

# Intersession Repeatability of Structural Biomarkers in Early and Intermediate Age-Related Macular Degeneration: A MACUSTAR Study Report

Marlene Saßmannshausen<sup>1,2</sup>, Sarah Thiele<sup>1,2</sup>, Charlotte Behning<sup>3</sup>, Maximilian Pfau<sup>1,2,4</sup>, Matthias Schmid<sup>3</sup>, Sérgio Leal<sup>5</sup>, Ulrich F. O. Luhmann<sup>6</sup>, Robert P. Finger<sup>1</sup>, Frank G. Holz<sup>1,2</sup>, and Steffen Schmitz-Valckenberg<sup>1,2,7</sup>, on behalf of the MACUSTAR Consortium

<sup>1</sup> Department of Ophthalmology, University of Bonn, Bonn, Germany

<sup>2</sup> GRADE Reading Center, University of Bonn, Bonn, Germany

<sup>3</sup> Institute of Medical Biometry, Informatics and Epidemiology, Medical Faculty, University of Bonn, Bonn, Germany

<sup>4</sup> Ophthalmic Genetics and Visual Function Branch, National Eye Institute, Bethesda, MD, USA

<sup>5</sup> Bayer AG, Berlin, Germany

<sup>6</sup> Roche Pharmaceutical Research and Early Development, Translational Medicine Ophthalmology, Roche Innovation Center Basel, Basel, Switzerland

<sup>7</sup> John A. Moran Eye Center, Department of Ophthalmology & Visual Sciences, University of Utah, Salt Lake City, UT, USA

**Correspondence:** Steffen Schmitz-Valckenberg, Department of Ophthalmology & Visual Sciences, John A. Moran Eye Center, University of Utah, 65 Mario Capecchi Drive, Salt Lake City, UT 84132, USA.  
e-mail: [steffen.valckenberg@utah.edu](mailto:steffen.valckenberg@utah.edu)

**Received:** November 19, 2021

**Accepted:** February 27, 2022

**Published:** March 25, 2022

**Keywords:** test–retest; drusen; pigment epithelium detachment; refractile deposits; grading

**Citation:** Saßmannshausen M, Thiele S, Behning C, Pfau M, Schmid M, Leal S, Luhmann UFO, Finger RP, Holz FG, Schmitz-Valckenberg S. Intersession repeatability of structural biomarkers in early and intermediate age-related macular degeneration: A MACUSTAR study report. *Transl Vis Sci Technol.* 2022;11(3):27. <https://doi.org/10.1167/tvst.11.3.27>

**Purpose:** To analyze the intersession repeatability of structural biomarkers in eyes with early and intermediate age-related macular degeneration (iAMD) within the cross-sectional part of the observational multicenter MACUSTAR study.

**Methods:** Certified site personnel obtained multimodal imaging data at two visits ( $38 \pm 20$  [mean  $\pm$  standard deviation] days apart), including spectral-domain optical coherence tomography (SD-OCT). One junior reader performed systematic and blinded grading at the central reading center, followed by senior reader review. Structural biomarkers included maximum drusen size classification ( $>63$  to  $\leq 125$   $\mu\text{m}$  vs.  $>125$   $\mu\text{m}$ ), presence of large pigment epithelium detachments (PEDs), reticular pseudodrusen (RPD), vitelliform lesions, and refractile deposits. Intrasession variability was assessed using Cohen's  $\kappa$  statistics.

**Results:** At the first visit, 202 study eyes of 202 participants were graded as manifesting with either early ( $n = 34$ ) or intermediate ( $n = 168$ ) AMD. Grading of imaging data between visits revealed perfect agreement for the maximum drusen size classification ( $\kappa = 0.817$ ; 95% confidence interval, 0.70–0.94). In iAMD eyes, perfect to substantial agreement was determined for the presence of large PEDs (0.87; 0.69–1.00) and RPD (0.752; 0.63–0.87), while intersession agreement was lower for the presence of vitelliform lesions (0.649; 0.39–0.65) and refractile deposits (0.342;  $-0.029$ –0.713), respectively.

**Conclusions:** Multimodal retinal imaging analysis between sessions showed a higher repeatability for structural biomarkers with predefined cutoff values than purely qualitative defined parameters.

**Translational Relevance:** A high repeatability of retinal imaging biomarkers will be important to implement automatic grading approaches and to establish robust and meaningful structural clinical endpoints for future interventional clinical trials in patients with iAMD.

## Introduction

Age-related macular degeneration (AMD) remains one of the most important causes of irreversible vision loss in industrialized countries.<sup>1,2</sup> Driven by the ongoing aging of the population, the prevalence is expected to increase further.<sup>3,4</sup> There is a growing unmet need to develop innovative interventional therapies in early and intermediate AMD (iAMD) that can delay progression to late-stage disease manifestation and associated severe vision loss. To assess potential therapeutic approaches, robust and reliable biomarkers on retinal structure and function are required, allowing for high repeatability between assessments as an essential prerequisite prior to acceptance as clinical endpoints by regulatory authorities, health care providers, and payers.

Previous studies report a great variability of optical coherence tomography (OCT) imaging based on retinal gradings, often depending on the underlying disease and difference in comparison to the variability of grading retinal markers in healthy participants.<sup>5,6</sup> In the presence of AMD, good repeatability and correlation of OCT imaging-based gradings were shown for the average central 1-mm subfield thickness in early AMD as well as for the foveal central subfield, foveal central point thickness, and presence of intra- or subretinal fluid in neovascular disease stages.<sup>7-9</sup> Good overall agreement on retinal thickness analysis in eyes with neovascular AMD has been demonstrated when comparing different imaging platforms.<sup>10</sup> Furthermore, high reproducibility of gradings on central drusen volume assessment within the 3-mm circle of the fovea has been reported in eyes with early AMD as well as iAMD stage among readers and in comparison to an automated software tool.<sup>11</sup>

However, more detailed data reports from multicenter prospective studies are needed to assess the reproducibility of human expert gradings on already established biomarkers by multimodal imaging on retinal microstructure in iAMD. A reliable reproducibility of retinal imaging biomarkers will further provide the basis for the implementation of automatic grading approaches, which will be required in the near future as multimodal retinal imaging approaches in clinical trials also lead to a growing number of available data sets.

Currently already established retinal imaging biomarkers include the presence of precursor lesions representing high-risk factors for disease progression, the presence of large pigment-epithelium detachment (PED), or reticular pseudodrusen (RPD), which can be readily detected by multimodal retinal imaging.<sup>12-16</sup>

Taking this into account, MACUSTAR is an ongoing European, low-interventional, multicenter study that aims to develop and validate clinical tests to detect subtle longitudinal disease progression in iAMD that can be used as structural and functional endpoints in upcoming interventional clinical trials.<sup>17,18</sup> As part of the study design, standardized multimodal retinal imaging is performed by certified technicians and analyzed according to standardized grading strategies in a reading center setting. The protocol also included a validation visit with additional retinal imaging acquired shortly after the screening and baseline visits. In the current analysis, based on the cross-sectional data set of the European MACUSTAR study, we assess the intersession repeatability of detecting qualitative high-risk structural lesions for AMD progression.

## Methods

### The MACUSTAR Study

MACUSTAR is a multicenter, low-interventional natural history study in patients with AMD that is conducted at 20 sites across seven European countries (ClinicalTrials.gov identifier: NCT03349801). Participants were enrolled from March 2018 to February 2020, and the last visit of the last patient is expected for February 2023. The details of the study design, including inclusion and exclusion criteria, have been published elsewhere.<sup>17,18</sup> In the cross-sectional study part, four groups of participants with different disease stages of early, intermediate, late, and no AMD were included. This study has been conducted according to the provisions of the Declaration of Helsinki, and all participants provided informed consent.

The cross-sectional part of MACUSTAR was designed for the technical evaluation of functional and structural outcome measures. In this subcohort of the study population, one additional validation visit (V3, day  $14 \pm 7$  days) was scheduled shortly after screening (V1, day  $-28$  to 0) and baseline (V2, day 0). For the current analysis, we included all participants in the cross-sectional part classified as either early AMD or iAMD at V1. Based on Ferris et al.,<sup>19</sup> early AMD was defined as presenting with medium-sized drusen ( $>63 \mu\text{m}$  and  $\leq 125 \mu\text{m}$ ) in the absence of any AMD pigmentary alterations and any signs of intermediate- or late-stage AMD manifestations in both eyes. For iAMD, both eyes had to exhibit large drusen ( $> 125 \mu\text{m}$ ) and/or any AMD pigmentary abnormalities. In addition to these definitions, any extrafoveal geographic atrophy (GA) lesion not larger than  $1.25 \text{ mm}^2$  could be present in the fellow eye.

## Retinal Imaging Protocol

Following pupil dilatation with tropicamide 0.5% and phenylephrine 2.5%, participants underwent multimodal imaging according to standard operational procedures by certified study site personnel. Retinal imaging included combined confocal scanning laser ophthalmoscopy (cSLO) (near-infrared reflectance [IR], multicolor, green and/or blue fundus autofluorescence imaging [FAF, automated real time mode (ART) at least 30 single frames]) and spectral-domain optical coherence tomography (SD-OCT; Heidelberg Engineering, Heidelberg, Germany) ( $20^\circ \times 20^\circ$ , 25 B-scans, distance 240  $\mu\text{m}$ , ART mode, four frames;  $30^\circ \times 25^\circ$  enhanced-depth imaging mode, 241 B-scans, distance 30  $\mu\text{m}$ , ART mode, nine frames) as well as color fundus photography (CFP). Furthermore, the average corneal curvature for each eye was obtained at V1 to enhance the precision of absolute measurements on combined cSLO and SD-OCT data for subsequent image analyses. All imaging data were transmitted to the GRADE Reading Center (University of Bonn, Germany) through a secure, web-based portal. After transmission, the imaging data were reviewed for completeness, technical quality, and adherence to the predefined imaging protocol. If deemed acceptable by data management, imaging data were assigned to medical readers for grading.

## Grading Strategy

Imaging data of each visit were systematically graded by one junior reader, followed by one senior reader review according to standardized and predefined grading procedures. All junior readers underwent structured training procedures for assessment of retinal imaging in eyes with AMD and had already gained experience in at least one clinical trial at a reading center. Senior readers were required to be certified board members in ophthalmology and/or were required to have extensive scientific expertise in the field of multimodal imaging and clinical AMD trials.

A total of nine junior and two senior readers were involved in the grading of imaging data. The mean number of gradings performed by each junior reader was 34 (range, 3–85) at V2 and 73 (range, 22–119; involvement of four junior readers) at V3, while a mean of 301 gradings at V2 (involvement of one senior reader) and 146 (range, 104–187) gradings at V3 were performed by the senior readers.

Before evaluation of any images obtained in the context of this study, readers underwent an initial training run on 15 representative AMD cases. At the screening visit, grading results of the reading center

had to confirm morphologic inclusion and exclusion criteria as defined in the study protocol prior to final enrollment into MACUSTAR.<sup>18</sup> To ensure independent grading of V1 versus V3 imaging data, data management relabeled the transmitted V3 imaging data (masking the original study identifiers) and assigned these data in a dedicated imaging review library. Without references images or grading results from other visits, V3 data were processed and documented by readers in the grading database.

## Structural Grading Parameters

Readers were instructed to use the dense SD-OCT raster scan as the base imaging modality for grading, while other modalities provided complementary and confirmatory information. To assess maximum drusen size, the dense SD-OCT raster scan was reviewed for the presence of dome-shaped elevations of the retinal pigment epithelium (RPE) band, separating the latter from Bruch's membrane. According to subjective judgment, readers were instructed to select the B-scan with the largest appearing druse. After changing to the 1:1  $\mu\text{m}$  display and increasing the magnification to 400%, the basal lateral diameter of the selected druse at the level of Bruch's membrane was determined using the caliper tool in the Heidelberg Eye Explorer (Heidelberg, Germany). Conversion from pixels into microns was based on the formula provided by the Heidelberg Eye Explorer software, taking the focus settings during acquisition and the individual corneal curvature measurements of each eye into account. This measurement was used to confirm and document the maximum drusen size, which allowed for classification into medium-sized ( $>63 \mu\text{m}$  and  $\leq 125 \mu\text{m}$ ) and large ( $>125 \mu\text{m}$ ) drusen. To meet the requirements for the presence of PEDs, the basal diameter of the RPE elevation had to be at least 1000  $\mu\text{m}$ . In addition, the minimum height at the highest point of the elevation had to be 200  $\mu\text{m}$ , as measured from the inner edge of Bruch's membrane to the outer edge of the RPE band. RPD were defined as hyperreflective irregularities and elevations above the RPE/Bruch's membrane SD-OCT band, showing medium-reflective mounds or cones at the level of the ellipsoid zone or between the ellipsoid zone and the RPE surface. RPD had to correspond to an ill-defined network of oval or roundish irregularities with a variable diameter of approximately 100  $\mu\text{m}$  on either cSLO FAF or IR imaging. RPD were graded to be present if a minimum of five individual lesions were visible. Vitelliform lesion were defined by accumulation of hyperreflective, amorphous material in the subretinal space as seen by SD-OCT imaging, typically associated with an increased signal in FAF imaging.

The presence of refractile deposits was determined either by an intense laminar hyperreflectivity ( $\geq 100 \mu\text{m}$ ) at the level of Bruch’s membrane or a pyramidal structure at the level of the outer retina (“ghost drusen”). Supportive evidence included glistening and a yellow-shining appearance by CFP, a corresponding hyperreflective signal by IR, and/or a mildly increased or mottled FAF signal. For the gradings, retinal parameters were graded as “yes” if the grader was more than 90% sure that a finding was positive. If grading categories were graded as “no”, the reader was more than 50% sure that the finding is negative.

### Statistical Analysis

All statistical analyses were performed using the R software environment (version 4.0.2; R Foundation, Vienna, Austria) with the psych package (version 2.1.6, Revelle W, 2021). Cohen’s  $\kappa$  statistics were used to assess the intersession repeatability of qualitative structural AMD biomarkers, including calculation of 95% confidence intervals (CIs). Intersession agreement was validated as fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), or almost perfect (0.81–1.00) according to the proposed categorization by Landis and Koch.<sup>20</sup>

## Results

### Study Cohort

In total, 301 study eyes from 301 participants were enrolled in the cross-sectional part of MACUSTAR. Their demographic data are provided in Table 1. Based on the final grading results at V1, the cohort included 202 study eyes with either early AMD ( $n = 34$ ; mean  $\pm$

SD age =  $71.7 \pm 6.38$  years; 27 = female, 7 = male) or iAMD ( $n = 168$ ; mean  $\pm$  SD age =  $71.2 \pm 7.55$  years; 106 = female, 62 = male). Out of these 202 study eyes, 7 ( $n = 6$  early,  $n = 1$  iAMD) were lost for the follow-up visit (V3). These were therefore not included in the repeatability analyses. The mean time of imaging data acquisition between V1 and V3 was  $38 \pm 20$  (range, 7–195) days. The grading results for both visits of the remaining 195 study eyes—as independently determined for at V1 and V3 time points—were subsequently used to analyze the intersession repeatability for maximum drusen size classification. For retest analysis of the other structural biomarkers, results are reported for eyes graded as iAMD at V1, as these structural parameters are mainly detected in eyes with iAMD.

Intersession agreement of maximum drusen size for distinction between early ( $>63 \mu\text{m}$  and  $\leq 125 \mu\text{m}$ ) and intermediate ( $>125 \mu\text{m}$ ) showed almost perfect agreement ( $\kappa = 0.817$ ; 95% CI, 0.70–0.94). In study eyes, the intersession repeatability to the validation visit (V3) for phenotypic hallmarks of iAMD varied between fair and almost perfect agreement, with highest consensus for the presence of large PEDs ( $\kappa = 0.869$ ; 95% CI, 0.69–1.0), followed by the presence of RPD ( $\kappa = 0.752$ ; 95% CI, 0.63–0.87) and vitelliform lesions ( $\kappa = 0.649$ ; 95% CI, 0.39–0.91). Lowest agreement was observed for the presence of refractile deposits ( $\kappa = 0.342$ ; 95% CI,  $-0.03$  to 0.71). In addition to study eye data, Table 2 also summarizes the results of fellow eyes, the latter showing overall similar results.

Figures 1 to 5 demonstrate examples of study eyes by multimodal retinal imaging (from left to right: CFP, blue light FAF, combined IR and SD-OCT imaging, enlarged window of B-scan) with agreements and disagreements on gradings between sessions for each of the risk features.

**Table 1.** Overview of Study Demographics Within the Early, Intermediate, Late, and No AMD Groups and Within the Overall Study Group at Screening Visit (V1)

Characteristic	Early AMD ( $n = 34$ )	iAMD ( $n = 168$ )	Late AMD ( $n = 43$ )	No AMD ( $n = 56$ )	Overall ( $n = 301$ )
Age, y					
Mean $\pm$ SD	71.7 $\pm$ 6.38	71.2 $\pm$ 7.55	74.9 $\pm$ 5.59	68.1 $\pm$ 6.35	71.2 $\pm$ 7.20
Median [minimum, maximum]	72.0 [57.0, 82.0]	72.0 [55.0, 88.0]	75.0 [64.0, 84.0]	68.0 [55.0, 80.0]	72.0 [55.0, 88.0]
Gender, $n$ (%)					
Female	27 (79.4)	106 (63.1)	21 (48.8)	33 (58.9)	187 (62.1)
Study eyes, $n$ (%)					
OD	13 (38.2)	78 (46.4)	23 (53.5)	33 (58.9)	147 (48.8)
BCVA, logMAR					
Mean $\pm$ SD	0.011 $\pm$ 0.083	0.024 $\pm$ 0.103	0.771 $\pm$ 0.247	$-0.041 \pm 0.084$	0.117 $\pm$ 0.297
Median [minimum, maximum]	0.02 [ $-0.18, 0.20$ ]	0.02 [ $-0.24, 0.68$ ]	0.84 [0.20, 1.24]	$-0.06$ [ $-0.24, 0.14$ ]	0.02 [ $-0.24, 1.24$ ]

BCVA, best-corrected visual acuity; logMAR, logarithm of the minimum angle of resolution; OD, oculus dexter.



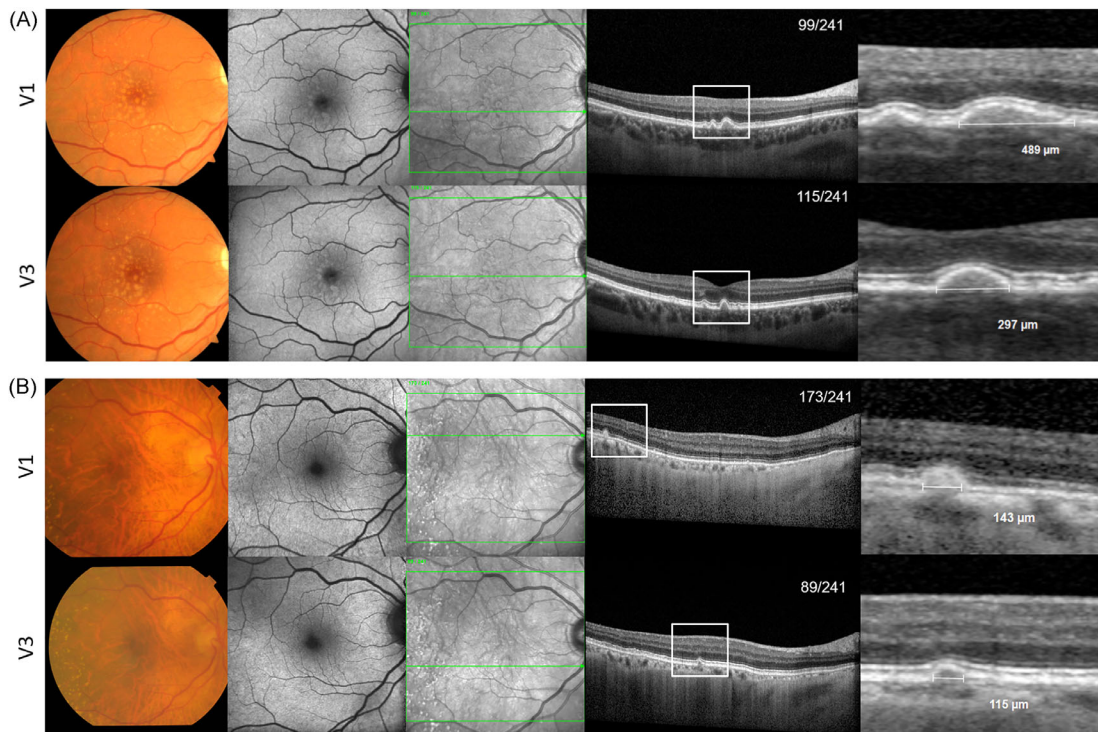
**Table 2.** Results of Cohen’s  $\kappa$  Analysis with the 95% Confidence Interval for Intersession Agreement on Qualitative Retinal Parameters between the Screening (V1) and the Retest Validation (V3) Visit

Retinal Parameter	Inter-Session Agreement	
	Cohen’s $\kappa$ [95% CI]	
	Study Eyes ( $n = 195$ )	Fellow Eyes ( $n = 187$ ) <sup>a</sup>
Maximum drusen size classification ( $>63\text{--}\leq 125 / > 125 \mu\text{m}$ )	0.817 [0.695–0.939] ( $n = 167$ )	0.782 [0.645–0.918] ( $n = 167$ )
Large PEDs	0.869 [0.689–1.00]	0.766 [0.586–0.946]
RPD	0.752 [0.633–0.870]	0.783 [0.666–0.900]
Vitelliform lesions	0.649 [0.389–0.649]	0.568 [0.283–0.853]
Refractile deposits	0.342 [–0.029–0.713]	0.512 [0.196–0.827]

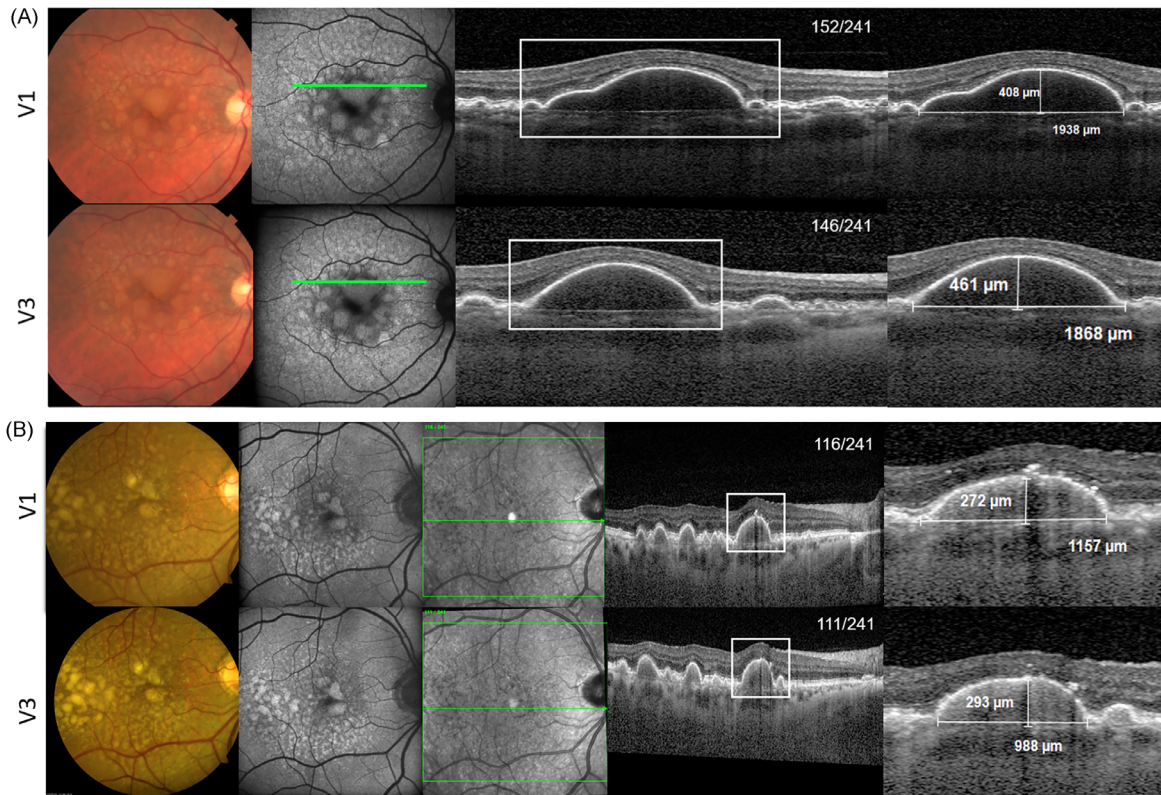
Gradings on maximum drusen size classification were performed in eyes with early and iAMD, while gradings on the presence of large PEDs, RPD, vitelliform lesions, or refractile deposits are reported for eyes graded as iAMD ( $n = 167$ ). Results are displayed for all study and fellow eyes with complete data sets available at V1 and V3. CI, confidence interval.

<sup>a</sup>Complete data were available for 187 fellow eyes with medium- or large-sized drusen.

For maximum drusen size (Fig. 1) and PED assessments (Fig. 2), reasons for discrepancy between gradings were related to not only the lateral diameter measurement of a specific lesion or by selection of different drusen with similar-appearing sizes throughout the scan field but also the reader selection of the representative B-scan within the dense raster scan that showed the identical lesion at both sessions (Fig. 2B).



**Figure 1.** (A) Agreement of gradings on maximum drusen size classification between sessions as maximum drusen size clearly exceeds  $>125 \mu\text{m}$  at V1 and V3. Grading results on maximum drusen size are shown in the enlarged window of the B-scan (righthand side). (B) Disagreement of gradings on maximum drusen size classification. At V1, an eccentric lesion in an area with concomitant presence of cuticular drusen was selected as the largest appearing druse, while at V3, a more central lesion was chosen. Both measurements were close to the cutoff value of 125 microns. Grading results on maximum drusen size are shown in the enlarged window of the B-scan (righthand side).



**Figure 2.** (A) Agreement of gradings on large PED presence between readers as measurement requirements for large PED presence were met at both sessions in SD-OCT imaging (basal diameter  $\geq 1000 \mu\text{m}$ , minimum height of  $200 \mu\text{m}$ ). Grading values of the horizontal and vertical extent of the lesion are presented on the *righthand side*. (B) Disagreement of gradings on large PED presence with the measurements on the basal diameter of RPE elevation close to the cutoff value of  $1000 \mu\text{m}$ . The disagreement appeared to be mainly related to the selection of a different B-scan (116/241 vs. 111/241).

By using the dense raster scan (distance of 30 microns between neighboring B-scans), only minor variation in the orientation and centration of the scan field occurred, which had no major impact on maximum drusen assessment. Overall, variations for determination of maximum drusen size or large PEDs only occurred if the lateral diameter was close to the predefined cutoff values of 125 microns and 1000 microns, respectively.

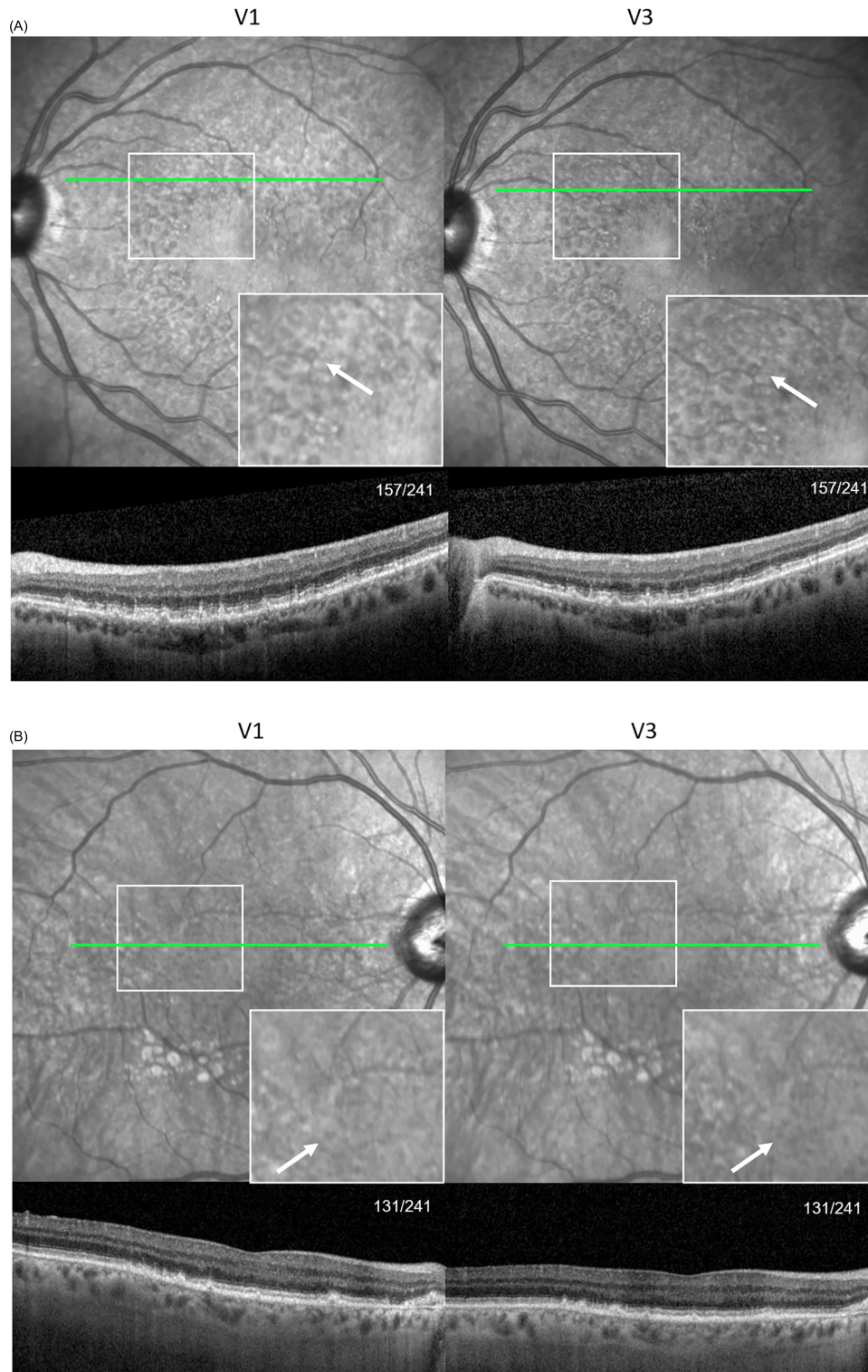
For gradings on the presence of RPD (Fig. 3), vitelliform lesions (Fig. 4), or refractile deposits (Fig. 5), reasons for disagreement were limitations of the current available multimodal retinal imaging data, especially in ambiguous cases with only minor changes. For refractile deposits, for example, the presence of RPE laminar intense hyperreflectivity was just detectable in one of the 241 B-scans in the dense raster scan field. At the same time, these lesions could not have been detected in the 25 B-scan OCT volume.

## Discussion

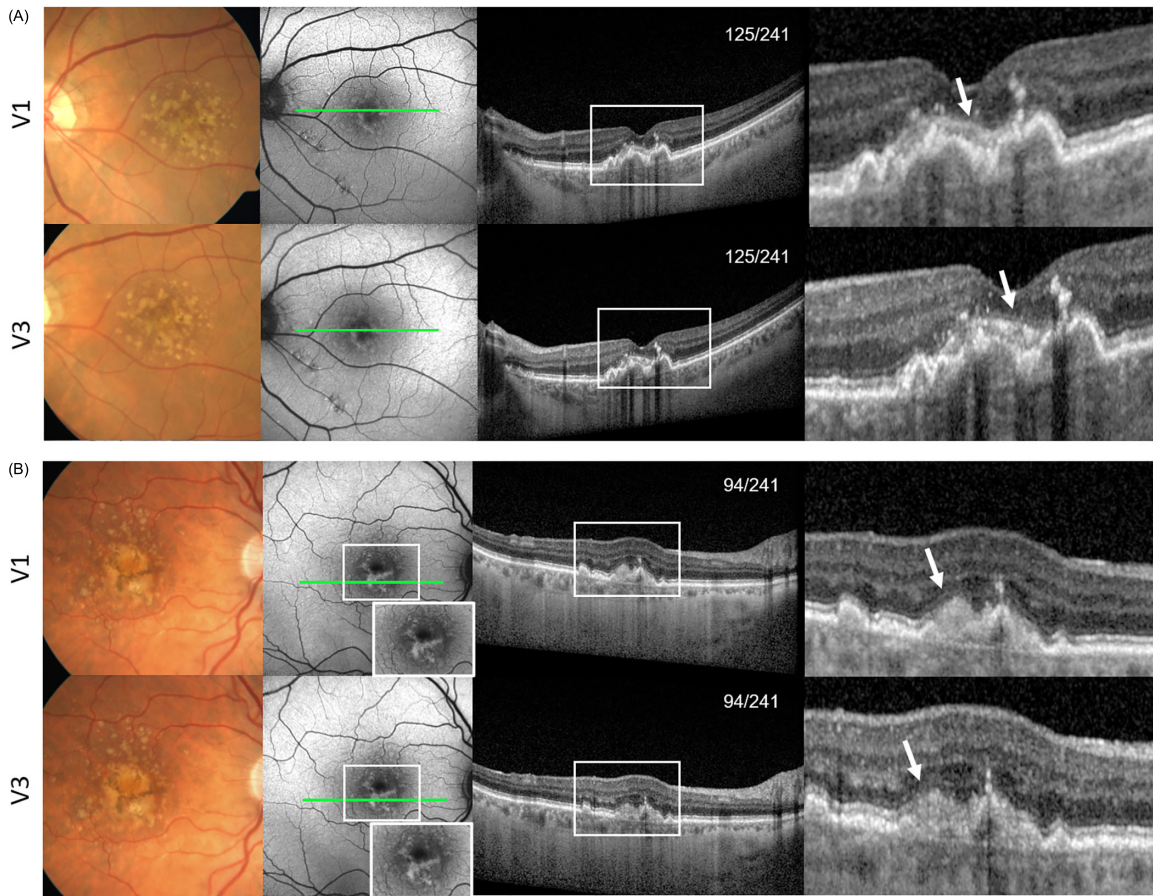
In this study, we systematically analyzed the intersession repeatability of gradings on previously established structural, imaging-based biomarkers in eyes with early AMD and iAMD within a central reading center setting in the context of the prospective, multi-center MACUSTAR study. Beyond the pure assessment of drusen size categories, we could demonstrate overall high reliability for the detection of PEDs and areas of RPD. Furthermore, this study identified reasons for disagreement and grading variabilities, which will improve further development of clinical endpoints and their assessment in future AMD studies.

For assessing parameters with clearly predefined cutoff values (e.g., drusen size classification or presence of large PEDs), we demonstrated almost perfect agreement between sessions with slightly higher  $\kappa$  values for gradings on large PED presence than drusen size





**Figure 3.** (A) Agreement on RPD presence as RPD are clearly detectable as oval or roundish irregularities in the infrared reflectance image (IR) (top row, white arrows in the enlarged window point at RPD) with a corresponding accumulation of hyperreflective material above the RPE level in the SD-OCT (bottom row). The green line indicates the position of the corresponding B-scan. (B) Disagreement on RPD presence (V1 graded as RPD absence, V3 as RPD presence) between sessions. The detection of the presence of single oval irregularities as the characteristic signs of RPD appears to be challenging; even at high magnification by the IR image (top row, white arrows in the enlarged window point at RPD). By OCT, typical presence of accumulated hyperreflective material above the RPE level is hardly to detect at both sessions. The green line indicates the position of the corresponding B-scan.



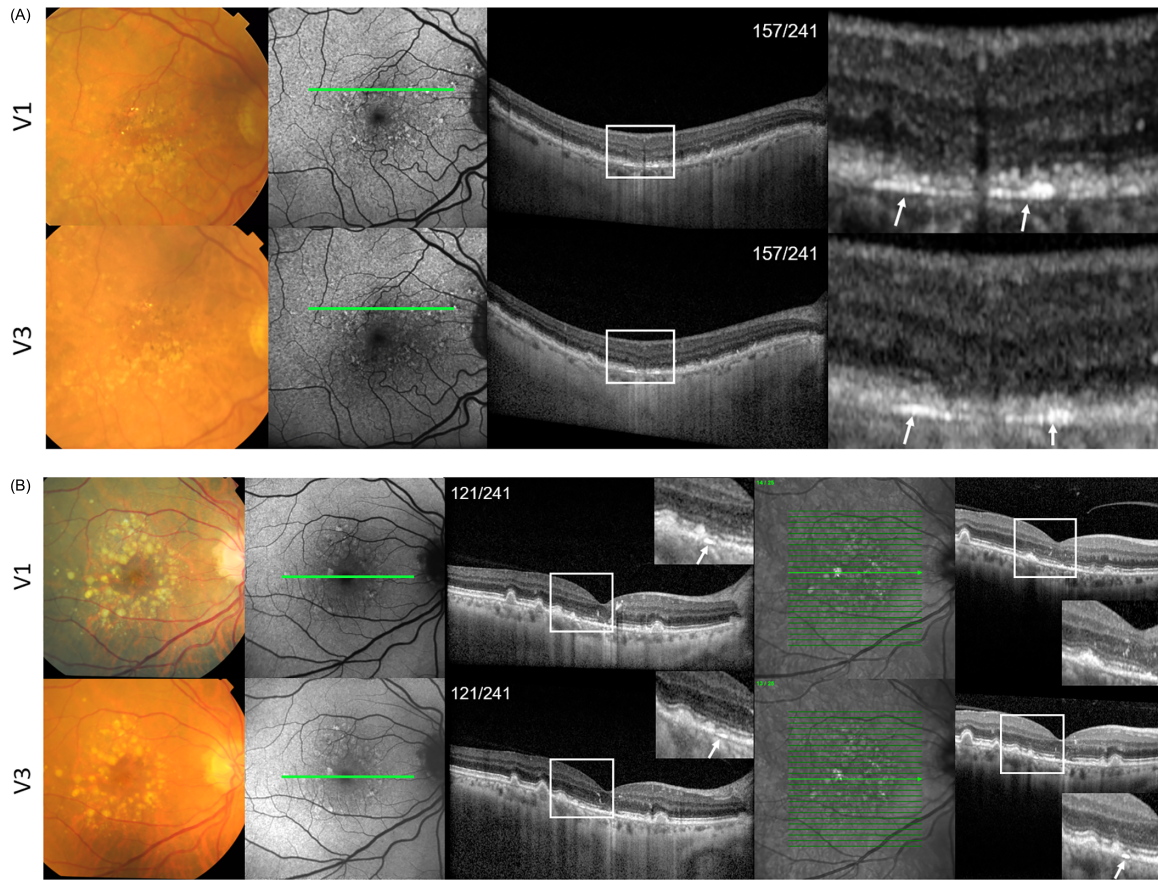
**Figure 4.** (A) Agreement on the presence of vitelliform lesions being well visible as an hyperreflective lesion in FAF imaging corresponding to an accumulation of hyperreflective material superior to the RPE in the SD-OCT. The *green line* indicates the position of the corresponding B-scan. (B) Disagreement on the presence of vitelliform lesions (V1 graded as nonpresent, V3 graded as present). At both visits, small amounts of hyperreflective, amorphous material in the subretinal space corresponding to an increased signal in FAF imaging are present. It appears that human reader assessment overlooked the presence of vitelliform material at V1. The *green line* indicates the position of the corresponding B-scan.

classification. The most important reason for disagreement related to drusen sizes close to the arbitrary cutoff value of 125  $\mu\text{m}$ . However, other factors need to be considered, such as density of the scan pattern and the selection of the largest-appearing druse within the scan field, the latter representing a subjective assessment by the reader and not an actual measurement. Among previous studies, Corvi et al.<sup>11</sup> showed comparable interession agreement ( $\kappa = 0.88$ ) on assessing central drusen volume within the central 3 mm of the Early Treatment Diabetic Retinopathy Study grid. Recently, Müller et al.<sup>21</sup> reported an intraclass correlation coefficient of 0.788 between readers for drusen detection with a minimum size of 1558.6  $\mu\text{m}^2$ , which corresponds to the minimal drusen diameter of 63  $\mu\text{m}$  and is well comparable to results in this study. Compared with the previous analysis, it is important to point out that we did not only perform an independent grading

of the imaging data acquired at a single time point in a single-center setting. Rather, interession analysis was based on imaging data obtained at two different sessions and acquired by multiple certified operators at clinical sites throughout different European countries. Furthermore, grading was performed at a central reading center, involving multiple readers who had detailed instructions and dedicated training. This setup allows the processing of the large amount of imaging data needed in multicenter interventional state-of-the-art clinical AMD trials. These differences in the design may be the reason interession repeatability for RPD presence (0.75) was lower than that found in previous reports. Cohen's  $\kappa$  values of 0.94 by FAF, 0.95 by multispectral imaging, and up to 0.96 in SD-OCT imaging have been obtained.<sup>22–24</sup>

The relatively low agreement for the presence of vitelliform lesions and refractile deposits may be due to





**Figure 5.** (A) Agreement on the presence of multiple lesions of refractile deposits being clearly detectable in multimodal retinal imaging. Multiple refractile deposits are detected as an accumulation of hyperreflective crystalline material in CFP and an increased signal in FAF imaging corresponding to the presence of laminar intense hyperreflectivity at the level of Bruch's membrane in the corresponding SD-OCT B-scan (highlighted by *white arrows* in the enlarged window of the B-scan). (B) Disagreement on the presence of refractile deposits (V1 graded as nonpresent, V3 graded as present). CFP imaging shows a single small hyperreflective lesion correlating to mildly increased signal in FAF imaging at both sessions. At the corresponding OCT B-scan, presence of intense laminar hyperreflectivity is shown in the dense 241 B-scans SD-OCT (highlighted by *white arrows* in the enlarged window of the B-scan). In contrast, this lesion is skipped in the 25 B-scans SD-OCT at V1.

various additional reasons. These represent less established anatomic parameters. Better definitions and more dedicated training may be helpful to improve interession agreement in the future. This assumption is supported by the observation of obvious human errors in grading and cases that appeared ambiguous and/or presented only as subtle lesions. Improvements in the review software, allowing to assess changes in multimodal imaging data set by highlighting corresponding regions in different modalities, may further reduce disagreement. Finally, another reason for relatively low  $\kappa$  values might be due to the low prevalence of these two features in the data set.

We advocate for the development of consensus definitions for high-risk features. Similarly, the Classification of Atrophy Meetings group recently reached a consensus for the definition of incomplete RPE and

outer retinal atrophy (iRORA) and complete RPE and outer retinal atrophy (cRORA).<sup>25,26</sup> This will also be necessary for the development and implementation of machine learning approaches and other artificial intelligence-based approaches for the detection of high-risk features of iAMD. In this context, previous studies reported on machine and deep learning (DL)-based methods in eyes with early and intermediate AMD to reliably detect structural biomarkers as lesions of hyperreflective foci (HRF) and RPD, to quantify drusen volumes, and to estimate the individual's eye risk toward disease progression.<sup>27–29</sup> In another study, Liefers et al.<sup>30</sup> even revealed findings on DL-based models exceeding the human performance on the detection of neovascular AMD-associated structural biomarkers. Interestingly, in the same data set, the accuracy rate for detecting drusen and drusenoid PEDs

was slightly higher for readers than for the model.<sup>30</sup> However, further studies validating machine learning approaches to differentiate disease phenotypes in multimodal retinal imaging as well as to detect early signs for disease progression as, for example, lesions of iRORA in eyes with iAMD are currently still ongoing. The availability of standardized gradings on various structural biomarkers as well as multimodal retinal imaging data within the MACUSTAR study can form the basis for the development of further automated grading approaches, which will also help to grade the increasing amount of multimodal retinal imaging data sets available in currently ongoing and future multicenter, observational clinical iAMD trials.

Several limitations of the current analysis must be considered. First, the retest analysis has been performed only in a limited number of eyes. Second, the retest analysis on retinal parameters was based on one OCT platform. It is not known if similar levels of agreement would be achieved when using other OCT devices. Third, drusen size measurements were performed on OCT scans and not based on CFP as it has been traditionally done in most previous studies. In this context, the recent report by Kim et al.<sup>31</sup> would suggest similar detection rates of drusen by both modalities while slightly larger measurements were seen by SD-OCT-based quantification. Furthermore, we did not provide levels of agreement of additional imaging-related biomarkers of AMD that are not included in the MACUSTAR study protocol but have been discussed in the past, such as intraretinal hyperreflective foci or volumetric data on photoreceptor layer integrity. Strengths of this study are the standardized retinal imaging acquisition at two independent visits with separate retinal imaging and the standardized grading of structural biomarkers of iAMD in a reading center setting. It is also important to mention that we enhanced the accuracy of absolute measurements by using the average corneal curvature measurements and taking the individual scaling factor for each eye into account. Moreover, results of this analysis derive from a multicenter- and multiexaminer-derived data set replicating the conditions of future potential trials and therefore provide a realistic expectation of grading results on intersession agreement.

In conclusion, intersession repeatability assessment of structural biomarkers in eyes with iAMD varies from almost perfect agreement for drusen size classification and presence of large PEDs to fair agreement for refractile deposits. Establishing reasons for disagreement will be helpful to refine further grading and image analysis strategies in current clinical settings, including those implementing artificial intelligence approaches.

Well-controlled and systematic analysis of imaging biomarkers represents an important prerequisite to further evaluate and validate potential clinical outcome measures for future interventional clinical trials in iAMD.

## Acknowledgments

Supported by the Innovative Medicines Initiative 2 Joint Undertaking under grant 116076. This joint undertaking receives support from the European Union's Horizon 2020 research and innovation program and European Federation of Pharmaceutical Industries and Associations (EFPIA). MS is supported by the Gerok Research Grant (BONFOR O-137.0030, Faculty of Medicine, University of Bonn, Bonn, Germany). The sponsors or funding organizations had no role in the design or conduct of the MACUSTAR study (project number: 116076) research. The communication reflects the authors' views. Neither IMI nor the European Union, EFPIA, or any associated partners are responsible for any use that may be made of the information contained therein.

MACUSTAR consortium members: H. Agostini, L. Altay, R. Atia, F. Bandello, P. G. Basile, C. Behning, M. Belmouhand, M. Berger, A. Binns, C.J.F. Boon, M. Böttger, C. Bouchet, J.E. Brazier, T. Butt, C. Carapezzi, J. Carlton, A. Carneiro, A. Charil, R. Coimbra, M. Cozzi, D.P. Crabb, J. Cunha-Vaz, C. Dahlke, L. de Sisternes, H. Dunbar, R.P. Finger, E. Fletcher, H. Floyd, C. Francisco, M. Gutfleisch, R. Hogg, F.G. Holz, C.B. Hoyng, A. Kilani, J. Krätzschmar, L. Kühlewein, M. Larsen, S. Leal, Y.T.E. Lechanteur, U.F.O. Luhmann, A. Lüning, I. Marques, C. Martinho, G. Montesano, Z. Mulyukov, M. Paques, B. Parodi, M. Parravano, S. Penas, T. Peters, T. Peto, M. Pfau, S. Poor, S. Priglinger, D. Rowen, G.S. Rubin, J. Sahel, D. Sanches Fernandes, C. Sánchez, O. Sander, M. Saßmannshausen, M. Schmid, S. Schmitz-Valckenberg, H. Schrinner-Fenske, J. Siedlecki, R. Silva, A. Skelly, E. Souied, G. Staurengi, L. Stöhr, D. Tavares, J. Tavares, D.J. Taylor, J.H. Terheyden, S. Thiele, A. Tufail, M. Varano, L. Vieweg, J. Werner, L. Wintergerst, A. Wolf, N. Zakaria.

Disclosure: **M. Saßmannshausen**, Heidelberg Engineering (F), CenterVue (F), Carl Zeiss Meditec (F); **S. Thiele**, Allergan (R), Bayer (R), Carl Zeiss Meditec AG (F), CenterVue (F), Heidelberg Engineering (R, F), Optos (F), Novartis (R, F); **C. Behning**, None; **M. Pfau**, Apellis (C); **M. Schmid**, None; **S. Leal**, Bayer (E); **U.F.O. Luhmann**, F.

Hoffmann-La Roche Ltd (E); **R.P. Finger**, Alimera (C), Bayer (C, F), Boehringer-Ingelheim (C), Bioqen (F), CenterVue (F), Ellex (C), Roche/Genentech (C), Heidelberg Engineering (F), Novartis (C, F), Santhera (C), Zeiss (F); **F.G. Holz**, Acucela (C, F), Allergan (F), Apellis (C, F), Bayer (C, F), Boehringer-Ingelheim (C), Bioeq/Formycon (F, C), CenterVue (F), Ellex (F), Roche/Genentech (C, F), Geuder (C, F), Graybug (C), Gyroscope (C), Heidelberg Engineering (C, F), IvericBio (C, F), Kanghong (C, F), LinBioscience (C), NightStarX (F), Novartis (C, F), Optos (F), Oxurion (C), Pixium Vision (C, F), Oxurion (C), Stealth BioTherapeutics (C), Zeiss (F, C); **S. Schmitz-Valckenberg**, AlphaRET (C), Apellis (C, R), Bayer (F), Bioeq (C), Carl Zeiss MediTec (F), Heidelberg Engineering (F, R), Katairo (C), Kubota Vision (C), Novartis (C, F), Oxurion (C), Pixium (C), Roche (C, F), SparingVision (C), STZ GRADE Reading Center (O)

## References

1. Lim LS, Mitchell P, Seddon JM, Holz FG, Wong TY. Age-related macular degeneration. *Lancet*. 2012;379(9827):1728–1738.
2. Fleckenstein M, Keenan TDL, Guymer RH, et al. Age-related macular degeneration. *Nat Rev Dis Primers*. 2021;7(1):31.
3. Li JQ, Welchowski T, Schmid M, Mauschitz MM, Holz FG, Finger RP. Prevalence and incidence of age-related macular degeneration in Europe: a systematic review and meta-analysis. *Br J Ophthalmol*. 2020;104(8):1077–1084.
4. Colijn JM, Buitendijk GHS, Prokofyeva E, et al. Prevalence of age-related macular degeneration in Europe: the past and the future. *Ophthalmology*. 2017;124(12):1753–1763.
5. Polito A, Del Borrello M, Isola M, Zemella N, Bandello F. Repeatability and reproducibility of fast macular thickness mapping with stratus optical coherence tomography. *Arch Ophthalmol*. 2005;123(10):1330–1337.
6. Krzystolik MG, Strauber SF, Aiello LP, et al. Reproducibility of macular thickness and volume using Zeiss optical coherence tomography in patients with diabetic macular edema. *Ophthalmology*. 2007;114(8):1520–1525.
7. Joeres S, Tsong JW, Updike PG, et al. Reproducibility of quantitative optical coherence tomography subanalysis in neovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2007;48(9):4300–4307.
8. Patel PJ, Chen FK, Ikeji F, Tufail A. Intersession repeatability of optical coherence tomography measures of retinal thickness in early age-related macular degeneration. *Acta Ophthalmol*. 2011;89(3):229–234.
9. DeCroos FC, Toth CA, Stinnett SS, Heydary CS, Burns R, Jaffe GJ. Optical coherence tomography grading reproducibility during the comparison of age-related macular degeneration treatments trials. *Ophthalmology*. 2012;119(11):2549–2557.
10. Folgar FA, Jaffe GJ, Ying G-S, Maguire MG, Toth CA. Comparison of optical coherence tomography assessments in the comparison of age-related macular degeneration treatments trials. *Ophthalmology*. 2014;121(10):1956–1965.
11. Corvi F, Srinivas S, Gupta Nittala M, et al. Reproducibility of qualitative assessment of drusen volume in eyes with age related macular degeneration. *Eye (Lond)*. 2021;35(9):2594–2600.
12. Marsiglia M, Boddu S, Bearely S, et al. Association between geographic atrophy progression and reticular pseudodrusen in eyes with dry age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2013;54(12):7362–7369.
13. Christenbury JG, Folgar FA, O’Connell RV, Chiu SJ, Farsiu S, Toth CA. Progression of intermediate age-related macular degeneration with proliferation and inner retinal migration of hyperreflective foci. *Ophthalmology*. 2013;120(5):1038–1045.
14. Schmidt-Erfurth U, Klimescha S, Waldstein SM, Bogunović H. A view of the current and future role of optical coherence tomography in the management of age-related macular degeneration. *Eye (Lond)*. 2017;31(1):26–44.
15. Spaide RF, Curcio CA. Drusen characterization with multimodal imaging. *Retina*. 2010;30(9):1441–1454.
16. Sleiman K, Veerappan M, Winter KP, et al. Optical coherence tomography predictors of risk for progression to non-neovascular atrophic age-related macular degeneration. *Ophthalmology*. 2017;124(12):1764–1777.
17. Finger RP, Schmitz-Valckenberg S, Schmid M, et al. MACUSTAR: development and clinical validation of functional, structural, and patient-reported endpoints in intermediate age-related macular degeneration. *Ophthalmologica*. 2019;241(2):61–72.
18. Terheyden JH, Holz FG, Schmitz-Valckenberg S, et al. Clinical study protocol for a low-interventional study in intermediate age-related macular degeneration developing novel clinical endpoints for interventional clinical trials



- with a regulatory and patient access intention—MACUSTAR. *Trials*. 2020;21(1):659.
19. Ferris FL, Wilkinson CP, Bird A, et al. Clinical classification of age-related macular degeneration. *Ophthalmology*. 2013;120(4):844–851.
  20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
  21. Müller PL, Liefers B, Treis T, et al. Reliability of retinal pathology quantification in age-related macular degeneration: implications for clinical trials and machine learning applications. *Transl Vis Sci Technol*. 2021;10(3):4.
  22. Hogg RE, Silva R, Staurengi G, et al. Clinical characteristics of reticular pseudodrusen in the fellow eye of patients with unilateral neovascular age-related macular degeneration. *Ophthalmology*. 2014;121(9):1748–1755.
  23. Alten F, Clemens CR, Heiduschka P, Eter N . Characterisation of reticular pseudodrusen and their central target aspect in multi-spectral, confocal scanning laser ophthalmoscopy. *Graefes Arch Clin Exp Ophthalmol*. 2014;252(5):715–721.
  24. Ueda-Arakawa N, Ooto S, Tsujikawa A, Yamashiro K, Oishi A, Yoshimura N. Sensitivity and specificity of detecting reticular pseudodrusen in multimodal imaging in Japanese patients. *Retina*. 2013;33(3):490–497.
  25. Wu Z, Pfau M, Blodi BA, et al. OCT signs of early atrophy in age-related macular degeneration: interreader agreement: Classification of Atrophy Meetings Report 6. *Ophthalmol Retina*. 2022;6(1):4–14.
  26. 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(3):394–409.
  27. Bogunovic H, Montuoro A, Baratsits M, et al. Machine learning of the progression of intermediate age-related macular degeneration based on OCT imaging. *Invest Ophthalmol Vis Sci*. 2017;58(6):BIO141–BIO150.
  28. Waldstein SM, Vogl W-D, Bogunovic H, Sadeghipour A, Riedl S, Schmidt-Erfurth U. Characterization of drusen and hyperreflective foci as biomarkers for disease progression in age-related macular degeneration using artificial intelligence in optical coherence tomography. *JAMA Ophthalmol*. 2020;138(7):740–747.
  29. Saha S, Nassisi M, Wang M, et al. Automated detection and classification of early AMD biomarkers using deep learning. *Sci Rep*. 2019;9(1):10990.
  30. Liefers B, Taylor P, Alsaedi A, et al. Quantification of key retinal features in early and late age-related macular degeneration using deep learning. *Am J Ophthalmol*. 2021;226:1–12.
  31. Kim DY, Loo J, Farsiu S, Jaffe GJ Comparison of single drusen size on color fundus photography and spectral-domain optical coherence tomography. *Retina*. 2021;41(8):1715–1722.