manual measurement using a scale factor calculation.

Cystoid spaces are identified as round, well-defined spaces within the neurosensory retina of at least 2 × 2 mm (60 × 60 µm) on the paper prints. A millimeter ruler was used to approximate the cyst size. The cyst cavity is typically nonreflective (dark) or minimally reflective. The presence or absence of cystoid spaces at the center point was evaluated, as well as their lateral extent with respect to the central subfield, using the 127-mm lateral scale. In addition, the height of the cystoid space at the center point was measured.

Subretinal fluid was identified on OCT as a predominantly nonreflective (dark) space between the posterior boundary of the neurosensory retina and an intact RPE/Bruch junction. The subretinal fluid is typically dome-shaped, and its greatest vertical height was measured. The location and lateral extent of the subretinal fluid were categorized in a manner similar to cysts.

Three types of vitreoretinal interface abnormalities were recorded: posterior vitreous detachment, epiretinal membrane, and macular holes. In an eye with posterior vitreous detachment, the posterior hyaloid membrane is visible on an OCT image as a thin, weakly reflecting membrane located in the dark area anterior to the retina. The attachment of the posterior hyaloid membrane to the internal limiting membrane was categorized as partially adherent or nonadherent. Tenting of the retinal tissue at or around the point of adherence was categorized as vitreomacular traction. Epiretinal membrane was identified by the presence of a well-delineated layer of increased density/reflectivity on the retinal surface. Other features supportive of an epiretinal membrane included corrugation of the retina, bridging of the innermost layer of retinal tissue, and flattening of the fovea. Retinal distortion caused by epiretinal membrane was characterized by scalloped or cleft-like areas with peaking of the retinal tissue at points of adherence or as irregular retinal tissue under the epiretinal membrane. A macular hole was identified as either a pseudohole, lamellar hole, or a full-thickness macular hole. No attempt was made to distinguish between pseudoholes and lamellar holes, both of which were identified by an abnormally wide foveal depression with a steep foveal contour. Intact outer layers of retinal tissue just above the RPE distinguish a pseudohole or lamellar hole from a full-thickness hole.

QUALITY CONTROL AND REPRODUCIBILITY

Various levels of quality control (QC) programs were performed to maintain intergrader reproducibility. An ongoing monthly QC program ensured that approximately 5% of the images were regraded every month. On a quarterly basis, intergrader agreements were generated for the whole group and for individual graders. Reproducibility data were used to identify the characteristic(s) for which the grader differed from the group, followed by focused training.

Continuous training programs in terms of bimonthly QC meetings also helped maintain excellent reproducibility. The grading team leaders and ophthalmologists at the reading center lead the meetings, in which images (randomly selected or images that focused attention on particular grading issues) were graded with the group's input, as an exercise. Difficulties in grading were handled through these meetings, which have proven to be an efficient way of reducing differences between the grading team members.

A randomly selected sample of SCORE Study images was identified for quality control with the intent to have them regraded annually through the course of the SCORE Study. A total of 106 images of 33 subjects were randomly selected from the SCORE Study OCT images that had grading completed by July 2006, irrespective of the visit. Eight graders participated in the annual exercise, and each scan was regraded by 2 or 3 graders. Each scan had a single grade of record that was exported for data analysis to the data coordinating center. Data from the QC exercises were compared using cross tabulations (grade of rec-
The result of the first 2 annual exercises is listed in the Table. Of the 370 QC grades available for 106 images, there was exact agreement for all categories in 336 (91%) in year 1 and 331 (89%) in year 2. The agreement rate for need of manual measurement was 92% in both years. The intraclass correlation for manually measured center point thickness was 0.994 in year 1 and 0.998 in year 2. The thickness value measured in the QC exercise was within 50 µm of the measurement in the grade of record in 90% of images in year 1 and 95% in year 2.

Of the morphological variables, cystoid spaces was the only variable recorded with a significant frequency; 55 (52%) had cystoid spaces present. The agreement rate for both presence and location of cystoid spaces was 83% in year 1 and 76% in year 2. Reproducibility could not be interpreted for subretinal fluid and vitreoretinal interface abnormalities owing to their limited presence in the sample.

### Table. Agreement Rates Between Grade of Record and Quality Control Grade in the First and Second Annual Retinal Vein Occlusion Temporal Drift Exercisea

<table>
<thead>
<tr>
<th>OCT Variable</th>
<th>Scale Steps and Distribution, No. (%) From Grade of Record</th>
<th>Exact Agreement, %</th>
<th>Unweighted κ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Year 1</td>
<td>Year 2</td>
</tr>
<tr>
<td>Overall quality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>84 (79)</td>
<td>0.78</td>
<td>0.75</td>
</tr>
<tr>
<td>Fair</td>
<td>5 (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borderline</td>
<td>17 (16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ungradable</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Need for manual measurement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>89 (84)</td>
<td>0.69</td>
<td>0.55</td>
</tr>
<tr>
<td>Yes</td>
<td>17 (16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cystoid spaces</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>51 (48)</td>
<td>0.69</td>
<td>0.55</td>
</tr>
<tr>
<td>Present outside CSF</td>
<td>3 (3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present within CSF</td>
<td>52 (49)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: OCT, optical coherence tomography; CSF, central subfield.

a n = 370 images.

The SCORE Study is the first large prospective study comparing the efficacy and safety of standard care with that of intravitreal injection(s) of triamcinolone acetonide for the treatment of macular edema secondary to central retinal vein occlusion and branch retinal vein occlusion. One of the secondary outcomes of the study is to use OCT to assess changes in retinal thickness at the center of the macula (center point thickness). In addition to assessing center point thickness, morphological assessment of OCT images provides important information that may help to better measure prognosis of the disease and response to treatment.

Artifacts affecting retinal thickness measurements are common,15,18,19 and it is, therefore, important to assess the quality of the OCT images to confirm the accuracy of the data. Clues that would support the accuracy of the thickness measurements from the map report include a high standard deviation of center point thickness, low signal strength, low confidence analysis message, and artifacts on the false color map. The map report should always be examined with the underlying B scans to confirm artifacts.

At the reading center, manual measurement of the center point is performed if the software-generated thickness is inaccurate. Manual measurement of center point thickness was performed in 28.9% of baseline OCTs in the SCORE Study and is substantially higher than the rate of manual measurement in a clinical trial of patients with diabetic macular edema (19%).20 This may be owing to the larger percentage of images with center-involved subretinal fluid in patients with retinal vein occlusion that could lead to boundary line errors, resulting in inaccurate center point thickness.9,19,21 Presence of intraretinal and subretinal blood and the abrupt transition at the fovea between edematous and nonedematous retina, particularly in eyes with branch retinal vein occlusion, are other possible causes of boundary line errors.

In the SCORE study, the method by which center point thickness was measured predated the availability of the Stratus OCT review software in which electronic calipers display the retinal thickness in micrometers. This study describes the method by which measurements in millimeters on OCT paper scans may be converted to micrometers to determine center point thickness. Correlation between caliper measurement of center point thickness on OCT prints and automated center point thickness in good-quality images showed excellent agreement (intraclass correlation, 0.956; data not published).

Assessment of morphological abnormalities on OCT images in retinal vein occlusion can be more difficult than in other retinal vascular diseases such as diabetic retinopathy, as severe superficial and deep retinal hemorrhages are present in many images. The high reflectivity of superficial hemorrhages leads to shadowing, making detection of intraretinal cysts and the outer retinal/RPE
Submitted for Publication: January 21, 2009; final revision received March 26, 2009; accepted April 6, 2009. Correspondence: Barbara A. Blodi, MD, Department of Ophthalmology and Visual Sciences, University of Wisconsin Madison, 2870 University Avenue, Room 206, Madison, WI 53705-3611 (bablodi@wisc.edu).

Financial Disclosure: None reported.

Funding/Support: The Standard Care vs Corticosteroid for Retinal Vein Occlusion (SCORE) Study was supported by National Eye Institute (National Institutes of Health, Department of Health and Human Services) grants 5U10EY014351, 5U10EY014352, and 5U10EY014404; and Allergan, Inc.

Role of the Sponsor: Allergan, Inc, donated the investigational drug and partially funded the monitoring visits and secondary data analyses.

boundary line difficult. In addition, superficial hemorrhages may be seen as hyperreflective inner retinal layer bands that can be mistaken for epiretinal membrane in some instances. Deep retinal hemorrhages may also interfere with identification of the outer retinal/RPE line. These issues may explain the moderate agreement rates for their detection.

The temporal drift QC evaluates the intergrader reproducibility of a predesigned sample over time. In long-term studies, this measurement variability is an important consideration when assessing the effect of treatment or natural history of a disease. These exercises help identify fluctuations in the grading group's methodology over the years of the study owing to changes in the grading group population during the follow-up period. In addition, changes in technology, the grading methodology itself, and experience gained over the years are some of the other factors that could affect temporal drift QC. Each image is evaluated independently by a single grader with no access to previous visit data or other images such as fundus photographs. Good contemporaneous inter-grader agreement and monitoring of the temporal drift is, therefore, essential to minimize error in longitudinal data analysis. There is little evidence of temporal drift for most OCT grading variables during the SCORE Study. There is a slight drift in the grading of cystoid spaces, with less agreement between the year 2 QC exercises and grade of record (83% in year 1 vs 76% in year 2). Images graded as having questionable cysts in the grade of record were regraded as having definite cysts in the second year. This could be attributed to the increase in graders' experience and confidence in identifying small cysts over the course of the study.

In summary, the reading center methodology for measuring center point thickness and morphologic features on paper OCT images are detailed in this study. The OCT image evaluation procedures used for the SCORE Study reproducibly measured center point thickness and were able to assess morphologic features in patients affected by retinal vein occlusion. This grading methodology can be used for other multicenter longitudinal studies or smaller studies of retinal vein occlusion.

REFERENCES


