Gradeability and Reproducibility of Geographic Atrophy Measurement in GATHER-1, a Phase II/III Randomized Interventional Trial

Angela S. Li, MD, Justin Myers, BS, Sandra S. Stinnett, DrPH, Dilraj S. Grewal, MD, Glenn J. Jaffe, MD

Objective: To validate GATHER-1 inclusion criteria and the study's primary anatomic end point by assessing the reproducibility of geographic atrophy (GA) measurements and factors that affect reproducibility.

Design: Post hoc analysis of phase II/III clinical trial.

Subjects: All 286 participants included in the GATHER-1 study.

Methods: For each subject, blue-light fundus autofluorescence (FAF), color fundus photographs, fluorescein angiograms, and OCT scans were obtained on the study eye and fellow eye. Geographic atrophy area and other lesion characteristics were independently graded by 2 experienced primary readers. If the 2 readers differed on gradeability, GA area (> 10%) or other lesion characteristics, the image was graded by an arbitrator whose measurement or characterization was the final grade.

Main Outcome Measures: The main outcome measures were gradeability and reproducibility of FAF imaging data. Imaging data included lesion area, confluence of GA with peripapillary atrophy (PPA), whether GA involved the foveal centerpoint, and type of hyperautofluorescence pattern.

Results: A total of 2004 images (1002 visits, 286 participants) were analyzed. Gradeability (90.5%) and interreader gradeability concordance (90.2%) were high across all visits. Study eye images were more gradable compared with fellow-eye images. A greater proportion of smaller lesions required arbitration, but interreader reproducibility was consistently high for all images. There was no difference in gradeability, gradeability concordance, or lesion-area concordance for images with PPA-confluent GA compared with those with nonconfluent PPA. Foveal centerpoint-involving lesions had lower gradeability and lesion-area concordance. Images with diffuse patterns of hyperautofluorescence had better gradeability and gradeability concordance than those with nondiffuse patterns but had no difference in lesion-area or lesion-area concordance.

Conclusions: There is high gradeability and excellent reproducibility measures across all images. These data support the validity of conclusions from GATHER-1 and the chosen inclusion criteria and end point.

Financial Disclosure(s): Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Science 2024;4:100383 © 2023 Published by Elsevier Inc. on behalf of 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.ophthalmologyscience.org.

Geographic atrophy (GA) is an advanced form of age-related macular degeneration (AMD) that may cause irreversible, severe vision loss.[1] Geographic atrophy is estimated to account for 20% of legal blindness (visual acuity of 20/200 or worse in both eyes) in the United States and has enormous impact on patients' functional status, quality of life, and independence.[2,3] In fact, patients who develop GA can lose 22 ETDRS visual acuity chart letters over 5 years.[4] The median time to develop central GA from the time of AMD diagnosis is only 2.5 years in 1 eye and 7 years in the fellow eye.[5] Geographic atrophy results in marked thinning and atrophy of retinal tissue from degeneration of macular photoreceptors, retinal pigment epithelial cells, and choriocapillaris.[2]

Unlike the multiple anti-VEGF therapies available for treatment of neovascular AMD, there is only 1 FDA-approved treatment for GA (intravitreal pegcetacoplan, a C3 inhibitor). Intravitreal avacincaptad pegol, a pegylated RNA aptamer that is a potent and specific inhibitor of complement C5, was studied in a phase II/III pivotal trial (GATHER-1).[6] It showed a statistically significant reduction in the mean GA growth rate over 12 months for both the 2 mg and 4 mg cohort when

compared with the corresponding placebo arms, and there were no major adverse events. Most recently, the pivotal, second phase III avacincaptad pegol study, GATHER2, met its primary study end point.[7] Besides pegcetacoplan, avacincaptad pegol is the only medication for GA that reduces GA lesion growth in phase III studies.[6] (*Investig Opthalmol Vis Sci.* 63:1500, 2022).

As the primary outcome for these pivotal trials, reduction in growth of GA, is based on the quantitative assessment of lesion area on autofluorescence imaging, there is a crucial need to assess the reproducibility of GA lesion-area measurements in these trials and determine whether this reproducibility supports the primary study end point. The present study examines data from the GATHER-1 study to evaluate overall gradeability and reproducibility of GA area as well as other parameters related to study inclusion criteria that could affect reproducibility, including the presence of peripapillary atrophy (PPA) confluent with GA, whether GA involved the foveal centerpoint, and a hyperautofluorescent pattern on fundus autofluorescence (FAF) imaging.

Methods

This study was approved by the institutional review board (Protocol 00046442) and adheres to the principles in the Declaration of Helsinki. Informed consent was obtained from all patients. All images from all GATHER-1 participants were included. For each subject, blue-light FAF, fundus photographs, and fluorescein angiograms were acquired on both eyes with the modified 3-field imaging protocol (field 1M: 30° field centered on the temporal aspect of the optic nerve; field 2: 30° field centered on the foveal center; and field 3M: 30° field centered 1−1.5 disc diameters temporal to the foveal centerpoint). A Heidelberg Spectralis or HRA (Heidelberg Engineering) system was used to obtain blue-light FAF (automatic real-time function = 15) and near-infrared, field-2 images. Spectral domain-OCT scans were obtained with Cirrus (Carl Zeiss Meditec) or Heidelberg Spectralis systems. Cirrus OCT scans were acquired with 512×128 macular cube and 5-line HD raster scan protocols. Spectralis scans were obtained with 97-line volume scan ($20° \times 20°$, high-resolution mode, ART = 9) and 73-line volume scan ($20° \times 15°$, high-resolution mode, ART = 9) protocols.

Masked readers at the Duke Reading Center independently analyzed and graded GA and associated features. Geographic atrophy was confirmed on OCT according to previously described criteria.[8] Readers used RegionFinder software (Heidelberg Engineering, version 2.6.4.0) to quantify GA area on FAF images from field 2.[9] Images from field 1M and field 3M were used to confirm that no GA extended beyond the boundaries of field 2, which would have rendered it ungradable for GA area. OCT and near-infrared imaging were used to supplement the FAF GA area analysis to help define GA boundaries and GA location relative to the foveal centerpoint. Two experienced, primary readers independently determined whether each variable was gradable, and, if not, the reason the variable could not be graded. To avoid reader bias at a given visit and ensure that the grades from each visit were independent from one another, the primary readers saved the markings as a PDF file to serve as a source document and then deleted all FAF image annotations on the RegionFinder-generated image as soon as they completed the grading. The readers were instructed not to refer to the original markings when grading subsequent visit images. Accordingly, the agreement at each visit did not depend on a previous visit.

Each reader assessed independently the following variables at each time point: area of GA from the RegionFinder report in mm^2, minimum distance from the GA lesion border to the foveal centerpoint (assessed on OCT), presence of PPA confluent with GA, presence of hyperautofluorescence on FAF, and the pattern of hyperautofluorescence on FAF (diffuse types including trickling, fine granular [dusty], fine granular [punctate], reticular, branching, and nondiffuse types, including banded, focal, or patchy).[10,11] For a study eye to be eligible, the total GA lesion area had to be between 2.5 and 17.5 mm^2 on initial visit, and, if multifocal, the area of at least 1 of the multifocal lesions had to be ≥ 1.25 mm^2. In addition, the GA also had to be nonfoveal-centered and contain either a banded or diffuse hyperautofluorescence pattern. Geographic atrophy lesions confluent with PPA were allowed.

The reproducibility of GA lesion area was assessed by inter-reader concordance ($< 10\%$ difference in sizes between 2 readers coefficient of reproducibility [CR]), also known as the smallest change in interreader measurement that can be interpreted as a true difference (also known as smallest real difference).[12]

If GA size was not concordant between readers or there was interreader disagreement in any of the other categorical variables, the image was arbitrated by the Reading Center Director (G.J.J.). Four certified primary readers and the Reading Center Director comprised the reader pool, although $> 95\%$ of all images were reviewed by 2 of the 4 primary readers. Across all variables, for discordant grades, the arbitrator's assessment was used as the final grade. For each participant, the study eye and fellow eye were analyzed in combination and separately.

For all gradeability analyses, interreader gradeability concordance, and interreader lesion-area concordance, either a chi-square test or a McNemar test (for paired data) was used to compare the 2 proportions. For quantitative data (lesion area or interreader absolute difference in mm^2), the nonparametric Wilcoxon rank sum test or Wilcoxon signed-rank test (for paired data) was used to compare the 2 distributions. All analyses comparing study eyes to fellow eyes were performed with paired tests. All statistical computations were performed in RStudio using the R programming language.

Results

Both eyes of all 286 participants in the GATHER-1 study (2004 images from 1002 participant visits) were included in this analysis. At each participant visit, both the study eye and fellow eye were imaged and subsequently analyzed.

Gradeability and Gradeability Concordance

Across all eyes, 1813 (90.5%) out of 2004 images were deemed gradable for GA area measurements (Fig 1A for a representative example). A significantly higher proportion of study eye images were gradable compared with fellow-eye images ($P < 0.001$, Table 1). Gradeability was uniformly high across all visits: 94.4% (540/572) for images obtained at the initial screening visit, 90.0% (434/482) at the 6-month visit, 90.7% (388/428) at the 12-month visit, and 88.9% (345/388) at the 18-month visit. Although the percentage of gradable images declined slightly from the initial screening visit to the 6-month visit ($P = 0.008$), there was no difference in percent gradeability between 6-month and subsequent visits.
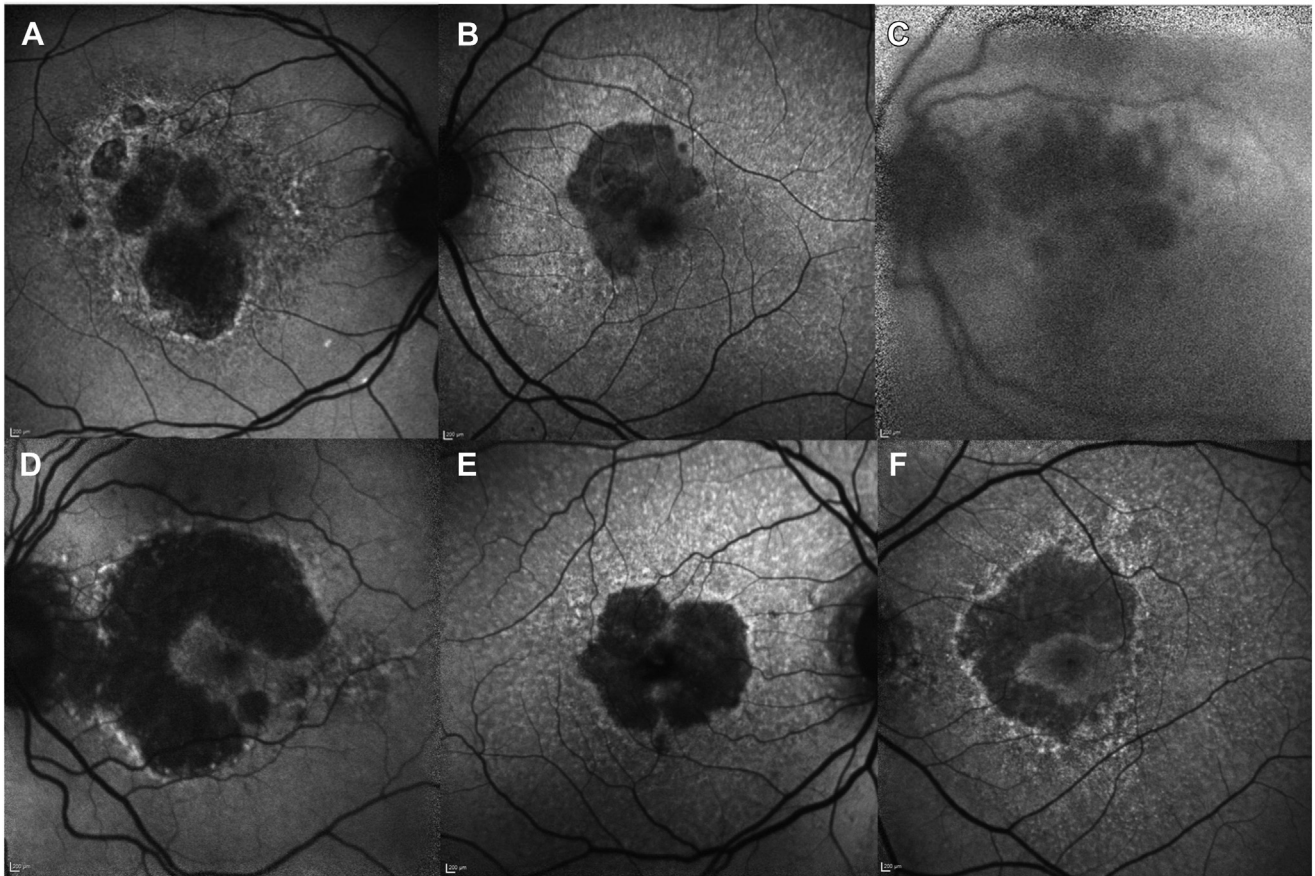
Figure 1. Example fundus autofluorescence images of gradable and ungradable geographic atrophy. **A,** Gradable image. **B,** Ungradable image because of indistinct margins around the fovea. **C,** Ungradable image because of poor image quality. **D,** Peripapillary atrophy confluent with geographic atrophy. **E,** Foveal centerpoint involvement. **F,** Diffuse, trickling, hyperautofluorescent pattern.

Interreader gradeability concordance (agreement between both readers on gradeability of an image), a measure of reproducibility, was excellent across all images and similarly high for both study eyes and fellow eyes (Table 1). In cases where the 2 readers disagreed on gradeability, the image was still gradable per the arbitrator in 72.1% of images. To enhance GA area measurement accuracy, readers used spectral domain-OCT and near-infrared reflectance images to define GA lesion borders. Among study images, gradeability was similarly high for those where GA measurement had been supplemented by Spectralis OCT (n = 861) compared with Cirrus OCT (n = 141), and there was no significant difference between gradeability concordance or size concordance for images whose gradeability was confirmed by either OCT system (Table S2, available at www.ophthalmologyretina.org).

There were 4 reasons that images were not gradable for GA area: ill-defined lesion borders, poor image quality, lesions exceeding image borders, or no observable GA (i.e., no GA seen on either FAF or OCT). Across the 191 ungradable images, the most common reasons were poor image quality (30.4%) or ill-defined lesion borders (30.4%).

Table 1. Gradeability and Interreader Gradeability Concordance

	All Images (N = 2004)	Study Eyes (n = 1002)	Fellow Eyes (n = 1002)	P Value
Gradable, n (%)	1813 (90.5)	946 (94.4)	867 (86.5)	P < 0.001
Gradeability concordance, n (%)	1807 (90.2)	907 (90.5)	900 (89.8)	P = 1.00

This table represents the proportion of gradable images (as defined by the arbitrator). The image is concordant if primary readers agree on its gradeability, and discordant if they disagree. *P* values compare the number of study eyes to fellow eyes using paired analysis. There is a significantly greater proportion of study eyes with gradable lesions compared with fellow eyes but no difference in gradeability concordance.

The most common reason among study eyes was poor image quality (51.8%), and the most common reason in fellow eyes was ill-defined lesion borders (34.8%) (Fig 1B, C). In addition, all 20 images with no observable GA were seen in fellow eyes (Table S3, available at www.ophthalmologyretina.org).

The Effect of GA Lesion Area on Gradeability and Gradeability Concordance

Across all gradable images (study and fellow eyes together), total GA area ranged from 0.115 mm^2 to 36.93 mm^2, with a median of 8.24 mm^2 and interquartile range of 5.13 mm^2 to 12.60 mm^2 (Fig 2). For nearly all study eyes, GA area ranged from 2.5 mm^2 to 17.5 mm^2 at the initial screening visit, which was an eligibility inclusion criterion. Of note, only a single, senior-reader GA measurement was used to determine if the image met eligibility criteria, but all GA measurements included in this analysis were based on subsequent dual-reader measurement. There were a total of 39 eyes (9 from study eyes and 30 from fellow eyes) for which the GA area became ungradable after the screening visit when the GA expanded to extend beyond the field 2 image borders. The average size at the screening visit for these 9 study eyes was 10.99 mm^2 (median, 12.81 mm^2). The average size at screening visit for the 30 fellow eyes was 10.97 mm^2 (median, 12.43 mm^2).

The mean GA area for all gradable images was 9.25 mm^2 (standard deviation [SD] = 5.42), with similar values of 9.09 mm^2 (SD = 4.54) in study eyes and 9.42 mm^2 (SD = 6.24) in fellow eyes.

There was no association between gradeability and GA lesion area. Overall, the mean GA size on images with discordant gradeability (as defined by the single reader who classified it as gradable) was similar to those that were gradable by both readers ($P = 0.98$). Among all with discordant gradeability, there was also no significant difference in the size of GA for images that were gradable per the arbitrator and those that were not ($P = 0.92$). Both of these observations also held true when study eyes and fellow eyes were analyzed separately (Table S4, available at www.ophthalmologyretina.org).

In GATHER-1, for a study eye to be eligible, the total GA lesion area had to be ≥ 2.5 mm^2, and, if multifocal, the area of at least 1 of the multifocal lesions had to be ≥ 1.25 mm^2. There was no difference in gradeability concordance for GA lesion area between 1.25 mm^2 and 2.5 mm^2 compared with GA size ≥ 2.5 mm^2. Gradeability and gradeability concordance for GA lesions between 1.25 mm^2 and 2.5 mm^2 compared with those ≥ 2.5mm^2 were similarly high, but the proportion of images with interreader size concordance was significantly lower when comparing those with lesion area between 1.25 mm^2 and 2.5 mm^2 compared with those ≥ 2.5 mm^2 (Table 5).

Finally, there was no significant lesion-area difference between gradable study eyes and the corresponding fellow eye in the same participant. The mean absolute raw difference in mm^2 between the study eye and the fellow eye on the same participant was 0.77 (study eye being larger), and the median difference was 0.38 (range, -24.67 to 21.21). The mean absolute percent difference was 5.22%, and the median difference was 2.38% ($P = 0.18$ for the paired comparison between study and fellow eye).

Lesion-Area Measurement Reproducibility

There were a total of 1490 images from 745 participant visits for which both the study eye and fellow eye were considered gradable by the arbitrator and 2 readers. These images were used to evaluate interreader reproducibility of lesion-area measurements. Lesion areas were concordant (difference < 10% between 2 readers) in 90.3% of these images, whereas the other 9.7% of images required arbitration because of interreader size-measurement difference > 10%. There was no difference in lesion-area concordance between study eyes and fellow eyes. The raw mean difference in lesion area (SD) as measured between the 2 readers was 0.126 mm^2 (0.720), and the mean interreader percent difference was 1.88% (7.76). The magnitudes of the differences in interreader lesion area, both absolute and percent, were similar in study eyes and fellow eyes (Table S6, available at www.ophthalmologyretina.org).

Smaller lesions were more likely to require arbitration for lesion area compared with larger lesions. For instance,
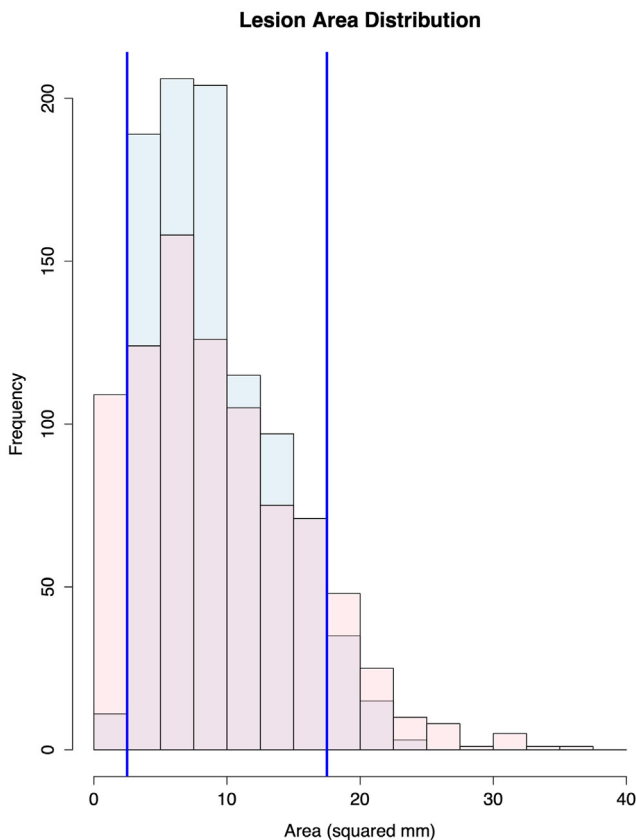

Figure 2. Distribution of geographic atrophy size across all gradable images. The blue bars represent the study eyes and the pink bars represent the fellow eyes. Areas of overlap are denoted in purple. Vertical blue lines correspond to the size inclusion criteria used in the study; on initial visit, all study eyes had to be between 2.5 mm^2 and 17.5 mm^2.

Table 5. Comparison of Reproducibility between 2 Different Size Cutoffs (1.25 mm^2 and 2.5 mm^2)

	< 1.25 mm^2	≥ 1.25 mm^2 and < 2.5 mm^2	≥ 2.5 mm^2	*P* Value
Gradable images (arbitrator only)	58	62	1693	
Gradeability concordance	52 (89.6%)	56 (90.3%)	1568 (92.6%)	0.50
Gradable images (arbitrator and both readers)	52	56	1563	
Size-concordant images	21 (40.4%)	29 (51.8%)	1356 (86.8%)	< 0.001

A total of 1813 images were gradable (per the arbitrator), and a total of 1671 images were gradable per the arbitrator and both readers. A difference of < 10% between the 2 readers' size measurements was considered size concordant. *P* values compare the difference between images with lesion size of 1.25 mm^2 to 2.5 mm^2 compared with those ≥ 2.5mm^2. There was no difference in gradeability or gradeability concordance between the 2 groups, but the smaller lesions did have lower size condordance.

59.6% of images with GA size < 1.25% mm^2 and 48.2% of images with GA size between 1.25 mm^2 and 2.5 mm^2 required arbitration, compared with only 13.4% of those with GA size ≥ 2.5mm^2 (*P* < 0.001; Table 5). In fact, GA size < 2.5 mm^2 made up 21.9% of all 265 size-discordant images, compared with only 3.41% of all 1406 size-concordant images (*P* < 0.001). Furthermore, the average lesion area (per the arbitrator) of all size-discordant images was 5.69 mm^2 (SD = 4.34) compared with 9.86 mm^2 (SD = 5.21) for all size-concordant images (*P* < 0.001). There was no significant difference in the proportion of size-concordant images with GA size < 1.25 mm^2 compared with those with GA size between 1.25 and 2.5 mm^2 (*P* = 0.32).

The coefficient of reproducibility (CR) (i.e., the value below which the absolute differences between 2 measurements would lie with 0.95 probability) was calculated using all 1643 gradable images that were interpreted by the 2 primary readers who read > 95% of all images. In this case, the CR was 1.47 mm^2, and 1600 of 1643 (97.4%) of the measurements were reproducible with interreader differences smaller than the CR. This held true regardless of GA area size. The other 43 (2.6%) of images were not reproducible per the CR, but there was no correlation between these nonreproducible images and lesion area (Fig 3A). The mean interreader difference across all 1643 images was 0.129 mm^2, similar to the mean interreader difference of 0.126 mm^2 for the subset of 1490 images. Furthermore, the CR was 1.32 mm^2 at the initial screening visit, 483 of 494 (97.8%) of the measurements were reproducible, and interreader differences were smaller than the CR (Fig 3B).

Association between GA Confluence with PPA, Gradeability, and Reproducibility

In GATHER-1, eyes could be included if there was PPA confluent with GA. Across all 2004 images, 1856 (92.6%) had measurable PPA. Of all eyes with PPA, 175 (9.4%) images had GA confluent with PPA (example in Fig 1D). These percentages were very similar for study eyes and fellow eyes. Overall, readers frequently agreed on whether an image had PPA confluent with GA; across all images, this percentage was 94.8%.

There was no association between gradeability or gradeability concordance and GA confluence with PPA

(Table S7, available at www.ophthalmologyretina.org). Specifically, across all images, eyes with confluent PPA had a slightly lower proportion of gradable images compared with those without confluent PPA, although this difference was not statistically significant after correction for multiple hypothesis testing. Among study eyes, PPA confluence was not associated with worse gradeability (*P* = 0.66). The gradeability concordance was also equally high for all eyes with confluent PPA compared with nonconfluent PPA (*P* = 0.41), and this held true for study eyes and fellow eyes.

The total GA lesion area among images with confluent PPA was significantly greater than that among images with nonconfluent PPA. Among all images, the average total lesion area for gradable images with confluent PPA was 14.33 mm^2 (SD = 5.41) compared with 8.88 mm^2 (SD =5.17 mm^2) with nonconfluent PPA (*P* < 0.001; Table S7, available at www.ophthalmologyretina.org). This difference was slightly less but still statistically significant for all study eyes. Notably, there was no association between confluence with PPA and interreader lesion-area concordance, both for all study participants and study eyes specifically.

Gradeability and Reproducibility of Foveal Centerpoint-Involving Lesions

For all study eye images taken with Spectralis, an additional analysis was performed to assess the minimum distance between the edge of the GA lesion and the foveal centerpoint. Those with a minimum distance of 0 were classified as involving the foveal centerpoint (example in Fig 1E). The minimum distance to the foveal centerpoint was quantifiable in 712 images. Of these, the foveal centerpoint was affected in 20.9% (149), and the foveal centerpoint was not affected in 79.1% (581).

Nearly all images undergoing foveal centerpoint analysis were by definition gradable according to the arbitrator, but the interreader gradeability concordance for those that involved the foveal-center point was significantly lower than those not involving the foveal-center point (*P* < 0.001; Table 8). The average size of foveal centerpoint-involving lesions was slightly larger at 10.02 mm^2 (SD = 5.08 mm^2), compared with the average size of nonfoveal centerpoint-involving lesions at 8.99 mm^2 (SD = 4.32 mm^2), although
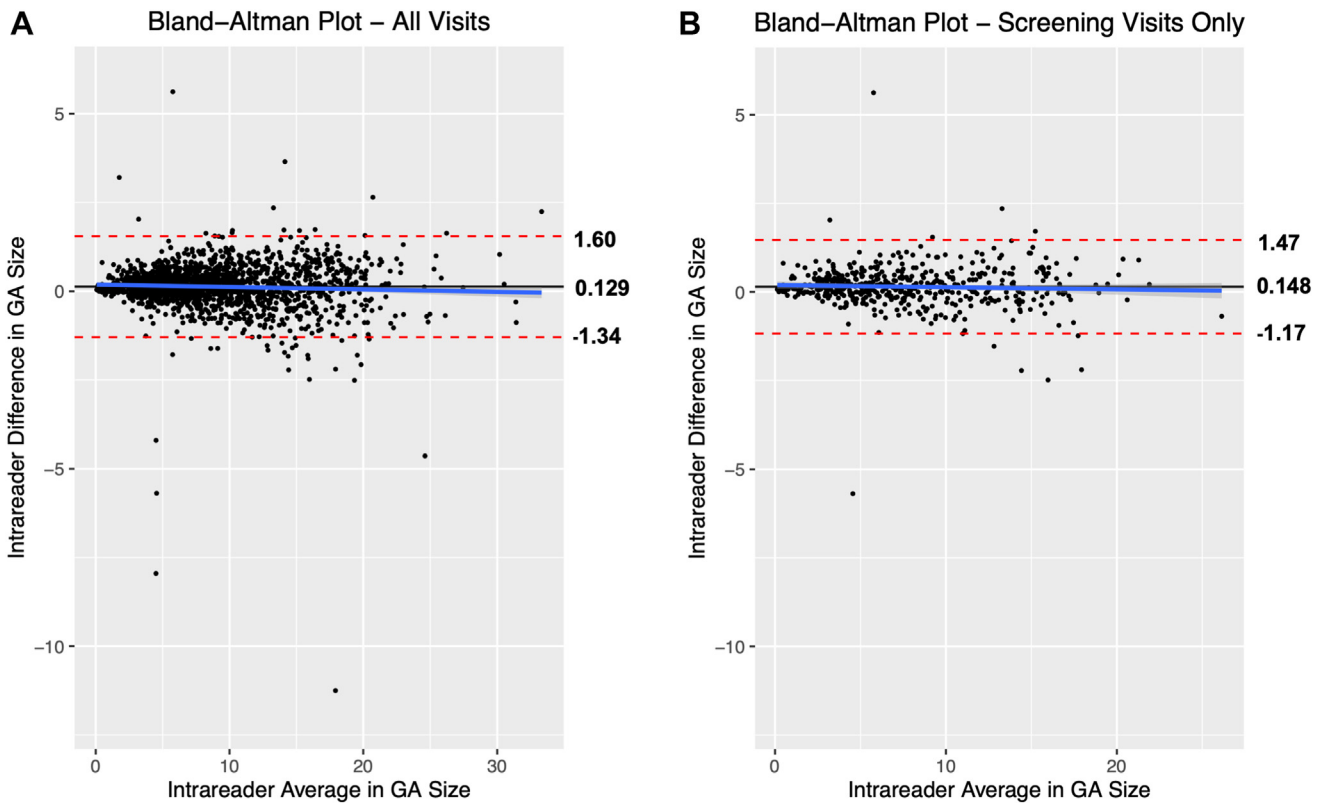
Figure 3. Bland Altman plot for all images (**A**) and screening-visit images (**B**) shows high interreader reproducibility across all image sizes. **A,** Across all images, the mean difference in lesion size between the 2 readers is 0.129 mm^2 (solid line), and dashed red lines represent the mean difference $+/-$ coefficient of reproducibility (CR), which was 1.47 mm^2. The vast majority (97.4%) of all interreader differences lay within the bounds of the CR and therefore are reproducible. **B,** For screening-visit images, the mean difference in lesion area between the 2 readers is 0.148 mm^2 (solid line), and the mean differences $\pm$ CR (1.32mm^2) are -1.17 mm^2 and 1.47 mm^2. Once again, 97.7% of all interreader differences lay within the range defined by the CR and, therefore, are reproducible. GA = geographic atrophy.

this difference was not statistically significant. For lesions that were gradable by the arbitrator and both readers, there was no significant difference in size concordance between lesions that involved the foveal-center point.

Gradeability and Reproducibility of Hyperautofluorescence Patterns

Of all images assessed, there was a determinable hyperautofluorescent pattern on FAF imaging in 79.7% (1598/2004), and the remaining 20.3% either did not have hyperautofluorescence or did not have a determinable pattern. Images were broadly categorized as having diffuse or banded hyperautofluorescence patterns or nondiffuse, nonbanded patterns. Among all images,1575 images had diffuse or banded patterns (example in Fig 1F), whereas 429 had nondiffuse, nonbanded patterns. Of all 1813 gradable images, 83.5% had diffuse or banded hyperautofluorescence patterns, and the other 16.5% had nondiffuse, nonbanded hyperautofluorescence patterns.

The vast majority (95.1%) of images with diffuse or banded patterns were gradable, compared with only 73.4% of images with nondiffuse, nonbanded patterns ($P < 0.001$; Table S9, available at www.ophthalmologyretina.org). The

Table 8. Associations between Lesions Involving the Foveal Centerpoint, Lesion Area, and Reproducibility

	Foveal Centerpoint-Involving	Minimum Distance to Fovea: 0–1.5 mm	P Value
Total study eyes with minimum distance analysis (n = 712)	149	563	
Gradeability concordance	122 (81.9%)	526 (93.4%)	< 0.001
Lesion area, mean (SD)	10.02 (5.08)	8.99 (4.32)	0.06
Lesion-area concordance	103/120 (85.8%)	450/526 (85.6%)	0.94

SD = standard deviation.
This analysis was only performed in 712 study eyes imaged with Spectralis. *P* values compare foveal centerpoint-involving lesions and those with a minimum distance to fovea > 0 mm. Images with lesions involving the foveal centerpoint were more likely to have lower gradeability concordance.

significant difference in gradeability also held true for both study eyes and fellow eyes. In addition, gradeability concordance was also significantly higher for diffuse patterns (91.8%) compared with nondiffuse patterns (86.5%; $P < 0.001$). This observation was also true for study eyes (92.1% vs. 86.0%; $P = 0.007$).

Discussion

In the current report, we assessed the gradeability and reproducibility of FAF imaging data in a randomized, controlled, interventional GA study. To the best of our knowledge, this type of analysis has not been reported previously in a clinical nonneovascular AMD GA trial. Overall, there was excellent reproducibility on gradeability and each reproducibility measure (gradeability concordance, lesion-area concordance, and CR) across all images.

In a clinical GA interventional trial, for which the GA area is the primary study end point, it is essential to have high quality, gradable images throughout the study to support the validity of the study results. In the current study, gradeability generally was very high throughout the study. The gradeability fell slightly between the initial screening (baseline) and 6-month visits but was still high at 6 months and did not decline further between 6 months, 12 months, and 18 months. There are several possible reasons that gradeability could decline over the course of a GA study. First, as lesion area increases over time, some of the lesions with GA margins near the image boundary at baseline may grow beyond the image boundaries so that it is not possible to measure lesion area. Indeed, we observed this phenomenon in 9 images in study eyes (7 subjects) and 30 fellow eyes (18 subjects). In addition, confounding pathology, such as cataracts or macular neovascularization that obscured the GA lesion boundaries, could have developed during follow-up. Furthermore, high-quality, gradable images were required as a study inclusion criterion, and it is possible that image quality could decline during follow-up visits. Indeed, a total of 29 of 56 (51.8%) of ungradable study images were because of poor image quality, and these increased over the course of the study. There were 6 poor-quality images at month 6, 7 at month 12, and 8 at month 18. An additional 8 images belonged to subjects who were terminated early from the study.

It is not possible to eliminate ungradable images in an interventional GA study or eliminate GA area measurement differences. However, we believe that several measures taken in GATHER-1 accounted for the overall high gradeability rate, and the high degree of concordance on GA lesion-area measurements. First, readers were trained by a rigorous process before they were approved to grade study images. For this training, readers read relevant literature on definitions of GA, underwent training on the use of RegionFinder software, reviewed with the Director of FAF Grading (G.J.J.) detailed in-house-developed GA grading instructions, and graded sample images whose grades were then reviewed by the Director of FAF Grading. The Director of Grading approved readers to grade study images once their image grades were comparable to that of other readers

and the Director of FAF Grading. Furthermore, an upper lesion area limit was 1 of the study inclusion criteria, which limited the number of eyes with GA that grew beyond the image boundaries. Eyes with cataracts at baseline that prevented good quality images, or those with cataracts likely to progress during the study were excluded. In addition, the Duke Reading Center−certified study site photographers were required to demonstrate their ability to obtain high-quality images before they could acquire study images. In addition, we provided prompt image quality feedback to site photographers on an ongoing basis to make the photographer aware of any correctable technical errors during image acquisition. Finally, the readers used OCT and near-infrared imaging to help define image boundaries, when needed, to assist in FAF GA area measurement, which helped to mitigate nongradeability based on indistinct lesion boundaries and ensure GA area measurement concordance.

Geographic atrophy lesion area did not affect image gradeability or gradeability concordance, but a higher proportion of smaller lesions (whether < 1.25 mm^2 or between 1.25 and 2.5 mm^2) exceeded a 10% difference between readers and, thus, required arbitration. However, GA lesion area for the majority of images, across all sizes, were considered reproducible because the raw interreader difference was within the limits of agreement as defined by the CR (no true difference between the 2 readers' size measurements). Smaller lesions had increased rates of arbitration simply because a smaller magnitude of interreader difference was more likely to be $> 10\%$ of lesion area. In fact, those that required arbitration had a significantly smaller lesion area compared with those that did not.

In GA trials, a lower size limit typically is chosen because lesions that are too small do not grow fast enough during a clinical trial to observe a treatment effect. However, it is then important that the chosen cutoff has a high degree of gradeability and lesion-area measurement reproducibility. There is no difference in gradeability concordance between the 2 cutoffs (minimum lesion area of at least 1 multifocal lesion > 1.25 mm^2 or total lesion area > 2.5 mm^2). Additionally, for either size cutoff (1.25 mm^2 or 2.5 mm^2), there was no difference in gradeability concordance when comparing lesions below or above the cutoff. This result supports the inclusion criteria used in GATHER-1, for which the minimum area of at least 1 multifocal GA lesion had to be ≥ 1.25 mm^2 and the total lesion area had to be ≥ 2.5 mm^2

In GATHER-1, GA lesions that were confluent with peripapillary GA were eligible for study inclusion, provided they met the other inclusion and exclusion criteria. Accordingly, it is important that these lesions can be measured reproducibly. In total, 92.6% of eyes had PPA, and, in 8.8% of eyes, the GA was confluent with the PPA, numbers that were similar to those that we observed in a GA natural history study, 86.4% and 7.1%, respectively.[13] We also confirmed that confluent PPA was associated with larger lesions, similar to that observed in the natural history study. In that study, we suggested that because of potential difficulty measuring GA lesion area in eyes with GA confluent with PPA or those likely to develop confluence that one should consider excluding those eyes from GA study enrollment. However, in the present

report, we found that there was no difference in gradeability, gradeability concordance, or lesion-area concordance between eyes with confluent PPA and those without. To increase the accuracy of the GA measurements, the readers established a predefined boundary at baseline between the PPA and the confluent macular GA that was used in subsequent follow-up visits, as we described previously.[13] We believe that this approach improved the GA area measurement reproducibility observed in GATHER-1 for confluent lesions.

Study eyes with foveal centerpoint involvement at the screening visit were excluded in GATHER-1. Lesions that affected the foveal centerpoint, either because of progression to the foveal centerpoint in study or fellow eyes or preexisting foveal centerpoint involvement in fellow eyes were more likely to have lower gradeability concordance, even though those involving the foveal centerpoint tended to be larger in size. This result is not surprising, because it is more difficult to determine GA lesion boundaries at the fovea. Blue-light autofluorescence was used in this study, and there is physiologic autofluorescence quenching by foveal xanthophyl pigment, which causes foveal hypoautofluorescence. When the pathologic GA hypoautofluorescence merges with the with physiologic foveal hypoautofluorescence, it may be difficult to distinguish the 2 types of hypoautofluorescence, rendering the GA boundary difficult to identify clearly. Ultimately, our data support GA trial designs, like GATHER-1, GATHER2 (NCT04435366), and the Derby and Oaks trials that exclude foveal-center involvement (*Investig Opthalmol Vis Sci.* 63:1500, 2022).

Diffuse hyperautofluorescent patterns are more gradable, but images without hyperautofluorescent patterns (or those with only focal or patchy patterns) should not necessarily be excluded on the basis of gradeability. Diffuse patterns have significantly better gradeability and gradeability concordance compared with nondiffuse hyperfluorescent patterns (either focal or none/not determinable). However, images with nondiffuse patterns of hyperfluorescence made up > 15% of all gradable images, so excluding these may eliminate a substantial proportion of gradable images. On the other hand, nondiffuse, nonbanded lesions tend to grow more slowly, which might make it more difficult to demonstrate a drug treatment effect if a large proportion of enrolled eyes have no associated hyperautofluorescent pattern.

Overall, these conclusions support the validity of the GATHER-1 results and the use of lesion area as an end point in GA trials. Furthermore, we believe the grading methodology used in the present study that resulted in a significant difference in lesion area growth in treated and sham groups, and the observations related to gradeability and lesion-area concordance will help inform the grading methods and inclusion/exclusion criteria for future GA studies.

Footnotes and Disclosures

Department of Ophthalmology, Duke University, Durham, North Carolina.

Disclosures:

All authors have completed and submitted the ICMJE disclosures form.

The authors have made the following disclosures: G.J.: Consultant − Roche/Genentech, Annexon, 4DMT, Regeneron.

D.G.: Consultant − Genentech, EyePoint, Regeneron.

HUMAN SUBJECTS: Human subjects were included in this study. This study was approved by the institutional review board (Protocol 0046442) and adheres to the principles in the Declaration of Helsinki. Informed consent was obtained from all patients.

No animal subjects were included in this study.

Author Contributions:

Conception and design: Jaffe

Data collection: Li, Myers, Grewal, Jaffe

Analysis and interpretation: Li, Stinnett, Grewal, Jaffe

Obtained funding: N/A

Overall responsibility: Li, Myers, Stinnett, Grewal, Jaffe

Abbreviations and acronyms:
AMD = age-related macular degeneration; **CR** = coefficient of reproducibility; **FAF** = fundus autofluorescence; **GA** = geographic atrophy; **PPA** = peripapillary atrophy; **SD** = standard deviation.

Keywords:
Geographic atrophy measurement, GATHER-1.

Correspondence:
Glenn J. Jaffe, MD, 2351 Erwin Rd, Durham, NC 27705. E-mail: jaffe001@mc.duke.edu.

References

1. Mitchell P, Liew G, Gopinath B, Wong TY. Age-related macular degeneration. *Lancet.* 2018;392:1147−1159.
2. Holz FG, Strauss EC, Schmitz-Valckenberg S, van Lookeren Campagne M. Geographic atrophy: clinical features and potential therapeutic approaches. *Ophthalmology.* 2014;121: 1079−1091.
3. Sivaprasad S, Tschosik EA, Guymer RH, et al. Living with geographic atrophy: an ethnographic study. *Ophthalmol Ther.* 2019;8:115−124.
4. Boyer DS, Schmidt-Erfurth U, van Lookeren Campagne M, et al. The pathophysiology of geographic atrophy secondary to age-related macular degeneration and the complement pathway as a therapeutic target. *Retina.* 2017;37:819−835.
5. Lindblad AS, Lloyd PC, Clemons TE, et al. Change in area of geographic atrophy in the Age-Related Eye Disease Study: AREDS report number 26. *Arch Ophthalmol.* Sep 2009;127: 1168−1174.

6. 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.

7. *Iveric Bio Announces Positive Topline Data from Zimura® GATHER2 Phase 3 Clinical Trial in Geographic Atrophy*; 2022. https://investors.ivericbio.com/news-releases/news-release-details/iveric-bio-announces-positive-topline-data-zimurar-gather2-phase. Accessed September 6, 2022.

8. 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. https://doi.org/10.1016/j.ophtha.2017.09.028.

9. Schmitz-Valckenberg S, Brinkmann CK, Alten F, et al. Semiautomated image processing method for identification and quantification of geographic atrophy in age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2011;52:7640−7646.

10. Holz FG, Bindewald-Wittich A, Fleckenstein M, et al. Progression of geographic atrophy and impact of fundus autofluorescence patterns in age-related macular degeneration. *Am J Ophthalmol*. 2007;143:463−472.

11. Fleckenstein M, Mitchell P, Freund KB, et al. The progression of geographic atrophy secondary to age-related macular degeneration. *Ophthalmology*. 2018;125:369−390.

12. Vaz S, Falkmer T, Passmore AE, et al. The case for using the repeatability coefficient when calculating test-retest reliability. *PLOS ONE*. 2013;8:e73990. https://doi.org/10.1371/journal.pone.0073990.

13. Chang P, Tan A, Jaffe GJ, et al. Analysis of peripapillary atrophy in relation to macular geographic atrophy in age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2016;57:2277−2282.