



OCT Signs of Early Atrophy in Age-Related Macular Degeneration: Interreader Agreement

Classification of Atrophy Meetings Report 6

Zhichao Wu, BAAppSc(Optom), PhD,^{1,2,*} Maximilian Pfau, MD,^{3,4,*} Barbara A. Blodi, MD,⁵ Frank G. Holz, MD,³ Glenn J. Jaffe, MD,⁶ Sandra Liakopoulos, MD,⁷ Srinivas R. Sadda, MD,⁸ Giovanni Staurenghi, MD,⁹ Elvira Bjelopera, MD,⁷ Tyler Brown,¹⁰ Petrus Chang, MD,³ John Choong, BS,⁶ Giulia Corradetti, MD,⁸ Federico Corvi, MD,⁸ Amitha Domalpally, MD, PhD,⁵ Cynthia Hurtenbach, MS,⁵ Muneeswar Gupta Nittala, MS,⁸ Anthony Olson, BS,¹⁰ Jeong W. Pak, PhD,⁵ Judith Pappe,⁷ Marlene Saßmannshausen, MD,³ Cindy Skalak, RN, COT,⁶ Sarah Thiele, MD,³ Robyn H. Guymer, MBBS, PhD,^{1,2,#} Steffen Schmitz-Valckenberg, MD^{3,10,#}

Purpose: To determine the interreader agreement for incomplete retinal pigment epithelium (RPE) and outer retinal atrophy (iRORA) and complete RPE and outer retinal atrophy (cRORA) and their related features in age-related macular degeneration (AMD).

Design: Interreader agreement study.

Participants: Twelve readers from 6 reading centers.

Methods: After formal training, readers qualitatively assessed 60 OCT B-scans from 60 eyes with AMD for 9 individual features associated with early atrophy and performed 7 different annotations to quantify the spatial extent of OCT features within regions of interest. The qualitative and quantitative features were used to derive the presence of iRORA and cRORA and also in an exploratory analysis to examine if agreement could be improved using different combinations of features to define OCT atrophy.

Main Outcome Measures: Interreader agreement based on Gwet's first-order agreement coefficient (AC_1) for qualitatively graded OCT features and classification of iRORA and cRORA, and smallest real difference (SRD) for quantitatively graded OCT features.

Results: Substantial or better interreader agreement was observed for all qualitatively graded OCT features associated with atrophy ($AC_1 = 0.63$ – 0.87), except for RPE attenuation ($AC_1 = 0.46$) and disruption ($AC_1 = 0.26$). The lowest SRD for the quantitatively graded horizontal features was observed for the zone of choroidal hypertransmission ($\pm 190.8 \mu\text{m}$). Moderate agreement was found for a 3-category classification of no atrophy, iRORA, and cRORA ($AC_1 = 0.53$). Exploratory analyses suggested a significantly higher level of agreement for a 3-category classification using (1) no atrophy; (2) presence of inner nuclear layer and outer plexiform layer subsidence, or a hyporeflective wedge-shaped band, as a less severe atrophic grade; and (3) the latter plus an additional requirement of choroidal hypertransmission of $250 \mu\text{m}$ or more for a more severe atrophic grade ($AC_1 = 0.68$; $P = 0.013$).

Conclusions: Assessment of iRORA and cRORA, and most of their associated features, can be performed relatively consistently and robustly. A refined combination of features to define early atrophy could further improve interreader agreement. *Ophthalmology Retina* 2022;6:4-14 © 2021 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Supplemental material available at www.opthalmologyretina.org.

Geographic atrophy (GA) is a common, late-stage disease manifestation of age-related macular degeneration (AMD), traditionally defined on color fundus photography and more recently on fundus autofluorescence imaging. Geographic atrophy as defined on both of these imaging methods is

accepted as a primary efficacy end point in nonneovascular AMD clinical trials.^{1–3}

Although no effective treatment currently exists to halt or slow the progression of GA lesions, the number of trials that test potential novel interventions have recently increased

rapidly, and several Phase 2 and 3 trials are currently underway.^{4–8} However, although a highly sought-after goal is to intervene at an earlier point in the disease process, very few clinical trials are designed to prevent the progression of intermediate AMD to GA. The feasibility and appetite to evaluate early interventions has been limited by the necessity for large and lengthy clinical trials because of the slowly progressive nature of AMD. However, this limitation could be addressed by establishing robust and reliable early clinical biomarkers of atrophy development. Such early atrophic features could identify individuals at a higher risk of progression, providing enriched populations to reduce the size and costs of such trials, providing new, earlier trial efficacy end points, or both. These features will also enable better risk stratification for progression in clinical practice.

Recent advances in multimodal imaging have provided an extraordinary opportunity to define AMD stages in more granular detail than has been previously possible. Enormous clinical research efforts have been underway to characterize anatomic features, especially those seen on OCT B-scans, that better depict disease severity and confer an increased risk of progression to sight-threatening late-stage disease.^{9–15} A summary of many of these features has been described in the Classification of Atrophy Meetings (CAM) Report 5 publication.¹⁶ In addition, characteristic OCT changes seen in the outer retina, which occur before the development of GA, have also been described.^{17,18}

The CAM group, an international group of AMD and retinal imaging experts whose aim is to arrive at consensus around new multimodal imaging definitions of AMD, have worked to provide a naming framework for these anatomic features that will provide a consensus nomenclature to help unify the field as it moves forward. The CAM group recommended that the atrophic stages of AMD should be named according to the affected anatomic layers on OCT.^{19–21} As such, the terms *complete retinal pigment epithelium (RPE) and outer retinal atrophy (cRORA)* and *incomplete RPE and outer retinal atrophy (iRORA)* were proposed as terms representing a combination of early OCT features of retinal cell death in eyes with drusen. Complete RPE and outer retinal atrophy was defined by the following criteria: (1) a region of choroidal hypertransmission of 250 μm or more in diameter; and (2) a zone of attenuation or disruption of the RPE of 250 μm or more in diameter; and (3) evidence of overlying photoreceptor degeneration, all occurring in the absence of signs of an RPE tear.²⁰ The term *iRORA* was introduced to describe a stage of AMD at which these OCT signs are present but do not fulfill all the size criteria for cRORA.¹⁹

These OCT-defined terms were intended to allow the AMD clinical and research community, as well as industry and regulatory agencies, to become familiar with the features that were associated with the development and initial progression of AMD-associated atrophy. The intent of the CAM group was for these OCT features to provide the backbone of a new, more refined grading system for AMD and provide additional biomarkers that could facilitate early intervention clinical trials in the early stages of AMD.

However, the CAM publications stressed that further work was required to determine, through a rigorous and

robust process, whether cRORA and iRORA, as well as the features used to define them, could be determined reliably in a reading center setting before they could be put forward as potential biomarkers of disease severity and structural clinical trial end points. Accordingly, it is necessary to determine the ability of readers to identify and, where relevant, reliably measure the features associated with these anatomic changes if they are adopted into AMD clinical trials.

We therefore trained readers at 6 established reading centers on the anatomic features that define cRORA and iRORA and thereafter assessed the value of this training. We then assessed the level of agreement among these readers in the qualitative evaluation of 9 features and quantitative measurements of 7 features associated with early atrophy. The results of this report provide a strong foundation to move forward when considering the requirements for more granular AMD grading schemes, robust biomarkers of disease severity, and OCT-defined atrophy end points.

Methods

Six established reading centers were invited to participate in this exercise to examine interreader agreement for features associated with atrophy. From the United States, these included the Doheny Imaging Reading Center, Duke Reading Center, Utah Retinal Reading Center, and Wisconsin Fundus Photograph Reading Center. Further, the Cologne Imaging Reading Center and Laboratory, Cologne, and GRADE Reading Center, Bonn, both in Germany, participated. At each reading center, the medical directors were asked to assign 2 experienced readers ($n = 12$ in total from all sites) to participate in the study, with the selection left to the medical directors' discretion. Before performing the actual grading task, readers participated in a formal and standardized training on iRORA and cRORA definitions and how to identify these features on OCT B-scans.

Reader Training

First, readers were asked to read the 2 previous CAM Reports 3 and 4.^{19,20} Second, they completed a web-based pretest that assessed their general knowledge about the definition and features of atrophy based on the CAM articles. Further, readers were also asked to indicate the presence of the following 7 features associated with atrophy on OCT B-scans in test cases: (1) region of choroidal signal hypertransmission; (2) zone of disruption or attenuation of the RPE; (3) ellipsoid zone (EZ) disruption; (4) external limiting membrane (ELM) disruption; (5) outer nuclear layer thinning; (6) outer plexiform layer (OPL) subsidence; and (7) hyporeflective wedge-shaped band within the limits of Henle's fiber layer. Third, 2 senior authors (R.H.G. and S.S.-V.) gave 2 web-based tutorials, each 1 hour long, to the readers (1 for United States-based reading centers and 1 for European-based reading centers), during which they reviewed and discussed 20 cases to illustrate the presence of features associated with iRORA or cRORA. These web conferences provided the readers an opportunity to clarify the definitions of each feature and discuss questions that arose from the pretest. Readers were then asked to perform the pretest again to examine changes in their assessments that were informed by these web conferences.

Grading Task

After the training was completed, all 6 reading centers received de-identified imaging data that contained combined near-infrared reflectance and spectral-domain OCT volume scan data (Spectralis HRA+OCT; Heidelberg Engineering GmbH, Heidelberg, Germany) of 60 eyes of 60 individuals with AMD. These imaging data had been obtained as part of natural history studies undertaken at the Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia, and the Department of Ophthalmology, University of Bonn, Bonn, Germany. These studies adhered to the tenets of the Declaration of Helsinki and were approved by the local ethics committees. Written informed consent was obtained from each participant before enrollment.

For each case, the region of interest to be evaluated was outlined on a single B-scan and provided to the readers as an image file. Readers were also provided with the entire corresponding OCT volume scan that could be reviewed using the review software of the imaging device (Heidelberg Eye Explorer; Heidelberg Engineering). The image samples from the 60 eyes (selected by 2 authors [Z.W. and S.S.-V.]) illustrated a variety of anatomic precursors and early AMD-related atrophic changes; approximately one third each consisted of cRORA, iRORA, and drusen lesions without any signs of atrophy. Note that the readers were masked to the severity distribution of the cases selected. These cases formed an image set that were to be graded in what we refer to as the *grading task*.

Each reader was provided with a grading manual that contained comprehensive predefined grading instructions as a guide to the grading task. Readers were asked to grade the preselected single B-scan image from each of the eyes. At the discretion of the reader, the entire OCT volume scan could also be reviewed. For each region of interest, readers were asked to determine the presence of 9 different qualitative features associated with atrophy, as outlined in [Supplemental Figure 1](#) (available at www.opthalmologyretina.org). Compared with the pretest, subsidence of the inner nuclear layer (INL) was added to OPL subsidence, and the attenuation and disruption of the RPE were assessed as 2 separate features. Response options for each feature included “yes,” “no,” “questionable,” or “cannot grade.”

Readers were also asked to annotate 7 different features to assess their ability to quantify the spatial extent of different features associated with atrophy, if present ([Supplemental Fig 2](#), available at www.opthalmologyretina.org). These annotations were used to define the 250- μ m cutoff in extent of choroidal hypertransmission and RPE attenuation or disruption that differentiated iRORA from cRORA. A custom-written tool in FIJI (an extension of the image processing software ImageJ, available at <http://fiji.se/>; National Institutes of Health, Bethesda, MD) was used to annotate the images.

Statistical Analysis

For the first part of the pretest (assessment of general knowledge about the definitions and features of atrophy), appropriate responses were determined based on a consensus among 3 senior authors (S.R.S., R.H.G., and S.S.-V.). The difference in the proportion of correct responses by each of the 12 readers before and after the training web conference was determined using a mixed-effects logistic regression model (to account for the repeated-measures nature of the data and correlation between the responses to multiple questions for a given reader). For the second part of the pretest (identification of features based on provided OCT B-scan data), the responses were binarized into 2 categories: (1) absent or questionable or (2) definitely present. The differences between the interreader agreement among the 12 readers before

and after the web conference was determined based on Gwet's first-order agreement coefficient (AC_1),²² with statistical significance determined using a bootstrap procedure ($n = 1000$ resamples). The AC_1 was used in this study over the widely used κ coefficient because the latter exhibits a well-known limitation of being affected by the prevalence of the trait being measured. For instance, κ coefficients can have low values despite high levels of agreement when the prevalence of the trait is either very high or low.²³ The AC_1 instead is a robust, chance-corrected measure of agreement that is more stable for variations in trait prevalence.²² The AC_1 values were interpreted as follows: poor agreement (< 0), slight agreement (0–0.20), fair agreement (0.21–0.40), moderate agreement (0.41–0.60), substantial agreement (0.61–0.80), and almost perfect agreement (0.81–1.0).²⁴ A P value of less than 0.05 was considered statistically significant.

For the grading task (consisting of the $n = 60$ cases), the qualitative grading for each of the features associated with atrophy was also binarized into the 2 categories described above. The prevalence of these features was determined based on the percentage of readings of whether a feature was present. The AC_1 statistic and percentage of readings in agreement was calculated to assess the interreader agreement for each of these 9 features evaluated.

For the quantitative parameters, the smallest real difference (SRD; also termed *repeatability coefficient*) was calculated for each feature.²⁵ The SRD denotes the upper limit (95% probability) for the absolute differences between 2 measurements attributable to chance and is derived from the average interreader variability within each B-scan as: $1.96 \times \sqrt{2} \times var_w$. To allow for comparison among SRD values, nonparametric bootstrap confidence intervals were generated. Intraclass correlation coefficients (2-way random, single measures, absolute agreement) were also calculated to assess the level of agreement for the quantitative features. To assess whether the measurement error increased with larger absolute values, Spearman's ρ was calculated to assess the relationship between the interreader variability and mean measurement across all cases.

Results from qualitative and quantitative grading parameters of the main test were also used to determine the presence of iRORA and cRORA in each case. The number of cases meeting these definitions was also presented based on the consensus response of the 12 readers. The interreader agreement between a 2-category classification of the atrophic grade (i.e., presence or absence of at least iRORA or presence or absence of cRORA) and 3-category classification (i.e., no atrophy vs. iRORA vs. cRORA) was also assessed using the AC_1 statistic.

Results

Pretest Results

The percentage of correct responses of the 12 readers improved slightly, but not statistically significantly, after the web-based tutorials for the first part assessing knowledge about the definitions and features of atrophy ($n = 18$ items; range, 81%–83%; $P = 0.522$). For the second part, involving the actual assessment of the OCT features associated with atrophy ($n = 98$ items), the interreader agreement improved significantly after the web conference (from $AC_1 = 0.59$ to $AC_1 = 0.79$; $P = 0.004$).

Grading Task Results

[Table 1](#) presents the prevalence of each feature examined based on the consensus among the 12 readers, highlighting how choroidal

Table 1. Prevalence of Features Associated with Atrophy among the Cases Evaluated in This Study (n = 60) and Interreader Agreement among the 12 Readers

Feature Present	Reading Prevalence		Agreement Rate (%)	Gwet's First-Order Agreement Coefficient (AC ₁)
	Yes (%)	No (%)		
Choroidal hypertransmission	66	34	80	0.63
RPE attenuation	71	29	68	0.46
RPE disruption	44	56	63	0.26
EZ disruption	89	11	89	0.87
ELM disruption	87	13	87	0.83
ONL thinning	86	14	81	0.75
OPL subsidence	60	40	82	0.65
INL subsidence	47	53	85	0.70
Hyporeflective wedge-shaped band*	27	73	86	0.77

ELM = external limiting membrane; EZ = ellipsoid zone; INL = inner nuclear layer; ONL = outer nuclear layer; OPL = outer plexiform layer; RPE = retinal pigment epithelium.

*Present within Henle's fiber layer.

hypertransmission, RPE attenuation, EZ and ELM disruption, and outer nuclear layer thinning were present in approximately two thirds or more of the cases. It also showed that RPE disruption and OPL and INL subsidence were present in approximately half of the cases, whereas the hyporeflective wedge-shaped band were present in approximately one quarter of cases.

All features showed substantial interreader agreement (AC₁ = 0.63–0.87), except for RPE disruption (AC₁ = 0.26) and attenuation (AC₁ = 0.46). These findings are also summarized in Table 1. Supplemental Figure 3 (available at www.opthalmologyretina.org) shows 1 example illustrating challenges in the determination of the zone of RPE attenuation and ELM disruption.

Agreement and Reliability of Quantitative Image Annotation

The SRD varied markedly among the horizontal features. The SRD estimates were the smallest for choroidal hypertransmission ($\pm 190.8 \mu\text{m}$ [95% confidence interval (CI), 130.8–272.6 μm]), followed by ELM disruption ($\pm 745.2 \mu\text{m}$ [95% CI, 593.1–929.9 μm]), then by EZ disruption ($\pm 869.1 \mu\text{m}$ [95% CI, 763.5–1017.6 μm]), and finally by RPE disruption ($\pm 1094.9 \mu\text{m}$ [95% CI, 930–1264 μm]; Fig 1). The SRD estimates for the vertical measurements were more uniform with $\pm 56.8 \mu\text{m}$ (95% CI, 44.5–72.8 μm) for the outer retinal thickness at the left margin, $\pm 46.8 \mu\text{m}$ (95% CI, 34.8–58.8 μm) for the outer retinal thickness at the right margin, and $\pm 54.8 \mu\text{m}$ (95% CI, 41.4–76.9 μm) for the outer retinal thickness at the thinnest location (Supplemental Fig 4, available at www.opthalmologyretina.org).

With adjustment for multiple testing, a significant relationship was observed between the measurement variability (variance between readers) and lesion size (mean value across readers) for RPE disruption (Spearman's $\rho = 0.56$; $P = 0.01$), EZ disruption ($\rho = 0.40$; $P = 0.04$), and ELM disruption ($\rho = 0.65$; $P < 0.001$) but not for choroidal hypertransmission ($\rho = 0.31$; $P = 1.0$). Further, analysis of the intraclass correlation coefficient, which considers measurement variability in relationship to the underlying variability of the lesion of interest in this cohort, revealed that measurements of choroidal hypertransmission showed a much better interreader agreement compared with the other horizontal lesions

(Supplemental Table 1, available at www.opthalmologyretina.org).

Interreader Agreement for Derived Presence of Incomplete Retinal Pigment Epithelium and Outer Retinal Atrophy and Complete Retinal Pigment Epithelium and Outer Retinal Atrophy

The qualitatively and quantitatively graded parameters were combined to derive the presence of iRORA or cRORA. Moderate interreader agreement (AC₁ = 0.58) was found for a 2-category classification of the atrophic features based on the presence or absence of at least iRORA, and substantial agreement (AC₁ = 0.68) was found based on the presence or absence of cRORA. Also moderate agreement (AC₁ = 0.53) was found for a 3-category classification of atrophy (no atrophy vs. iRORA vs. cRORA). These findings are summarized in Supplemental Table 2 (available at www.opthalmologyretina.org), along with the prevalence of each atrophic classification stage. Figure 2 shows 3 representative examples with full consensus for the different classification stages of OCT-defined atrophic AMD.

The impact of the region-of-interest retinal location was also examined, because anecdotally concern has arisen regarding some of the OCT features associated with atrophy being more difficult to identify close to the fovea, given the difference in anatomic features in this region. Eight of the 60 cases (13%) had the region of interest falling within the central 1000 μm of the foveal center. Statistical analysis revealed no difference in the level of interreader agreement for a 3-category classification of atrophy (no atrophy vs. iRORA vs. cRORA) between cases within (AC₁ = 0.47) and outside (AC₁ = 0.54) 1000 μm of the foveal center ($P = 0.743$). Supplemental Figure 5 (available at www.opthalmologyretina.org) shows 1 example demonstrating a high level of agreement for the determination of atrophic features for a region of interest at the foveal center.

Exploratory Analyses

Based on the findings of the qualitative and quantitative graded results, we explored additional combinations of features and measurement cutoffs to define an alternative staging system for

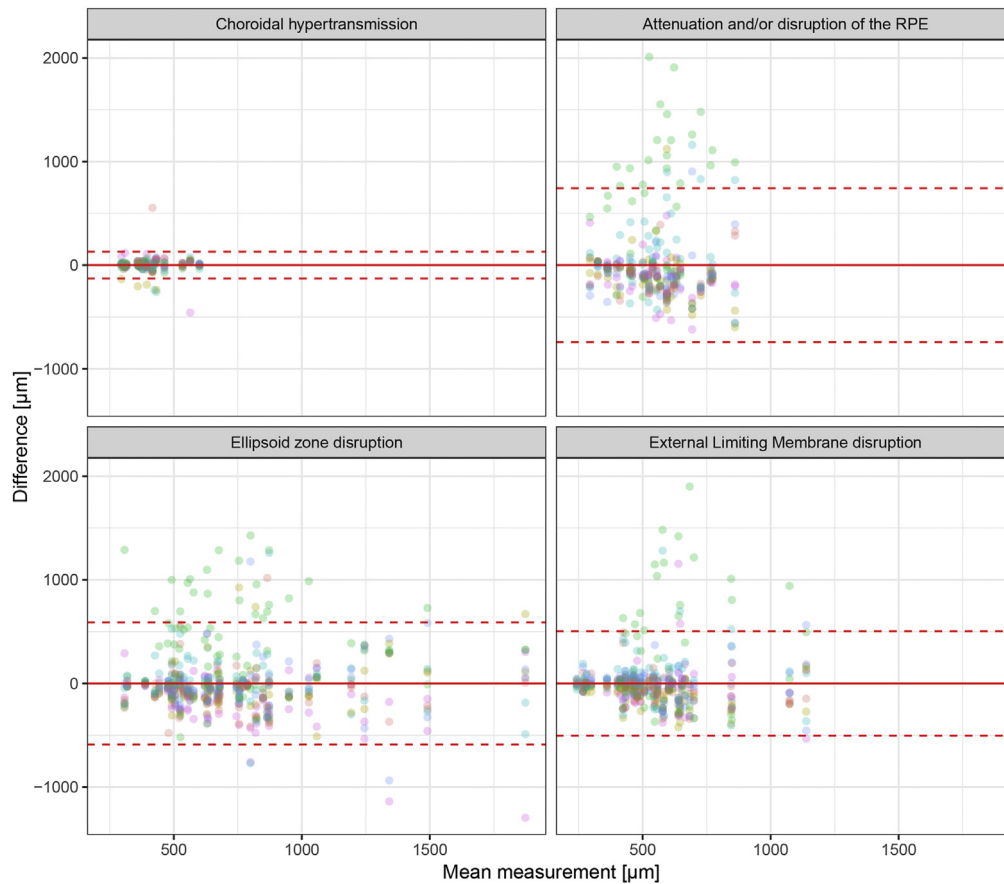


Figure 1. Bland-Altman plots showing the interreader agreement for the 4 horizontal quantitative imaging features. The y-axis shows the difference from the mean measurement (of all 12 readers) and the x-axis shows the mean measurement. Different reading centers are highlighted by different colors. The solid red lines denote the mean difference, and the dashed red lines denote the 95% limits of agreement with the mean for all readers. All points were plotted semitransparently to illustrate each point. Notably, the zone of choroidal hypertransmission constituted the most reliable horizontal measure of disease extent across, both overall and on a reading center level. Note that the 95% limits of agreement with the mean (i.e., agreement between an individual reader and mean values across readers) describe the upper limit of measurement variability of an individual reader. Hence, the smallest real difference estimates shown in Supplemental Table 1, which describe the upper limit for the absolute difference between 2 individual readers, are higher by a factor of $\sqrt{2}$.

OCT atrophy to determine if interreader agreement could be improved (Supplementary Table 3, available at www.ophtalmologyretina.org).

Considering the features indicative of photoreceptor loss, we considered a classification of early atrophy that required either (1) only EZ or ELM disruption or (2) only the subsidence of the OPL and INL, or hyporeflective wedge-shaped band, but did not require other features to define photoreceptor loss. When these features of photoreceptor degeneration were combined with the definitive presence of RPE attenuation or disruption, choroidal hypertransmission, or both to consider the presence or absence of at least a less severe grade of OCT atrophy, nonsignificant improvements in interreader agreement were achieved ($AC_1 = 0.58$ – 0.67 ; all $P > 0.05$) compared with the original features required for at least iRORA presence or absence ($AC_1 = 0.58$). A larger improvement in agreement among readers was seen if only the presence of subsidence of the OPL and INL, or a hyporeflective wedge-shaped band, was used to consider the presence or absence of at least a less severe grade of OCT atrophy ($AC_1 = 0.75$; $P = 0.010$).

In another exploratory analysis, we also examined whether defining photoreceptor loss based on the 2 approaches above, as well as excluding the requirement for RPE attenuation or disruption, choroidal hypertransmission, or both can improve interreader agreement when considering the presence or absence of a more severe grade of OCT atrophy in a 2-category classification. When compared with a classification based on the presence or absence of cRORA, interreader agreement was not significantly different for all alternative combinations ($AC_1 = 0.45$ – 0.71 ; all $P > 0.05$), but the highest agreement was observed for the combination based on the presence of (1) choroidal hypertransmission of $250 \mu\text{m}$ or more and (2) the subsidence of the OPL and INL, or the hyporeflective wedge-shaped band ($AC_1 = 0.71$).

When considering 2 different combinations for an earlier and later grade of the atrophic features in a 3-category classification scheme (no atrophy vs. stage 1 based on presence of subsidence of the OPL and INL, or hyporeflective wedge-shaped band, vs. stage 2 based on stage 1 plus choroidal hypertransmission of $\geq 250 \mu\text{m}$), the interreader agreement ($AC_1 = 0.68$) was significantly better compared with a 3-category classification scheme based on no

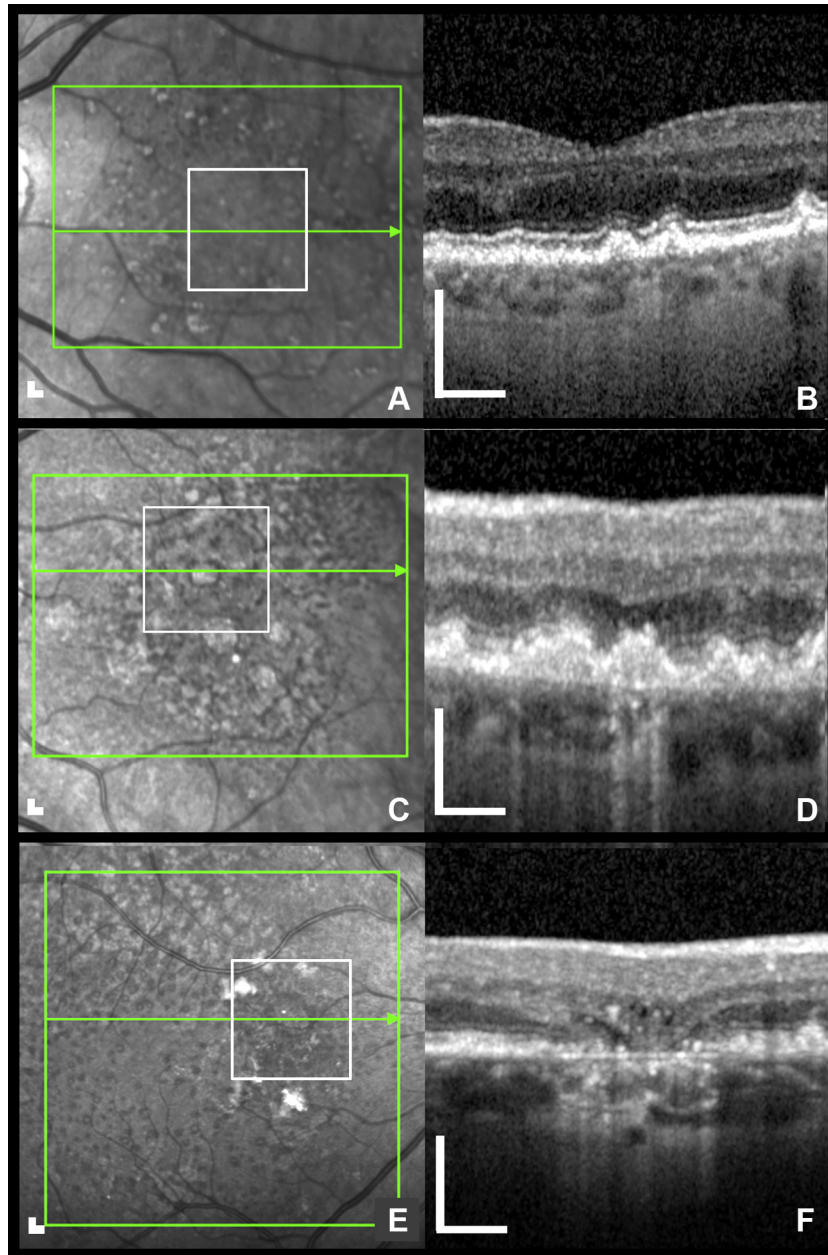


Figure 2. Three examples showing excellent agreement among readers for the absence or presence of atrophic features according to Classification of Atrophy Meetings definitions, with the near-infrared confocal scanning ophthalmoscopy image and corresponding spectral-domain OCT B-scan at the region of interest (green line through the white box) shown for each example. The white scaling bars at each left lower corner represent 250 μm in both lateral and vertical dimensions. **A, B**, Full consensus for absence of incomplete retinal pigment epithelium (RPE) and outer retinal atrophy (iRORA) and complete RPE and outer retinal atrophy (cRORA) among all readers (“no atrophy”). At the same time, a relatively high level of disagreement occurred for the individual feature of disruption of the ellipsoid zone: 7 readers determined that it was present, and 5 readers believed that it was absent. **C, D**, Full consensus for presence of iRORA among all readers. Highest disagreement occurred for the individual feature of inner nuclear layer subsidence; an equal number of readers considered this feature to be present or absent. **E, F**, Full consensus for presence of cRORA among all readers, with high agreement for the presence of all individual atrophic features.

atrophy versus iRORA versus cRORA ($AC_1 = 0.53$; $P = 0.013$; Supplemental Table 3). A similar 3-category classification scheme based on the 2 different combinations above, except requiring (1) EZ or ELM disruption rather than (2) subsidence of OPL and INL, or hyporeflective wedge-shaped band, for defining photoreceptor degeneration, performed similarly to the classification based on no

atrophy versus iRORA versus cRORA ($AC_1 = 0.57$; $P = 0.349$; Supplemental Table 3).

Figure 3 shows 2 representative examples with low agreement when using iRORA and cRORA, but full consensus among readers when only the features of subsidence of the OPL and INL, or a hyporeflective wedge-band, was included.

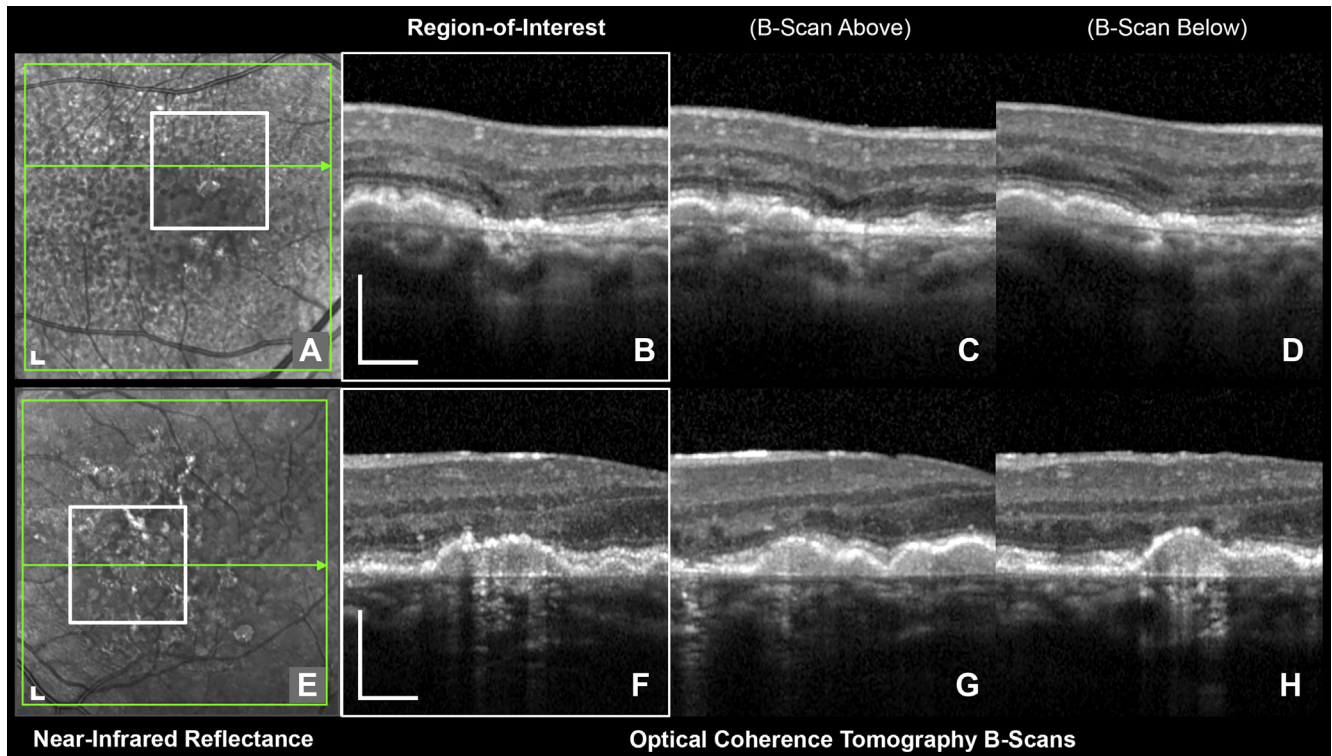


Figure 3. Two examples illustrating the lack of consensus among readers when applying the Classification of Atrophy Meetings definitions, with the (A, E) en face near-infrared reflectance (NIR) image, (B, F) OCT B-scan obtained through the region of interest (location on the NIR image indicated by the green line through the white box), and (C, D and G, H) its 2 neighboring B-scan sections (distance between individual B-scans is approximately 120 μm) shown. The white scaling bars at each left lower corner of the NIR and OCT B-scan of the region of interest represent 250 μm in both lateral and vertical dimensions. For the first example (A–D), 6 of 12 readers determined that choroidal hypertransmission and retinal pigment epithelium (RPE) attenuation or disruption was either present or absent. This disagreement on choroidal hypertransmission resulted in a lack of consensus in the determination of the presence of an atrophic stage derived from the qualitative and quantitative gradings, because incomplete RPE and outer retinal atrophy (iRORA) would have been considered to be present by 6 readers and absent by the other 6 readers. In contrast, the atrophic grade including only the presence of outer plexiform layer (OPL) and inner nuclear layer (INL) subsidence or the presence of a hyporeflective wedge-shaped band reached full consensus. For the second example shown on the right (E–H), 3 readers would not have identified both iRORA and complete RPE and outer retinal atrophy (cRORA) when these atrophy grades were derived from their combined qualitative and quantitative grading (triggered by grading the absence of choroidal hypertransmission), whereas another 3 readers would have identified iRORA and 6 readers would have identified cRORA. The disagreement between presence of iRORA and cRORA was caused by the determination of the diameter of the zone of choroidal hypertransmission. Although some readers concluded that the zone was continuous with a diameter of 250 μm or more, 3 others identified 2 separate zones, leading to a measurement of less than 250 μm . Based again on an atrophic grade derived from only the presence of OPL and INL subsidence, or the presence of a hyporeflective wedge-shaped band, full consensus of absence of atrophy was reached.

Note also that for the above-mentioned 3-category classification scheme based on the different combination of OCT features, there was no evidence of a significant difference in agreement between cases that were within ($AC_1 = 0.76$) and outside ($AC_1 = 0.71$) 1000 μm of the foveal center ($P = 0.282$).

Discussion

This study showed a moderate to substantial level of agreement ($AC_1 = 0.53$ – 0.68) among 12 readers for the assessment of 2- and 3-category classification schemes of atrophic grades based on iRORA and cRORA criteria. These findings indicate that evaluation of high-resolution OCT B-scans allow for a relatively consistent and robust assessment of structural changes in early atrophic AMD in a reading center setting. This study also revealed that all

features associated with atrophy showed substantial or better agreement, except for the assessment of RPE attenuation or disruption. The exploratory analyses also suggest that substantial agreement for the 2- and 3-category classification of atrophic features ($AC_1 = 0.68$ – 0.75) could be achieved by using alternative combinations of features for defining atrophy where photoreceptor loss was defined based on the presence of subsidence of the OPL and INL, or a hyporeflective wedge-shaped band.

The ability to detect structural changes in features seen on OCT imaging likely depends in large part on the training, background, and experience of readers and the robustness of the definitions and guidelines for their interpretation. Thus, we sought to minimize these confounding factors by performing a formal and intensive training, selecting 12 readers from 6 established reading centers, and providing detailed instructions for performing the grading task. The observed

improvement of interreader agreement in the pretest after formal and intensive training (to ensure that readers understood and were able to apply the iRORA and cRORA definitions, including the grading of their associated features) highlights the value and importance of such training. This also supports the anecdotal feedback we received from the readers on the helpfulness of the virtual web conferences to clarify of the grading strategy.

After reader training, in the grading task, we observed substantial agreement for all qualitative features associated with atrophy, except for the assessment of RPE attenuation or disruption. These observations are consistent with feedback by readers during the web conferences, at which the edges of RPE disruption and attenuation were often challenging to define. These observations are supported by the quantitative grading results, which showed that RPE disruption or attenuation showed the lowest level of agreement.

The results from the quantitative grading also revealed that determination of the zone of choroidal hypertransmission was the most robust quantitative horizontal feature, and the level of agreement did not vary based on the average absolute size (e.g., smaller areas of hypertransmission did not show poorer interreader agreement than larger areas). This level of performance was observed despite the fact that the continuity of hypertransmission may sometimes be interrupted, particularly by blockage of material anterior in the light path (Fig 3E–H), resulting in a so-called “barcode” appearance. This performance level may have been achieved because we included clear examples and descriptions in the grading instructions of how to measure the extent of hypertransmission when such an appearance was present. Nonetheless, we observed in a minority of cases within the atrophic disease spectrum that it was difficult for readers to detect and measure choroidal hypertransmission consistently. Despite these exceptions, the relatively robust nature of this quantitative measurement could provide a potentially sensitive and meaningful biomarker to assess the progression of atrophic lesions. The potential usefulness of this measure is supported further by ours and others’ previous findings that the extent of choroidal hypertransmission is associated highly with the extent of atrophy measured on fundus autofluorescence imaging,^{26–28} which is currently accepted as an anatomic outcome measure by international regulatory agencies.²⁹ Strategies that could improve the reliability of these measurements to explore in the future include: (1) requiring a minimum of 2 neighboring B-scans meeting the size criterion, instead of relying on single B-scan quantification for the zone of choroidal hypertransmission; (2) requiring 2 independent measurements of 2 readers within a certain tolerance level; or (3) using artificial intelligence-based automated image analysis methods to improve consistency in quantification of the choroidal hypertransmission zone.

We combined the qualitatively and quantitatively graded parameters to determine whether iRORA or cRORA was present in each case and showed that the level of interreader agreement for a 2-category classification of cRORA presence versus absence was substantial ($AC_1 = 0.68$), whereas

agreement was moderate ($AC_1 = 0.58$) for a classification of the presence of at least iRORA versus its absence. It is conceivable that the interreader agreement was better for cRORA because it represents a more advanced and well-defined lesion, thus being easier to assess. The moderate level of agreement ($AC_1 = 0.53$) for the 3-category classification of atrophic features based on no atrophy versus iRORA versus cRORA is reasonable, given that κ coefficients for interobserver agreement in the assessment of other ophthalmic and nonophthalmic imaging-based features have been reported to range from 0.13 to 0.96.^{30–33} In the field of retina, for example, fair agreement for interobserver variations (main pairwise $\kappa = 0.37–0.40$) have been reported for classifying types of choroidal neovascularization.³⁰

This study provided an opportunity to examine whether different combinations of features associated with the early structural OCT atrophic changes may lead to improved interreader agreement of atrophic AMD severity. The exploratory analyses demonstrated that significantly higher levels of agreement for an arbitrary, less severe stage of atrophy could be achieved by including only the presence of subsidence of the OPL and INL, or a hyporeflective wedge-shaped band, with the inclusion of a zone of choroidal hypertransmission as a criterion for defining a more severe stage of atrophy. These alternative definitions showed better agreement when excluding the requirement for RPE disruption or attenuation, which were the features that showed the lowest level of interreader agreement in this study, despite attempts to maximize agreement through careful training and instructions. However, improved interreader agreement could not be achieved when using similar definitions for a less and more severe stage of atrophy and basing the definition of photoreceptor degeneration on EZ or ELM disruption, despite the fact that these individual features showed the highest level of agreement. This is likely because of the very high prevalence of these features (91% of readings had either EZ or ELM disruption), meaning that the resulting classification of atrophy would be relatively similar whether EZ or ELM disruption was required to be present. Notably, it was the presence of (1) subsidence of the OPL and INL; (2) a hyporeflective wedge-shaped band; or (3) both that were the only features required for the original description of nascent GA.¹⁷ These features were recently shown to confer a 78-fold increased risk of progression to color fundus photography-defined GA in a prospective longitudinal study, underscoring its predictive validity.¹⁸

The findings of this study will help the field move toward a robust granular staging system for AMD by demonstrating that it is possible to discern different severities of atrophy on OCT imaging with robust levels of interreader agreement. These findings will also help to clarify the features that may best serve to establish earlier end points and inclusion or exclusion criteria for earlier-stage interventional trials in AMD. The choice of features to include in defining atrophic biomarkers and end points will vary depending on the aims of the studies being undertaken.

A strength of this study is its design, which sought to be as representative of a real-world reading center scenario as possible. This was achieved by providing the readers with

the full OCT volume scan, in addition to the OCT B-scan with an outlined region of interest. Thus, readers had the opportunity to use information from the neighboring B-scans for assessments of the region of interest. It is possible that the availability of other imaging methods, such as color fundus photography or fundus autofluorescence imaging, could improve the level of interreader agreement, but this warrants further investigation. Note that although the assessments in this study were limited to regions of interest, they can be extrapolated to the assessment of the presence of atrophy in the entire OCT volume scan of an eye in a clinical trial scenario.

Although we sought in this study to select a set of AMD cases that comprised the entire spectrum of changes seen in regions where atrophy develops (that ranged from no atrophic changes to an established area of atrophy), it is important to note that the prevalence of individual features associated with atrophy varied in the cases. Such variability likely does not represent selection bias but rather largely reflects the variability of the disease manifestation of atrophy itself. For example, the high prevalence of outer nuclear layer thinning and EZ and ELM disruption seen in 86% or more of readings indicates that these features commonly occur in eyes both with and without a full suite of atrophic changes. In contrast, a hyporeflective wedge-shaped band was present in only approximately one quarter of the readings.

The limitations of this study include the number of cases assessed in the grading task. This study also only assessed interreader agreement with just 1 of the widely used OCT instruments, and the generalizability of these findings to OCT scans obtained with other devices remains to be determined. It is also possible that interreader agreement for the different features assessed in this study may vary based

on the image quality, but this factor could not be assessed because we excluded all poor-quality images for this exercise. Finally, the level of interreader agreement may also have differed if the cases selected spanned a different range of disease severity. We chose to include an approximately equal proportion of cases with cRORA, iRORA, and no atrophy to provide a sufficiently large number of cases to assess the features associated with them. This limitation should be considered when attempting to generalize the findings of this study to the assessment of these features to other scenarios, such as in the assessment of these features more generally in a cohort with intermediate AMD.

In conclusion, this study demonstrated that a moderate to substantial level of interreader agreement could be achieved for iRORA and cRORA. It revealed a substantial or better agreement for the assessment of individual features associated with atrophy, except for RPE attenuation or disruption. We demonstrated in an exploratory analysis that the agreement for early atrophic stages could be improved significantly when the presence of INL and OPL subsidence, a hyporeflective wedge-shaped band, or both were determined for an earlier atrophic stage and when the presence of the zone of choroidal hypertransmission ($\geq 250 \mu\text{m}$) was included to define a more advanced atrophic stage. This study also highlighted the potential value of quantitative measurements of choroidal hypertransmission as a measure of disease progression, a concept that warrants further investigation in prospective longitudinal studies. The evaluation of interreader agreement in a reading center setting represents an important step toward a more detailed classification system for atrophic AMD and helps to inform the design and analysis of early intervention clinical trials in AMD.

Footnotes and Disclosures

Originally received: February 17, 2021.

Final revision: March 10, 2021.

Accepted: March 12, 2021.

Available online: March 23, 2021. Manuscript no. ORET-2021-84.

¹ Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia.

² Ophthalmology, Department of Surgery, The University of Melbourne, Melbourne, Australia.

³ GRADE Reading Center and Department of Ophthalmology, University of Bonn, Bonn, Germany.

⁴ Ophthalmic Genetics and Visual Function Branch, National Eye Institute, National Institutes of Health, Bethesda, Maryland.

⁵ Fundus Photograph Reading Center, Department of Ophthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin.

⁶ DUKE Reading Center and Department of Ophthalmology, Duke University, Durham, North Carolina.

⁷ Cologne Image Reading Center and Laboratory and Department of Ophthalmology, University of Cologne, Faculty of Medicine and University Hospital Cologne, Cologne, Germany.

⁸ Doheny Imaging Reading Center and Doheny Eye Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California.

⁹ Department of Biomedical and Clinical Science “Luigi Sacco,” Luigi Sacco Hospital University of Milan, Milan, Italy.

¹⁰ Utah Retinal Reading Center, John A. Moran Eye Center, University of Utah, Salt Lake City, Utah.

*Both authors contributed equally as first authors.

#Both authors contributed equally as senior authors.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s): M.P.: Nonfinancial support - Heidelberg Engineering, Carl Zeiss Meditec, CenterVue

F.G.H.: Financial support - Heidelberg Engineering, CenterVue, Zeiss, Optos, Bayer, Novartis, Acucela, Apellis, Kanghong, Roche/Genentech, Pixium Vision, Allergan, Lin Bioscience, Oxurion, NightStar X, Bioequ/Formycon, Stealth Therapeutics, Kodiak, Geuder, Grayburg Vision

G.J.J.: Consultant - Novartis, Regeneron, Iveric, Allegro

S.L.: Financial support - Novartis, Bayer, Allergan, Zeiss, Heidelberg Engineering

S.R.S.: Financial support - Amgen, Allergan, Roche/Genentech, Novartis, Bayer, Regeneron, Astellas, Oxurion, 4DMT, Heidelberg, Optos, Carl Zeiss Meditec, Nidek; Nonfinancial support: Heidelberg, Optos, Carl Zeiss Meditec, CenterVue, Nidek, Topcon

M.S.: Nonfinancial support - Carl Zeiss Meditec, Heidelberg Engineering

S.T.: Financial support - Allergan, Novartis, Bayer, Heidelberg Engineering; Nonfinancial support - CenterVue, Heidelberg Engineering, Optos, Carl Zeiss Meditec

R.H.G.: Consultant - Roche Genentech; Advisory board - Novartis, Bayer, Apellis

S.S.-V.: Financial support - Acucela/Kubota Vision, Allergan, Bayer, Gailmedix, Heidelberg Engineering, Katario, Novartis, Roche, Oxurion, Bioeq/Formycon, Apellis, SparingVision, Pixium; Nonfinancial support - Carl Zeiss Meditec, Centervue, Heidelberg Engineering

SriniVas R. Sadda, an editorial board member of this journal, was recused from the peer-review process of this article and had no access to information regarding its peer-review.

Supported in part by Research to Prevent Blindness, Inc., New York, New York (unrestricted grant to the Department of Ophthalmology & Visual Sciences, University of Utah [T.B., A.P.O., S.S.V.]); the Federal Ministry of Education and Research of Germany (scholarship FKZ 13N10349 [S.T., S.S.V.]), Berlin, Germany; the National Health & Medical Research Council of Australia (fellowship grant no.: GNT1103013 [R.H.G.]), Canberra, Australia; and the German Research Foundation (grant no.: PF950/1-1 [M.P.]), Bonn, Germany.

HUMAN SUBJECTS: Human subjects were included in this study. The human ethics committees at Royal Victorian Eye and Ear Hospital and University of Bonn approved the study. All research adhered to the tenets of the Declaration of Helsinki. All participants provided informed consent.

No animal subjects were included in this study.

Author Contributions:

Conception and design: Wu, Pfau, Blodi, Holz, Jaffe, Liakopoulos, Sadda, Staurengi, Chang, Thiele, Guymer, Schmitz-Valckenberg

Analysis and interpretation: Wu, Pfau, Blodi, Holz, Jaffe, Liakopoulos, Sadda, Staurengi, Guymer, Schmitz-Valckenberg

Data collection: Wu, Pfau, Blodi, Holz, Jaffe, Liakopoulos, Sadda, Staurengi, Bjelopera, Brown, Chang, Choong, Corradetti, Corvi, Domalpally, Hurtenbach, Gupta Nittala, Olson, Pak, Pappé, Saßmannshausen, Skalak, Thiele, Guymer, Schmitz-Valckenberg

Obtained funding: N/A

Overall responsibility: Wu, Pfau, Blodi, Holz, Jaffe, Liakopoulos, Sadda, Staurengi, Bjelopera, Brown, Chang, Choong, Corradetti, Corvi, Domalpally, Hurtenbach, Gupta Nittala, Olson, Pak, Pappé, Saßmannshausen, Skalak, Thiele, Guymer, Schmitz-Valckenberg

Abbreviations and Acronyms:

AC₁ = Gwet's first-order agreement coefficient; **AMD** = age-related macular degeneration; **CAM** = Classification of Atrophy Meetings; **CI** = confidence interval; **cRORA** = complete retinal pigment epithelium and outer retinal atrophy; **ELM** = external limiting membrane; **EZ** = ellipsoid zone; **INL** = inner nuclear layer; **iRORA** = incomplete retinal pigment epithelium and outer retinal atrophy; **OPL** = outer plexiform layer; **RPE** = retinal pigment epithelium; **SRD** = smallest real difference.

Keywords:

Nascent geographic atrophy, Outer retinal atrophy, Reading center, Retinal pigment epithelium, Variability.

Correspondence:

Steffen Schmitz-Valckenberg, MD, Department of Ophthalmology & Visual Science, John A. Moran Eye Center, University of Utah, 65 North Mario Capecchi Drive, Salt Lake City, UT 84312. E-mail: steffen.valckenberg@utah.edu.

References

- Bird AC, Bressler NM, Bressler SB, et al. An international classification and grading system for age-related maculopathy and age-related macular degeneration. The International ARM Epidemiological Study Group. *Surv Ophthalmol*. 1995;39:367–374.
- Age-Related Eye Disease Study Research Group. The Age-Related Eye Disease Study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the Age-Related Eye Disease Study report number 6. *Am J Ophthalmol*. 2001;132:668–681.
- Holz FG, Strauss EC, Schmitz-Valckenberg S, van Lookeren Campagne M. Geographic atrophy: clinical features and potential therapeutic approaches. *Ophthalmology*. 2014;121:1079–1091.
- Cheng QE, Gao J, Kim BJ, Ying GS. Design characteristics of geographic atrophy treatment trials: systematic review of registered trials in ClinicalTrials.gov. *Ophthalmol Retina*. 2018;2:518–525.
- Rosenfeld PJ, Dugel PU, Holz FG, et al. Emixustat hydrochloride for geographic atrophy secondary to age-related macular degeneration: a randomized clinical trial. *Ophthalmology*. 2018;125:1556–1567.
- Jaffe GJ, Westby K, Csaky KG, et al. C5 inhibitor avacincaptad pegol for geographic atrophy due to age-related macular degeneration: a randomized pivotal phase 2/3 trial. *Ophthalmology*. 2021;128:576–586.
- Nebbio M, Lambiase A, Cerini A, et al. Therapeutic approaches with intravitreal injections in geographic atrophy secondary to age-related macular degeneration: current drugs and potential molecules. *Int J Mol Sci*. 2019;20:1693.
- Liao DS, Grossi FV, El Mehdi D, et al. Complement C3 inhibitor pegcetacoplan for geographic atrophy secondary to age-related macular degeneration: a randomized phase 2 trial. *Ophthalmology*. 2020;127:186–195.
- Veerappan M, El-Hage-Sleiman AM, Tai V, et al. Optical coherence tomography reflective drusen substructures predict progression to geographic atrophy in age-related macular degeneration. *Ophthalmology*. 2016;123:2554–2570.
- Thiele S, Nadal J, Pfau M, et al. Prognostic value of intermediate age-related macular degeneration phenotypes for geographic atrophy progression. *Br J Ophthalmol*. 2020;105:239–245.
- Thiele S, Pfau M, Larsen PP, et al. Multimodal imaging patterns for development of central atrophy secondary to age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2018;59:AMD1–AMD11.
- Tan ACS, Pilgrim MG, Feam S, et al. Calcified nodules in retinal drusen are associated with disease progression in age-related macular degeneration. *Sci Transl Med*. 2018;10:eaat4544.
- Christenbury JG, Folgar FA, O'Connell RV, et al. Progression of intermediate age-related macular degeneration with proliferation and inner retinal migration of hyperreflective foci. *Ophthalmology*. 2013;120:1038–1045.
- Ouyang Y, Heussen FM, Hariri A, et al. Optical coherence tomography-based observation of the natural history of drusenoid lesion in eyes with dry age-related macular degeneration. *Ophthalmology*. 2013;120:2656–2665.
- Ferrara D, Silver RE, Louzada RN, et al. Optical coherence tomography features preceding the onset of advanced age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2017;58:3519–3529.
- Jaffe GJ, Chakravarthy U, Freund KB, et al. Imaging features associated with progression to geographic atrophy in age-

- related macular degeneration: CAM report 5. *Ophthalmol Retina*. 2020. In press.
17. Wu Z, Luu CD, Ayton LN, et al. Optical coherence tomography-defined changes preceding the development of drusen-associated atrophy in age-related macular degeneration. *Ophthalmology*. 2014;121:2415–2422.
 18. Wu Z, Luu CD, Hodgson LAB, et al. Prospective longitudinal evaluation of nascent geographic atrophy in age-related macular degeneration. *Ophthalmol Retina*. 2020;4:568–575.
 19. Guymer RH, Rosenfeld PJ, Curcio CA, et al. Incomplete retinal pigment epithelial and outer retinal atrophy in age-related macular degeneration: Classification of Atrophy Meeting report 4. *Ophthalmology*. 2020;127:394–409.
 20. Sadda SR, Guymer R, Holz FG, et al. Consensus definition for atrophy associated with age-related macular degeneration on OCT: Classification of Atrophy report 3. *Ophthalmology*. 2018;125:537–548.
 21. Schmitz-Valckenberg S, Sadda S, Staurenghi G, et al. Geographic atrophy: semantic considerations and literature review. *Retina*. 2016;36:2250–2264.
 22. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol*. 2008;61:29–48.
 23. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol*. 1990;43:543–549.
 24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometris*. 1977;33:159–174.
 25. Bland M, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327:307–310.
 26. Schmitz-Valckenberg S, Fleckenstein M, Göbel AP, et al. Optical coherence tomography and autofluorescence findings in areas with geographic atrophy due to age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2011;52:1–6.
 27. Sayegh RG, Simader C, Scheschy U, et al. A systematic comparison of spectral-domain optical coherence tomography and fundus autofluorescence in patients with geographic atrophy. *Ophthalmology*. 2011;118:1844–51.
 28. Hu Z, Medioni GG, Hernandez M, et al. Segmentation of the geographic atrophy in spectral-domain optical coherence tomography and fundus autofluorescence images. *Invest Ophthalmol Vis Sci*. 2013;54:8375–8383.
 29. U.S. Food and Drug Administration. FDA Cellular, Tissue, and Gene Therapies Advisory Committee CTGTAC Meeting #52. Cellular and Gene Therapies for Retinal Disorders. June 29, 2011. Available at: <https://wayback.archive-it.org/7993/20170113010811/http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/BloodVaccinesandOtherBiologics/CellularTissueandGeneTherapiesAdvisoryCommittee/UCM259087.pdf>.
 30. Holz FG, Jorzik J, Schutt F, et al. Agreement among ophthalmologists in evaluating fluorescein angiograms in patients with neovascular age-related macular degeneration for photodynamic therapy eligibility (FLAP-study). *Ophthalmology*. 2003;110:400–405.
 31. Kim SH, Lee EH, Jun JK, et al. Interpretive performance and inter-observer agreement on digital mammography test sets. *Korean J Radiol*. 2019;20:218–224.
 32. Siddiqui MR, Gormly KL, Bhoday J, et al. Interobserver agreement of radiologists assessing the response of rectal cancers to preoperative chemoradiation using the MRI tumour regression grading (mrTRG). *Clin Radiol*. 2016;71:854–862.
 33. Scott IU, Blodi BA, Ip MS, et al. SCORE Study report 2: interobserver agreement between investigator and reading center classification of retinal vein occlusion type. *Ophthalmology*. 2009;116:756–761.