Optical Coherence Tomography Grading Reproducibility during the Comparison of Age-related Macular Degeneration Treatments Trials

Francis Char DeCroos, MD,[1,2] Cynthia A. Toth, MD,[2] Sandra S. Stinnett, DrPH,[2] Cynthia S. Heydary, BA,[2] Russell Burns, AAS,[2] Glenn J. Jaffe, MD,[1] for the CATT Research Group

Objective: To report reading center reproducibility during grading of Stratus optical coherence tomography (OCT) (Carl Zeiss Meditec, Dublin, CA) images obtained during the Comparison of Age-Related Macular Degeneration Treatments Trials (CATT).

Design: Prospective, clinical trial.

Participants: Independent reading teams reevaluated 270 OCT scans randomly sampled from the first 2 years of CATT enrollment. To assess temporal drift, a cohort of 23 scans submitted during the initial portion of the CATT study was longitudinally followed with serial reproducibility analysis.

Intervention: The CATT readers performed standardized grading of OCT images. A reader team, composed of 2 independent readers and a senior reader, evaluated each scan. Grading included the CATT OCT end points of total thickness at the foveal center point and intraretinal fluid (IRF), subretinal fluid (SRF), and subretinal pigment epithelium (RPE) fluid. Independent reading teams masked to the results of initial grading reevaluated scans to determine the reproducibility of qualitative grading and measurements.

Main Outcome Measures: Categorical grading agreement was reported using percent agreement and kappa statistic, and measurement agreement was reported using intraclass correlations and paired differences.

Results: Reading center teams reproducibly graded IRF (percent agreement = 73%, kappa = 0.48; 95% confidence interval [CI], 0.38–0.58), SRF (percent agreement = 90%; kappa = 0.80; 95% CI, 0.73–0.87), and sub-RPE fluid (percent agreement 88%; kappa = 0.75; 95% CI, 0.67–0.83). For independent reading center team measurements of total thickness at the foveal center point, the intraclass correlation was 0.99 (95% CI, 0.99–0.99), and the mean paired difference between reading center teams was 4 μm (95% limits of agreement, -55 to 47 μm). There was no qualitative or quantitative grading drift.

Conclusions: The standardized protocols used to evaluate OCT scans from the CATT study were reproducible. The methods used are suitable to monitor OCT imaging data from a large, neovascular age-related macular degeneration, interventional, multicenter study.

Financial Disclosure(s): The author(s) have no proprietary or commercial interest in any materials discussed in this article. *Ophthalmology 2012;119:2549–2557 © 2012 by the American Academy of Ophthalmology.*

The Comparison of Age-Related Macular Degeneration Treatments Trials (CATT) is a prospective, randomized, multicenter, clinical trial that compares the relative safety and efficacy of intravitreal bevacizumab with intravitreal ranibizumab as interventions for neovascular age-related macular degeneration (NVAMD).[1] This trial also examines the relative efficacy of different dosing schedules of each agent. Various imaging modalities, including optical coherence tomography (OCT) (Carl Zeiss Meditec, Dublin, CA), were used to monitor CATT study patient response to therapy.

Optical coherence tomography provides a noninvasive way to obtain cross-sectional images of the retina. Anatomic changes associated with NVAMD, such as intraretinal fluid (IRF), subretinal fluid (SRF), hyperreflective material under the retina, and pigment epithelial detachment (PED), can be readily visualized on OCT.[2–4] These pathologic changes[5] and the efficacy of various treatments[6–9] can be followed longitudinally on OCT. Furthermore, OCT-facilitated determination of the presence or absence of macular fluid associated with choroidal neovascularization (CNV) also has been used to rationally direct intravitreal pharmacologic therapy.[10–13]

In the CATT, macular fluid, defined as 1 or more of the following: IRF, SRF, or subretinal pigment epithelium (RPE) fluid, was an eligibility prerequisite, re-treatment criteria, and secondary study end point. Accordingly, it is important to accurately and reproducibly identify macular

Published by Elsevier Inc.

ISSN 0161-6420/12/$–see front matter
http://dx.doi.org/10.1016/j.ophtha.2012.06.040

fluid to ensure appropriate study enrollment and treatment and to correctly interpret study results.

To evaluate CATT OCT images, we adopted a novel team-based grading approach: a pool of CATT readers was chosen, and 2 readers selected from this pool independently graded each scan. Any discrepancies between the 2 readers were arbitrated by a senior reader. We report the reproducibility of the CATT OCT grading protocol and whether the grading changed over time.

Materials and Methods

Approval for this study was obtained from the Duke Institutional Review Board. All experimental procedures adhered to the tenets of the Declaration of Helsinki, and all participants engaged in an informed consent process and signed a written consent document before enrollment in the CATT (ClinicalTrials.gov Identifier: NCT00593450). For the CATT, the qualitative OCT end point was the presence of macular fluid, and the quantitative end point was thickness at the foveal center. A description of OCT acquisition procedures, site technician and reader certification, and grading methodology can be found in Chapter 18 of the CATT Manual of Procedures (available at: http://www.med.upenn.edu/cpob/studies/CATT.shtml; accessed May 26, 2012).

Reader Certification

Certified readers reviewed all scans. To become certified, readers were required to review an OCT grading manual, complete a training curriculum, pass an OCT reader knowledge assessment test, and be closely supervised by a senior reader until grading was determined to be accurate. When the CATT was initiated, a pool of 2 senior readers (readers who have fulfilled all of the training requirements of a reader and have completed additional prespecified advanced training activities) and 3 readers were designated as CATT readers. As the study scan volume increased, these numbers were expanded to include a pool of 4 senior readers and 8 readers. At any 1 time, 3 to 8 readers and senior readers concurrently analyzed study scans.

Optical Coherence Tomography Scan Acquisition

All study scans were acquired by CATT-certified OCT technicians using Stratus OCT machines (Carl Zeiss Meditec, Dublin, CA). To become CATT certified, a technician successfully completed a knowledge assessment test and received image acquisition training that emphasized appropriate focus, scan saturation, line length, and line placement. The technician submitted 16 certification scans to reading center imaging specialists who evaluated the scans to verify that the scans were of high quality and were obtained according to the study scan protocol. Certification was awarded once the technician had successfully completed these requirements. An automated e-mail feedback system reported scan quality, placement, and individually identified scans of concern to OCT technicians for all scans submitted to the reading center during the CATT.

Before submission of OCT scans to the reading center, all patient-identifying data were removed in compliance with Health Insurance Portability and Accountability Act guidelines. All eyes were imaged with both the fast macular thickness map (FMTM) and the macular thickness map (MTM) scan protocols. Less than 1% of scans submitted to the reading center did not adhere to this submission protocol.

Optical Coherence Tomography Scan Grading

Each of 12 (6 from the FMTM and 6 from the MTM) radial line images was assessed during grading. All OCT scans were analyzed for the presence of the following parameters: vitreomacular attachment, epiretinal membrane (ERM), IRF, SRF, subretinal hyperreflective material (SHRM), and RPE elevation (RPEE) (Figs 1A–F and 2A–F). Vitreomacular attachment was defined as vitreous attachment and focal separation from the inner retina within a 3-mm diameter centered at the middle of the fovea. Standardized reference images were compiled to illustrate examples of each morphologic feature and were made available to all CATT readers.

For each morphologic feature evaluated, 1 of the following grades was assigned: feature present, feature absent, not interpretable (because of incorrect scan placement or poor scan saturation), or absent scan. The OCT morphologic features graded as present ranged from subtle to obvious. Examples of obvious and subtle morphologic findings are shown in Figure 1A–F and Figure 2A–F.

If a particular OCT morphologic feature was graded as present on a scan, a reader was always required to further subcategorize the feature. For example, if ERM or vitreomacular attachment was graded present, a reader recorded the presence of any associated deformation of the central 1 mm of the retina (Fig 3A and B). If RPEE was present, a reader recorded whether sub-RPE fluid was present. If macular fluid ($\geq$1 of IRF, SRF, or sub-RPE fluid) was graded as present, a reader determined whether that specific type of macular fluid was present anywhere within the central 1 mm of the OCT scan and whether any macular fluid was present at the foveal center point (Fig 3C–H). Finally, if SHRM was graded as present, a reader determined whether SHRM was present anywhere within the central 1 mm of the retina (Fig 3I).

After morphologic grading was completed, morphometric analysis was performed on each scan. Quantitative values for morphometric variables were preferentially recorded from the 6 radial line images produced by the MTM protocol, although if these were of not acceptable quality, individual images from the FMTM protocol could be substituted. The largest horizontal and vertical dimensions for RPEE were measured from each of the 6 radial line scans, and the maximum value on a single radial scan for both dimensions was reported. We defined RPEE height (vertical dimension) from Bruch's membrane to the basal RPE surface of the RPE and RPEE width (horizontal dimension) from the point where the RPE started to separate from the choroid and become elevated to the point where the RPE was flat against Bruch's membrane and no longer elevated.

For each radial line scan evaluated, thickness (vertical dimension) at the foveal center point was reported for each of the following: retina, SRF, and CNV-PED complex. The CNV-PED complex thickness was defined as the sum of RPE thickness, RPEE thickness, and SHRM thickness because individual borders of these features were difficult to consistently delineate with accuracy (Fig 4). The sum of retinal thickness, SRF thickness, and CNV-PED complex thickness at the foveal center point defined the CATT quantitative OCT end point of total thickness at the foveal center point.

Vertical dimension measurements of retinal, SRF, and CNV-PED complex thickness were performed on all 6 radial line scans until April 2009. From April 2009 onward, measurements of vertical dimension of the retina, SRF, and CNV-PED complex thickness were performed on all 6 radial line scans for study visits at weeks 0, 4, 8, 12, 24, and 52. The mean thickness measurements averaged from scans 1 and 4 were determined by the CATT Coordinating Center to be approximately equal to mean thickness measurements derived from the average of 6 radial line scans (data not shown). Thus, to increase grading efficiency and minimize unnecessary measurements for the remaining year 1 CATT study visits submitted after April 2009, vertical dimensions were mea-

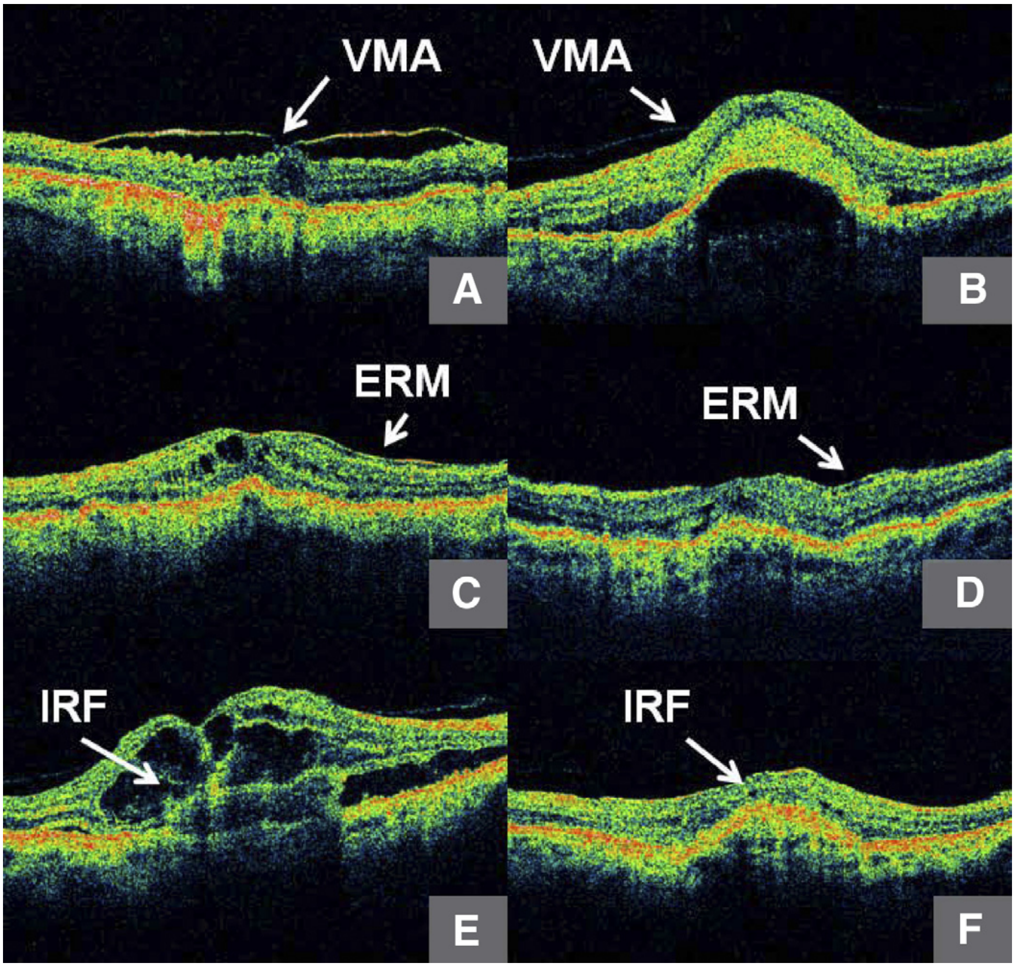

Figure 1. Representative morphologic features from optical coherence tomography images produced by the macular thickness map protocol. **A,** Obvious vitreomacular attachment (VMA) or vitreous attachment, and focal separation from the inner retina within a 3-mm diameter horizontal region centered at the middle of the fovea. **B,** Subtle VMA. **C,** Obvious epiretinal membrane (ERM). **D,** Subtle ERM. **E,** Obvious intraretinal fluid (IRF). **F,** Subtle IRF.

sured for retinal, SRF, and CNV-PED complex thickness on radial line scans 1 and 4 alone. Thickness measurements were performed manually on a standardized monitor at defined image size with a ruler and then converted to micrometers at the CATT Coordinating Center.

Team-Based Grading

Two masked readers individually graded all OCT scans in parallel. An independent data transcriptionist identified discrepant values between the paired readers. Morphometric data were considered discrepant if the vertical measurement differed by more than 65 μm or the horizontal measurement differed by more than 220 μm. The director of grading and senior readers established these values for horizontal and vertical measurement discrepancies after analysis of aggregated Stratus OCT grading data from a prior interventional study of eyes with exudative age-related macular degeneration. All graded scan pairs with discrepant data were then presented to a senior reader for arbitration. During the arbitration process, a senior reader reconciled all discrepancies between the initial reader pair. Any concordant reader grades that were deemed inaccurate by the senior reader were likewise corrected. Senior readers also reviewed all OCT scans for the presence of any macular fluid because this fluid was a study end point. Any finding or value that remained controversial after arbitration was forwarded to the director of grading for final decision.

Team Agreement Analysis

Grading reproducibility between several different pairs of reading center teams was analyzed on scans uploaded to the reading center between July 2009 and February 2010. From a subset of 274 scans randomly selected by computer from a comprehensive archive, 270 were available for reproducibility analysis. A pair of readers other than those who had performed the initial review and a senior reader who had not performed original arbitration performed the reproducibility grading. All new readers were masked to the results of the first reading team. The values obtained by the second reading team were then compared with those obtained by the first reading team (Fig 5, available at http://aaojournal.org). Of note, any morphologic feature graded not interpretable by 1 reading team and graded present or absent by another team was recorded as disagreement.

To test for grading drift over time, a subset of 23 scans uploaded during the initial portion of the CATT study underwent serial inter-team agreement analysis. These reproducibility studies

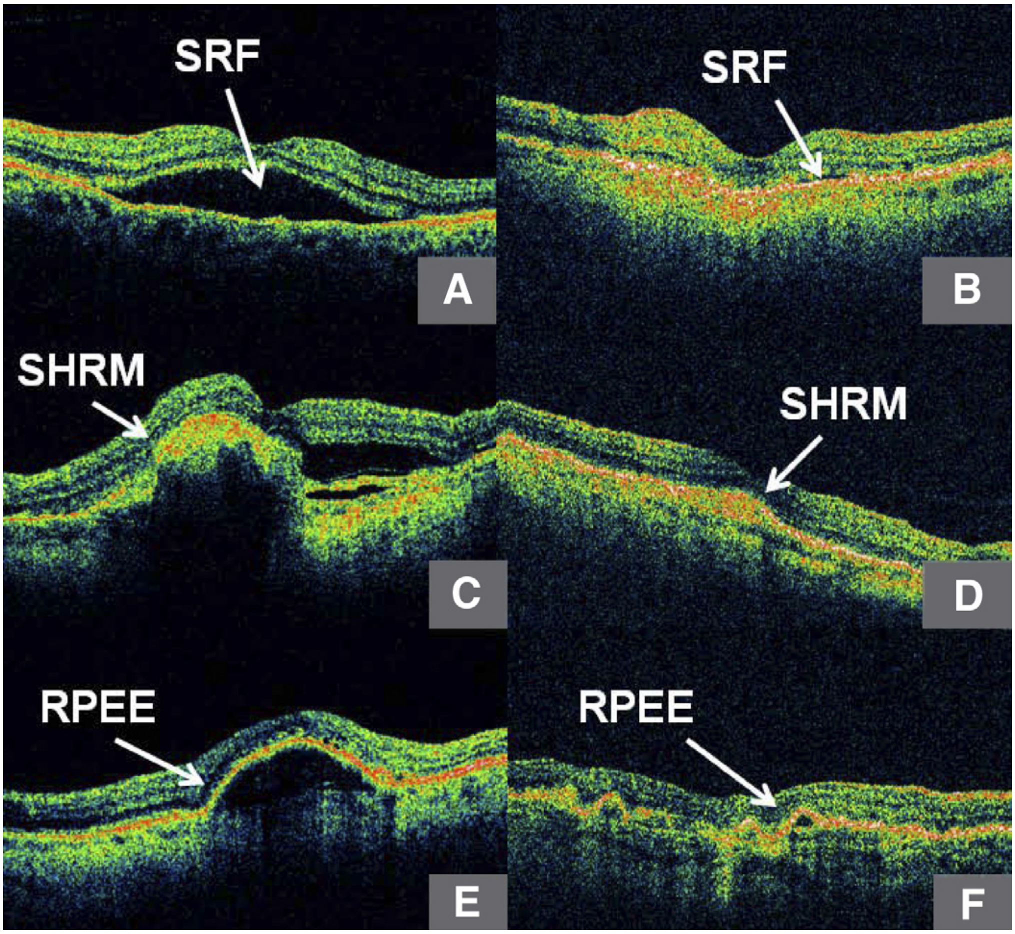

Figure 2. Representative morphologic features from optical coherence tomography images produced by the macular thickness map protocol. **A,** Obvious subretinal fluid (SRF). **B,** Subtle SRF. **C,** Obvious subretinal hyperreflective material (SHRM). **D,** Subtle SHRM. **E,** Obvious subretinal pigment epithelium elevation (RPEE). **F,** Subtle RPEE.

were performed at approximately 4- to 6-month intervals over the study duration.

Quantitative Intraretinal Fluid Analysis

From the 270 scans that underwent reproducibility analysis, both reading center teams agreed that IRF was present on 108 scans, and only a single reading center team reported IRF on 70 scans. To determine whether the single largest intraretinal cystoid hyporeflective cross-sectional area differed between these 2 groups, we performed a comparative analysis of cross-sectional area on 35 scans randomly selected from each group. All 6 images from the FMTM protocol and 6 images from the MTM protocol were reviewed to determine the largest horizontal and vertical dimensions from a single radial line image. Stratus software–based calipers were used to quantify the maximal horizontal and vertical dimensions of the single largest cross-sectional area of IRF for a specific scan. Cross-sectional area of single largest IRF was approximated as an ellipse using the following formula: area = Π × (horizontal dimension/2) × (vertical dimension/2). The sample size was calculated on the basis of the IRF area.

Quality Control

Several measures facilitated consistent analysis. First, all OCT scans were obtained by CATT-certified OCT technicians from OCT machines using standardized software packages (v. 4.0 or greater). Next, all data entry by transcriptionists into a centralized database was verified via an independent data entry team. Finally, ongoing monthly meetings ensured adherence to study grading protocols, addressed general discrepancies, and allowed for consensus opinion regarding controversial scans. It is worthwhile to note that only 1 scan of the 270 that underwent reproducibility analysis was discussed at a monthly meeting within 3 months of the actual reproducibility exercise.

Statistical Analysis

For categoric measures, the percent agreement (grading concordance between reading center teams) was computed to determine agreement. Percent agreement was computed as the number of concordant grading pairs divided by the total number of grading pairs multiplied by 100. Kappa statistics and respective 95% confidence intervals (CIs) were reported using the guidelines proposed by Koch et al[14] and Landis and Koch[15]: >0.80 = near perfect agreement, 0.61–0.80 = substantial agreement, 0.41–0.60 = good agreement, and 0.21–0.40 = fair agreement.

For assessing reproducibility of continuous measures, paired differences were computed. The mean (standard deviation) of the paired difference and 95% limits of agreement was calculated. The significance of the paired differences was assessed using the Wil-

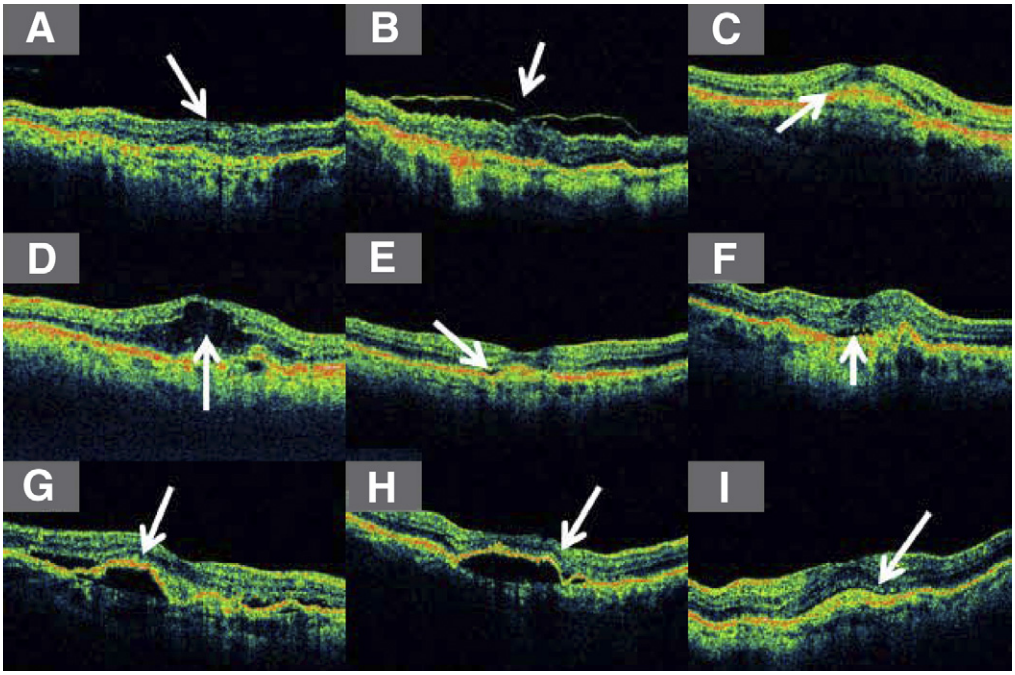

Figure 3. Examples of grading subcategories for morphologic features noted on optical coherence tomography images produced by the macular thickness map protocol. **A,** Epiretinal membrane (ERM) present (white arrow) with any deformation of the central 1 mm (horizontal dimension) of the retina. **B,** Vitreomacular attachment present (white arrow) with any deformation of the central 1 mm of the retina. **C,** Any intraretinal fluid (IRF) present (white arrow) within central 1 mm of the retina. **D,** IRF present (white arrow) at the foveal center point. **E,** Any subretinal fluid (SRF) present within central 1 mm of the retina. **F,** SRF present (white arrow) at the foveal center point. **G,** Any subretinal pigment epithelium (RPE) fluid present (white arrow) within central 1 mm of the retina. **H,** Sub-RPE fluid present (white arrow) at the foveal center point. **I,** Any subretinal hyperreflective material present (white arrow) within central 1 mm of the retina.

coxon signed-rank test of median difference equal to zero, and intraclass correlations were used to summarize the agreement of continuous measures.

A Wilcoxon rank-sum test of difference in medians was used to compare the difference in IRF area for eyes with concordant IRF grades with those with discordant grades. All analyses were performed with SAS 9.2 software (SAS Inc., Cary, NC).

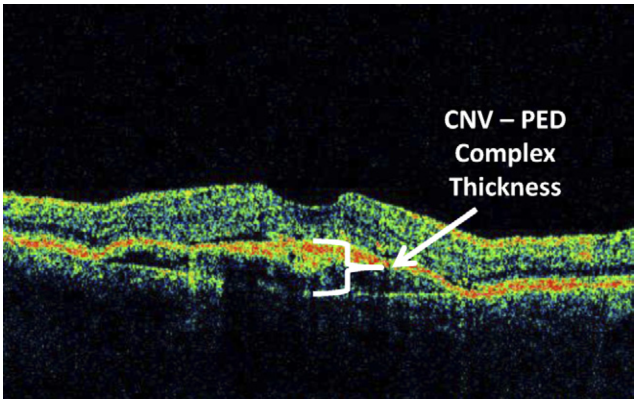

Figure 4. Representative scan demonstrating choroidal neovascularization (CNV)–pigment epithelial detachment (PED) thickness measurement at the foveal center point. The measurement was performed from the outer boundary of Bruch's membrane to the inner boundary of the CNV.

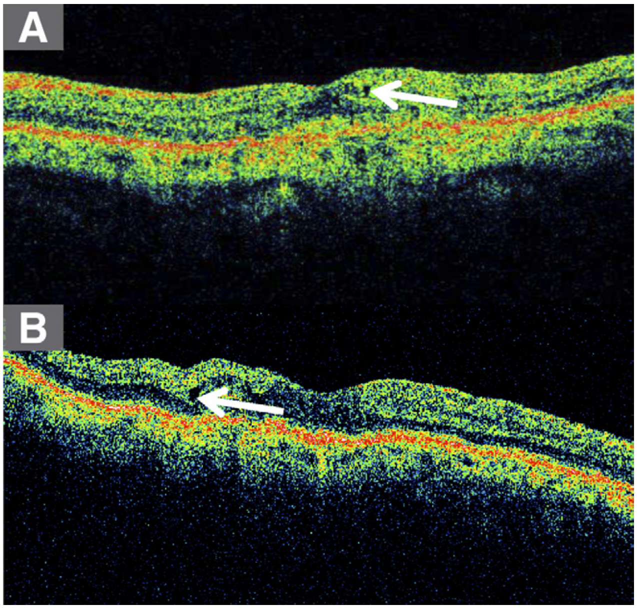

Figure 6. Comparison of the single largest intraretinal cystoid hyporeflective cross-sectional area from optical coherence tomography scans. **A,** Median single largest intraretinal cystoid hyporeflective cross-sectional area (white arrow) where only 1 reading team reported intraretinal fluid (IRF) as present. **B,** Median single largest intraretinal cystoid hyporeflective cross-sectional area (white arrow) for scans where both reading teams agreed on the presence of IRF.

Results

Reading Center Team Agreement

Table 1 (available at http://aaojournal.org) summarizes the grading agreement between reading center teams for evaluation of all OCT morphologic features in 270 OCT scans. Percent agreement between grading teams for macular fluid was 84%. Percent agreement for IRF, SRF, and sub-RPE fluid was 73%, 90%, and 88%, respectively. Independent reading center teams demonstrated good or better levels of agreement, based on kappa statistics, for grading of morphologic features. For IRF, SRF, and sub-RPE fluid, kappa statistics were 0.48, 0.80, and 0.75, respectively. The kappa statistic for macular fluid was 0.55.

Table 2 (available at http://aaojournal.org) details the agreement between reading center teams for all OCT quantitative measurements. For mean total thickness at the foveal center point, the intraclass correlation between reading center teams was 0.99 (95% CI, 0.99–0.99). For mean retinal thickness at the foveal center point, mean SRF thickness at the foveal center point, and mean CNV-PED complex (RPE + RPEE + SHRM) thickness at the foveal center point, the intraclass correlations between reading center teams were 0.93, 0.90, and 0.98, respectively. For total thickness at the foveal center point, the mean paired difference between reading center teams was −4 μm (95% limits of agreement, −55 to 47 μm). For mean retinal, SRF, and CNV-PED complex (RPE + RPEE + SHRM) thickness at the foveal center point, paired differences (95% limits of agreement in micrometers) between reading center teams were −3 μm (−62 to 56 μm), 0.6 μm (−27 to 28 μm), and −2 μm (−61 to 57 μm), respectively. The mean paired differences between reading center teams for all OCT measurements are shown in Table 2 (available at http://aaojournal.org).

Analysis of Temporal Drift Grading

Serial grading of a cohort of scans demonstrated comparable levels of inter-team agreement over time (Table 3, available at http://aaojournal.org). For macular fluid, percent agreement ranged from 78% to 83%. For IRF, SRF, and sub-RPE fluid, percent agreement was 57% to 70%, 83% to 100%, and 78% to 91%, and respectively.

For mean total thickness at the foveal center point, intraclass correlations between reading center teams over time were 0.97 at all 3 time points. For mean retinal thickness at the foveal center point, mean SRF thickness at the foveal center point, and mean CNV-PED complex (RPE + RPEE + SHRM) thickness, the intraclass correlations between reading center teams over time ranged between 0.95 and 0.97, 0.98 and 1.00, and 0.97 and 0.98, respectively. The agreement for each morphometric feature undergoing longitudinal analysis is shown in Table 4 (available at http://aaojournal.org). For mean total thickness at the foveal center point, the mean paired differences between reading center teams over time ranged between −10 and −3 μm. For mean retinal thickness at the foveal center point, mean SRF thickness at the foveal center point, and mean CNV-PED complex (RPE + RPEE + SHRM) thickness at the foveal center point, the mean paired differences between reading center teams over time ranged between −1 and −0.2 μm, 0 and 0 μm, and −9 and −1 μm, respectively. The mean paired measurement differences for all OCT measurements undergoing longitudinal analysis are shown in Table 4 (available at http://aaojournal.org).

Quantitative Intraretinal Fluid Analysis

The median single largest intraretinal cystoid hyporeflective cross-sectional area (median, 11.7×10^{-3} mm^2; range, $1.9–135.0\times10^{-3}$ mm^2) on 35 randomly sampled scans where reading center teams agreed on IRF presence was larger ($P = 0.001$) when compared with the median single largest intraretinal cystoid hyporeflective cross-sectional area (median, 5.5×10^{-3} mm^2; range, 1.3 – 570.8×10^{-3} mm^2) on 35 randomly sampled scans where only 1 reading center team graded IRF as present. Figure 6 shows representative images depicting the median single largest intraretinal cystoid hyporeflective cross-sectional area for scans where both reading center team agreed on the presence of IRF and where only 1 reading center team reported fluid.

Discussion

In this study, we have shown that well-trained reader teams in a reading center setting can reproducibly grade OCT qualitative and quantitative features in a large multicenter, randomized, interventional NVAMD treatment trial. Of the CATT OCT end point macular fluid variables, agreement was best for subretinal and sub-RPE fluid. Reproducibility was generally excellent for quantitative parameters. The reproducible results that we obtained resulted from rigorous reader certification requirements, collectively understood definitions of morphologic characteristics, and consistently applied quantitative measurement protocols.

We previously showed that OCT images generated from 132 eyes in an interventional NVAMD trial were reproducibly interpreted in a reading center setting.[16] In the present study, we observed 73% and 90% team grading agreement for IRF and SRF, respectively, comparable to the 84% to 85% and 90% to 91% inter-reader agreement for IRF and SRF, respectively, that we reported previously.[16] For total thickness measurement at the foveal center in the current work, we noted a median paired difference of 0 μm between teams, which was less than the 21 to 64 μm range of inter-reader median measurement differences reported in the previous study. These modest disparities may be due to differences in the trial enrollment criteria, OCT scan acquisition protocol, and grading methodology.

Our reading center has established a team-based grading approach that includes arbitration by a senior reader to maximize grading consistency during the study. This process also allows a senior reader to review a higher volume of scans and to establish a closed feedback loop with newer readers to enhance grading consistency. Prior series detailing OCT grading protocols have used individual readers[17,18] and paired readers in parallel,[16,19] whereas other large clinical trials using OCT grading by a reading center have not published detailed grading protocols.[20–22]

The “double grading” protocol for baseline fundus photographs used in the Early Treatment Diabetic Retinopathy Study (ETDRS) most resembles our team-based OCT scan grading protocol. During the ETDRS study, only baseline color fundus photographs underwent review by a pair of independent readers. One step of disagreement (of 3 possible steps in the ETDRS fundus photograph grading scale) was averaged together, and 2 steps or more of disagreement were returned to the initial graders for repeat evaluation. A masked ETDRS senior grader resolved any persistent disagreements. For subsequent study visits, a single reader alone evaluated follow-up fundus photographs, and grading was monitored using "haphazardly selected reading lists" of 10 eyes each.[23] Our image grading protocol differed in that an independent grading team evaluated both baseline and follow-up images, senior readers arbitrated all grading inconsistencies, and reproducibility studies were systematically performed on reading center teams. Although the ETDRS "double grading" has similarities to our team-based grading protocol, our evaluation methods more stringently address grading discrepancies and reproducibility.

Reading center grading was reproducible for morphologic features. Agreement was highest for SRF and less for IRF and ERM. Cystoid hyporeflective areas within the retina on OCT represent IRF from NVAMD.[24,25] However, a variety of factors may compromise IRF identification. There may be increased hyporeflective pixels within the retina, which in hyporeflective layers of the retina may have the appearance of small cystoid changes when none are actually present on scans with low signal intensity due to media opacity, low signal strength, or other factors. We have termed this finding a *pixel void*. Even the normal foveal center often appears slightly hyporeflective on OCT and can mimic subtle IRF, especially when coupled with decreased scan signal intensity. Finally, underlying active choroidal neovascular membranes may result in SRF at the CNV–retinal interface, making it difficult to discriminate IRF from SRF.

We found that the single largest cystoid hyporeflective area was smaller when only 1 reading center team reported fluid. It is not surprising that smaller true cystoid spaces are more challenging to grade consistently. These smaller areas of fluid are more difficult to differentiate from pixel voids than those with a larger cross-sectional area.

Epiretinal membranes can be difficult to visualize on Stratus OCT, especially when tractional changes are not visualized at the inner retina. In addition, a jagged, discontinuous inner retinal boundary that mimics an ERM can be seen when OCT image saturation is decreased. Hallmarks of ERM, such as focal points of attachment, optical reflectivity difference, and visible tufts or edges,[26] may not be visible on Stratus OCT during grading. One group reported a 30% increase in ERM detection rate when using ultrahigh resolution spectral domain (SD) OCT compared with Stratus OCT.[27]

Reading center teams demonstrated high levels of quantitative grading agreement. For all thickness measurements at the foveal center point, we observed relatively small mean paired thickness measurement differences less than 5 μm and high intraclass correlations between 0.90 and 0.99. For the trial end point, total thickness at the foveal center point, the mean (± standard deviation) of the paired difference was 3.9 ± 25.7 μm ($P = 0.025$). Although this difference was statistically significant, a reading team measurement difference of less than 4 μm is likely not clinically significant. These minimal differences and high levels of measurement agreement are especially notable in light of Bruch's membrane obscuration by overlying CNV or disruption of the RPE layer by CNV. These pathologic changes common to NVAMD can make accurate segmentation of the outer retina more difficult. To minimize these segmentation difficulties, our protocol aggregated RPE thickness, any RPEE, and SHRM thickness as a single measurement termed *CNV-PED complex* thickness.

Reading center teams also demonstrated excellent agreement when measuring maximal RPEE height (intraclass correlation = 0.97) and lower agreement when grading maximal RPE elevation width (intraclass correlation = 0.81). The heterogeneous changes induced by CNV in the subretinal space as visualized on OCT may partly account for the reduced reproducibility in grading RPE elevation width. For example, within an area of RPE elevation, CNV-mediated RPE fragmentation can make it difficult to consistently identify the exact separation point of the RPE from Bruch's membrane. In addition, overlying SHRM can sometimes obscure the borders of underlying RPE elevation. Finally, in scans with multiple adjacent RPE elevations, it can be challenging to confirm whether a single RPE elevation is discrete or contiguous with adjacent RPE elevations because of difficulty visualizing each potential point of RPE attachment to Bruch's membrane.

We evaluated reader agreement over time in a cohort of subjects followed from the initiation of CATT to monitor temporal grading drift. No obvious temporal drift was identified. We hypothesize that ongoing reader training and feedback during the study helped to minimize variations in reader grading over time.

Study Limitations

There are limitations to this study. The data were derived from a single reading center. Accordingly, reader reproducibility reported may not be readily generalized to other reading centers. Nonetheless, we believe that readers in other settings could adopt our team-based approach, with ongoing reader training and feedback and standardized grading protocols to produce reproducible grading data. Prior work demonstrated generally high levels of OCT grading agreement between independently trained reader pairs at 2 different reading centers.[19] Next, a senior reader reviewed all scans analyzed by primary readers for IRF, SRF, and sub-RPE fluid, key morphologic variables in the CATT. However, for other morphologic variables, if the grade assigned by the 2 primary readers was not discrepant, the senior reader did not necessarily review the scans. Accordingly, it is conceivable that if a variable was ascribed an identical inaccurate value by both primary readers, the senior reader might not correct the inaccuracy. However, we believe that these instances are likely rare and, for several reasons, would have minimal impact on the study results. First, a reader was not consistently matched with a particular second reader. Although 2 individuals may make similar grading errors, the likelihood of several readers all making an identical error for the same grading variable is small. Next, independent reading center teams showed high levels of agreement with one another. By discounting widespread and systematic biases across the entire reading center, the chances of 4 to 6 independent readers obtaining identical erroneous values for a particular finding is low. Finally, senior readers corrected erroneous values consistently reported by a reader pair if these values were determined to be inaccurate during arbitration. These scans were then returned to the reader pair for mandatory review to maintain grading consistency across readers.

During categoric grading analysis of all OCT morphologic features, we reported both percent agreement and kappa statistic in consideration of the innate limitations of this second analysis method. In particular, case distribution could result in high percent agreement but low values for kappa statistic. In the event that cases are common or rare, the kappa statistic can differ widely from percent agreement.[28,29] This phenomenon was apparent in this study for less commonly observed morphologic features, such as vitreomacular adhesion (94% agreement; kappa = 0.74) and ERM (95% agreement; kappa 0.53). The disparity between percent agreement and kappa statistic was more pronounced for the even less frequently observed grading variables vitreomacular adhesion with foveal deformation (82% agreement; kappa = 0.49) and ERM with foveal deformation (90% agreement; kappa = 0.46).

Future investigations will capitalize on the numerous advantages offered by SD-OCT technology. Compared with conventional time domain OCT, such as Stratus OCT used for this study, SD-OCT offers increased image resolution, improved registration, and faster data acquisition, resulting in decreased motion artifact.[30,31] These advantages may result in increased detection of important retinal features, such as IRF, SRF, and sub-RPE fluid.[32,33] If so, reader reproducibility may have been even higher than that reported in this study. An SD-OCT substudy has been initiated in CATT, and definitive answers to questions regarding reader reproducibility with SD-OCT when compared with time domain OCT will be forthcoming when the substudy has been completed.

In conclusion, because clinical studies for retinal diseases increasingly incorporate OCT to better understand treatment effect, reproducible analysis of imaging data is crucial to understand the efficacy of an intervention and to consistently evaluate an individual's response to therapy. This study demonstrates that reading center teams can reproducibly grade OCT images to facilitate monitoring of therapeutic effect in a large, prospective, multicenter, interventional treatment trial for NVAMD. A standardized training, grading, and feedback protocol can employ readers with differing levels of experience and obtain consistent results while maintaining quality over time.

References

1. Martin DF, Maguire MG, Ying GS, et al. Ranibizumab and bevacizumab for neovascular age-related macular degeneration. N Engl J Med 2011;364:1897–908.
2. Hee MR, Baumal CR, Puliafito CA, et al. Optical coherence tomography of age-related macular degeneration and choroidal neovascularization. Ophthalmology 1996;103:1260–70.
3. Jaffe GJ, Caprioli J. Optical coherence tomography to detect and manage retinal disease and glaucoma. Am J Ophthalmol 2004;137:156–69.
4. Ting TD, Oh M, Cox TA, et al. Decreased visual acuity associated with cystoid macular edema in neovascular age-related macular degeneration. Arch Ophthalmol 2002;120: 731–7.
5. Hee MR, Puliafito CA, Wong C, et al. Quantitative assessment of macular edema with optical coherence tomography. Arch Ophthalmol 1995;113:1019–29.
6. Rogers AH, Martidis A, Greenberg PB, Puliafito CA. Optical coherence tomography findings following photodynamic therapy of choroidal neovascularization. Am J Ophthalmol 2002; 134:566–76.
7. Kaiser PK, Blodi BA, Shapiro H, Acharya NR. Angiographic and optical coherence tomographic results of the MARINA study of ranibizumab in neovascular age-related macular degeneration. Ophthalmology 2007;114:1868–75.
8. Avery RL, Pieramici DJ, Rabena MD, et al. Intravitreal bevacizumab (Avastin) for neovascular age-related macular degeneration. Ophthalmology 2006;113:363–72 e5.
9. Rich RM, Rosenfeld PJ, Puliafito CA, et al. Short-term safety and efficacy of intravitreal bevacizumab (Avastin) for neovascular age-related macular degeneration. Retina 2006;26:495–511.
10. Gupta OP, Shienbaum G, Patel AH, et al. A treat and extend regimen using ranibizumab for neovascular age-related macular degeneration clinical and economic impact. Ophthalmology 2010;117:2134–40.
11. Dadgostar H, Ventura AA, Chung JY, et al. Evaluation of injection frequency and visual acuity outcomes for ranibizumab monotherapy in exudative age-related macular degeneration. Ophthalmology 2009;116:1740–7.
12. Rothenbuehler SP, Waeber D, Brinkmann CK, et al. Effects of ranibizumab in patients with subfoveal choroidal neovascularization attributable to age-related macular degeneration. Am J Ophthalmol 2009;147:831–7.
13. Fung AE, Lalwani GA, Rosenfeld PJ, et al. An optical coherence tomography-guided, variable dosing regimen with intravitreal ranibizumab (Lucentis) for neovascular age-related macular degeneration. Am J Ophthalmol 2007;143:566–83.
14. Koch GG, Landis JR, Freeman JL, et al. A general methodology for the analysis of experiments with repeated measurement of categorical data. Biometrics 1977;33:133–58.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
16. 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.
17. Krebs I, Hagen S, Brannath W, et al. Repeatability and reproducibility of retinal thickness measurements by optical coherence tomography in age-related macular degeneration. Ophthalmology 2010;117:1577–84.
18. 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;127:1461–7.
19. Ritter M, Elledge J, Simader C, et al. Evaluation of optical coherence tomography findings in age-related macular degeneration: a reproducibility study of two independent reading centres. Br J Ophthalmol 2011;95:381–5.
20. Glassman AR, Beck RW, Browning DJ, et al. 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: 560–6.
21. 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;117:1134–46 e3.
22. 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;117:1124–33 e1.

23. Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification. ETDRS report number 10. Early Treatment Diabetic Retinopathy Study Research Group. Ophthalmology 1991;98:786–806.
24. Ahlers C, Michels S, Elsner H, et al. Topographic angiography and optical coherence tomography: a correlation of imaging characteristics. Eur J Ophthalmol 2005;15:774–81.
25. Coscas F, Coscas G, Souied E, et al. Optical coherence tomography identification of occult choroidal neovascularization in age-related macular degeneration. Am J Ophthalmol 2007;144:592–9.
26. Wilkins JR, Puliafito CA, Hee MR, et al. Characterization of epiretinal membranes using optical coherence tomography. Ophthalmology 1996;103:2142–51.
27. Falkner-Radler CI, Glittenberg C, Hagen S, et al. Spectral-domain optical coherence tomography for monitoring epiretinal membrane surgery. Ophthalmology 2010;117:798–805.
28. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543–9.
29. Crewson PE. Reader agreement studies. AJR Am J Roentgenol 2005;184:1391–7.
30. 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 e1–14.
31. Wojtkowski M, Bajraszewski T, Gorczynska I, et al. Ophthalmic imaging by spectral optical coherence tomography. Am J Ophthalmol 2004;138:412–9.
32. Keane PA, Bhatti RA, Brubaker JW, et al. Comparison of clinically relevant findings from high-speed Fourier-domain and conventional time-domain optical coherence tomography. Am J Ophthalmol 2009;148:242–8 e1.
33. Sayanagi K, Sharma S, Yamamoto T, Kaiser PK. Comparison of spectral-domain versus time-domain optical coherence tomography in management of age-related macular degeneration with ranibizumab. Ophthalmology 2009;116:947–55.

Footnotes and Financial Disclosures

Originally received: March 28, 2012.
Final revision: May 26, 2012.
Accepted: June 25, 2012.
Available online: August 28, 2012. Manuscript no. 2012-447.

[1] Wills Eye Institute (Mid Atlantic Retina), Philadelphia, Pennsylvania.

[2] Duke University Eye Center, Durham, North Carolina.

Financial Disclosure(s):
The author(s) have no proprietary or commercial interest in any materials discussed in this article.

Funding: National Institutes of Health 5U10EYO17825. This work also was supported in part by the Heed Foundation.

Presented at: Association for Research in Vision and Ophthalmology, May 6–10, 2012, Ft. Lauderdale, Florida.

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