GREEN-LIGHT AUTOFLUORESCENCE VERSUS COMBINED BLUE-LIGHT AUTOFLUORESCENCE AND NEAR-INFRARED REFLECTANCE IMAGING IN GEOGRAPHIC ATROPHY SECONDARY TO AGE-RELATED MACULAR DEGENERATION

Maximilian Pfau,1,2 Lukas Goerdt,1 Steffen Schmitz-Valckenberg,1,2 Matthias M. Mauschitz,1,2 Divyansh K. Mishra,3 Frank G. Holz,1,2 Moritz Lindner,1,4 and Monika Fleckenstein1,2

1Department of Ophthalmology, University of Bonn, Bonn, Germany
2GRADE Reading Center, Bonn, Germany
3Sankara Eye Hospital, Varthur Main Road, Kundalahalli Gate, Bangalore, Karnataka, India
4The Nuffield Laboratory of Ophthalmology, Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

Correspondence: Monika Fleckenstein, Department of Ophthalmology, University of Bonn, Ernst-Abbe-Str. 2, Bonn 53127, Germany; Monika.Fleckenstein@ukbonn.de.

MP and LG contributed equally to the work presented here and should therefore be regarded as equivalent authors.

Submitted: February 28, 2017
Accepted: May 11, 2017

PURPOSE. To compare the intermodality and interreader agreement for geographic atrophy (GA) lesion size quantification in green-light fundus autofluorescence (GAF; excitation = 518 nm) versus combined blue-light fundus autofluorescence (BAF; excitation = 488 nm) and near-infrared reflectance (NIR; 820 nm) –based grading.

METHODS. Confocal scanning laser ophthalmoscopy (cSLO) GAF, BAF, and NIR images of 40 eyes from 29 patients (mean age 79.7 years) with GA secondary to AMD were recorded according to a standardized protocol. GA areas were analyzed in GAF, BAF combined with NIR (BAF+NIR), or BAF alone, by four independent readers using semiautomated software (RegionFinder; Heidelberg Engineering, Heidelberg, Germany). A mixed-effects model was used to assess the effect of image modality on the measured square-root lesion area. The coefficient of repeatability (CR) and intraclass correlation coefficient (ICC) were assessed for the square-root lesion area, lesion perimeter, and circularity.

RESULTS. GAF-based measurements were on average 0.062 mm (95% confidence interval [CI] 0.04–0.08 mm) larger than BAF+NIR-based measurements and 0.077 mm (95% CI 0.06–0.10 mm) larger than BAF-based measurements. Interreader agreement was highest for GAF-based analysis ([CR, ICC] 0.196 mm, 0.995) followed by BAF+NIR (0.232 mm, 0.992) and BAF alone (0.263 mm, 0.991). The same was noted for the lesion perimeter and circularity. Post hoc review revealed that interreader differences were associated with media opacification interfering with lesion boundary demarcation to a larger extent in BAF than in GAF.

CONCLUSIONS. cSLO-based GAF and combined BAF+NIR imaging with semiautomated lesion delineation allow for an accurate and reproducible quantification of GA. The slightly better interreader agreement using cSLO GAF suggests that its use may be preferable in clinical trials examining the change in lesion size as a clinical endpoint.

Keywords: geographic atrophy, age-related macular degeneration, fundus autofluorescence, green-light autofluorescence
spared fovea in the NIR image and subsequent semi-automated quantification of GA areas. Further, the so-called “shadow correction” can be used for the assessment of the fovea in BAF images.

In contrast to BAF imaging, green-light autofluorescence (GAF, excitation 518 nm) cSLO imaging is not significantly affected by macular pigment due to a lack of absorption. Thus, GAF imaging would probably result in an even more precise assessment of small, central changes, including the differentiation between foveal atrophy and foveal sparing. Only one previous study has compared BAF with GAF imaging in GA. However, this study did not compare combined BAF+NIR imaging (as used in currently ongoing clinical trials) with GAF imaging. Furthermore, no manual constraints and no “shadow correction” were used to exclude regions of foveal sparing from the lesion area measured. Recently, the lesion perimeter (lesion circumference) and lesion circularity (the ratio of area to perimeter squared) have been reported to be prognostic biomarkers for upcoming GA progression (Pfau M, et al. IOVS 2016;57:ARVO E-Abstract 1613). However, no data on the interreader agreement of these biomarkers in GAF, BAF, and NIR imaging have so far been published.

The aim of this study was to systematically compare the intermodality and interreader agreement for cSLO GAF, cSLO BAF, and cSLO NIR, respectively, based on semi-automated delineation of GA in a reading center setting. We tested the hypothesis that there are no differences in lesion size measurements among the assessed modalities. Further, we hypothesized that GAF-based grading exhibits the highest interreader agreement for the lesion area, perimeter, and circularity, because fewer manual constraints are necessary with regard to measuring the foveal region.

**Methods**

**Patients**

Patients were recruited from the prospective longitudinal, natural history DSGA (Directional Spread in Geographic Atrophy) and cross-sectional SIGHT (Sparing of the Fovea in Geographic Atrophy Progression) studies (NCT02051998 and NCT02332343; http://clinicaltrials.gov). The inclusion and exclusion criteria for DSGA have been described previously. For inclusion into SIGHT, the study eye had to show contiguous well-demarcated GA either in a complete ring around the spared fovea or in a horseshoe pattern. Patients had to exhibit either uni- or multifocal GA in at least one eye. Exclusion criteria included any history of retinal surgery, laser photocoagulation, radiation therapy, or other retinal diseases in the study eye as well as eyes with an area of atrophy exceeding the $30^\circ \times 30^\circ$ cSLO image frame. If both eyes of a patient met the inclusion criteria, both eyes were selected as study eyes.

**Imaging**

BAF and NIR images were obtained with an HRA 2 or Spectralis (Heidelberg Engineering, Heidelberg, Germany) device. BAF images were taken with an excitation wavelength of 488 nm and an emission spectrum of 500 to 700 nm using the high-speed mode. NIR images were obtained at 820-nm wavelength. Further, GAF images (GAF excitation 518 nm) were obtained using the Spectralis device. The field of view was set to $30^\circ \times 30^\circ$ with a resolution of $768 \times 768$ pixels and was centered on the fovea. Single BAF, NIR, and GAF images were automatically aligned and averaged (up to 100 single frames) to maximize the signal-to-noise ratio using the manufacturer’s software.

**Grading**

The readers (R1, R2, R3, and R4) were trained according to reading center standard operating procedures (GRADE Reading Center, Bonn, Germany). Measurements of atrophy areas were performed using the RegionFinder software (version 2.6.3; Heidelberg Engineering) as previously described. Briefly, the readers were asked to set at least one seeding point inside of each atrophic region by selecting the pixel with the lowest FAF signal (darkest gray value). Thereafter, the readers had to increase the growth power for each seeding point, which resulted in the inclusion of adjacent pixels depending on the gray value, until the delineation just exceeded the lesion boundaries. Finally, the growth power had to be decreased by one increment below this threshold. The growth limit function was used if the segmentation algorithm included the edges of the image frame. Further, retinal vessels or macula pigment were excluded from the measured lesion area through vessel detection and “shadow correction” or by placing manual constraints. For BAF+NIR grading in foveal sparing, the readers were asked to delineate the spared fovea in the NIR image semiautomatically before semiautomated quantification of GA areas in the corresponding BAF image.

Each visit was graded by each reader with BAF images only, BAF+NIR images, and GAF images only. The grading task was carried out on separate days and in random order. With the currently available software version, combined GAF+NIR-based grading was not possible. The graded annotated images were transferred to ImageJ (http://image.nih.gov/ij/) provided in the public domain by the National Institutes of Health, Bethesda, MD, USA) to measure the (cumulative) lesion circularity and (cumulative) lesion perimeter using a custom-built plug-in (Pfau M, et al. IOVS 2016;57:ARVO E-Abstract 1613). Further, eyes were classified into foveal atrophy, extrafoveal atrophy, and foveal sparing according to the extent of GA in/near the fovea. Foveal sparing was defined as an intact, residual foveal island being surrounded by more than 270° of well-demarcated GA areas.

**Outcome Measures and Statistical Analyses**

Statistical analyses were performed using the software environment R. Area measurements were square-root transformed to obtain normally distributed data. A mixed-effects model considering imaging modality as fixed effect (GAF versus BAF+NIR versus BAF) and visit as well as reader as random effects was used to assess whether measured lesion size is dependent on the image modality. For each imaging modality (GAF, BAF+NIR, and BAF) the intraclass correlation coefficient (ICC; two-way random, absolute agreement), the 95% coefficient of repeatability (CR), and the coefficient of variation (CV) were determined. Moreover, the ICC, CR, and CV were also determined for the perimeter and circularity measurements. For visualization, Bland-Altman graphs were plotted. Spearman’s rank correlation coefficient ($\rho$) was calculated between the absolute differences and the mean values to determine whether measurement variability increases with lesion size.

**Results**

**Cohort Characteristics**

A total of 40 visits of 40 eyes from 29 patients (age [mean ± SD] 79.7 ± 6.2 years, 20 female) with GA secondary to AMD were included and graded (Fig. 1). Foveal sparing was present in 32 (55%) of 40 of these eyes. Of the 40 eyes, 31 (77%) were pseudophakic (Table 1; Fig. 2).
Lesion Size in Dependence of Grading Modality

A mixed-effects model considering reader and visit as random effects disclosed that the grading modality (GAF versus BAF+NIR versus BAF) significantly affected lesion size measurements ($\chi^2(2) = 55.257, P < 0.001$). Hereby, the square-root lesion area was on average 0.062 mm (95% CI 0.04–0.08 mm) larger for GAF-based measurements than for BAF+NIR-based measurements. Similarly, GAF-based measurements were on average larger by 0.077 mm (95% CI 0.06–0.10 mm) than BAF-based measurements. There were no significant differences in the square-root lesion areas between BAF- and BAF+NIR-based measurements (0.01 mm; 95% CI –0.01 to 0.04 mm).

The differences between the measurements were plotted against their respective mean value (Bland-Altman plots) for graphic analysis (Fig. 3). To assess whether measurement variability increased with lesion size, the Spearman’s rank correlation coefficient ($\rho$) for absolute differences and mean values was calculated. It indicated for GAF- versus BAF+NIR- ($\rho = 0.195, P = 0.23$), for GAF- versus BAF- ($\rho = 0.172, P = 0.29$), and for BAF+NIR- versus BAF-based ($\rho = 0.148, P = 0.36$) measurements that lesion size did not significantly affect the intermodality variability (Fig. 3). In line with the mixed-effects model, the mean differences of the Bland-Altman plots indicated that GAF-based measurements were larger than BAF+NIR- (0.062 mm) or BAF-based (0.077 mm) measurements (Fig. 3).

Because this difference between GAF- and BAF+NIR- or BAF-based grading was largely caused by five visits from five eyes (Fig. 3), a detailed post hoc analysis of the images was carried out. The eye with the greatest GAF-BAF discrepancy is shown in Figure 4. The 84-year-old, pseudophakic, female patient presented with posterior capsular opacification. The contrast of the lesion as compared with the background signal was higher for the GAF than the BAF image. Especially the temporal lesion boundary was better demarcated in the GAF image. Because readers were instructed to increase the growth power of each seed until the defined area exceeded the lesion boundaries, the measurements tended to be slightly larger for the GAF images. In the BAF grading, the readers had to stop increasing the growth power prematurely due to low-contrast segments of the lesion boundary. Further, some foci of...
questionably decreased autofluorescence (especially at the nasal margin of the lesion) were visible only in the GAF image (Fig. 4). Figure 5 shows another eye with very low intermodality agreement. Both the GAF and BAF image did not allow for an accurate delineation of the lesion, whereas the lesion boundaries were clearcut in the NIR image. The assessment of foveal atrophy also resulted in some interreader differences. As shown in Figure 2, foveal GA foci can have a similar appearance compared with macular pigment in BAF images. This led to omission of foveal GA foci in some BAF-based measurements or to an incorrect grading taking macular pigment for atrophy.

**Interreader Agreement for Lesion Size Measurements**

The CR (i.e., the value below which the difference between two measurements will lie with a probability of 0.95) for the square-root area was 0.196 mm for the GAF-based grading, 0.232 mm for the BAF+NIR-based grading, and 0.263 mm for the BAF-based grading. Likewise, the CV (2.87% for GAF, 3.49% for BAF+NIR, 3.98% for BAF) and ICC (0.995 for GAF, 0.992 for BAF+NIR, 0.991 for BAF), which take into account the underlying lesion size, indicated that GAF-based grading has the highest interreader agreement (Table 2, Figure 6).

The highest interreader variability was observed for BAF-based measurements in the subset of eyes with foveal sparing (CR of 0.274 mm) followed by the BAF-based measurements in the subset of eyes without foveal sparing (CR of 0.265 mm). In contrast, in BAF+NIR-based measurements (CR of 0.218 [foveal sparing] and 0.248 [nonfoveal sparing]) and GAF-based measurements (CR of 0.175 [foveal sparing] and 0.211 [nonfoveal sparing]), the interreader variability was lower in eyes with foveal sparing as compared with eyes without foveal sparing.

**Interreader Agreement for Lesion Circularity and Perimeter**

Despite equal lesion area measurements, the actual underlying delineations may differ, because minor grading differences seem to balance out. Therefore, the lesion perimeter (cumulative circumference) and lesion circularity were analyzed with regard to interreader reliability, as these lesion shape descriptive factors are more susceptible to small differences of the actual underlying delineations. For the perimeter, the GAF-based grading (CR = 3.92 mm; CV = 6.94%; ICC = 0.983) exhibited the highest interreader agreement followed by the BAF+NIR-based grading (CR = 5.04 mm; CV = 9.03%; ICC = 0.972) and BAF-based grading (CR = 5.25 mm; CV = 9.69%; ICC = 0.971). Likewise, GAF-based grading exhibited the best interreader agreement for lesion circularity (Table 3).

Major sources of interreader disagreement regarding perimeter and circularity were the extent of foveal involvement for
BAF-based grading as exemplified in Figure 2. The foveal involvement was difficult to assess using only BAF due to macular pigment interference. Even with the shadow correction tool, which partially allowed for assessment of foveal involvement, the delineation of the spared fovea differed among readers because manual constraints had to be used. In contrast, BAF+NIR imaging allowed for an accurate recognition of foveal sparing; however, (semi)automated constraints had to be used to delineate the boundary of the spared fovea. In GAF-based grading, the least amount of constraints had to be used, as illustrated in Figure 1. Generally, the use of constraints for BAF, BAF+NIR, and GAF delineation, which was necessary for GA measurements in some eyes, appeared to be associated with a lower interreader agreement. Thus, GAF-based measurements relied mostly on the (semi)automatically identified boundaries and were least dependent on manual or semi-automated constraint placements resulting in the highest interreader agreement.

**DISCUSSION**

This study demonstrates that GAF and BAF+NIR imaging allow for an accurate and reproducible quantification of GA lesions, and, therefore, qualify as measurement tools for clinical trials testing the efficacy of interventions aiming at a slowing down of GA progression. Hereby, GAF-based quantification exhibited the best interreader agreement. BAF-based measurements also resulted in an excellent interreader agreement. Yet, the interreader agreement was markedly lower than those obtained from GAF or BAF+NIR imaging, especially when measuring circularity and perimeter. Besides, GAF-based lesion size measurements tended to be minimally larger than BAF+NIR- (or BAF-) based measurements.

To date, only one study compared BAF versus GAF imaging in GA. However, this study did not compare GAF-based grading with BAF+NIR-based grading, which serves as the primary outcome measure in currently ongoing phase II and III trials (NCT02247531, NCT02247479, NCT02087085). Further, the authors concluded that lesion sizes in BAF images were larger than in GAF images because of centrally decreased blue-light autofluorescence due to macular pigment. However, it is conceivable that the newer version of the Region-Finder software with “shadow correction” and manual constraints excludes more precisely regions of foveal sparing from the measured lesion area. Indeed, in our study, there was no relevant mean difference between BAF+NIR- and BAF-based grading. The fact that GAF-based grading resulted in minimally larger lesion size measurements than BAF- or BAF+NIR-based grading is most likely attributable to the minimally sharper contrast at lesion boundaries in a subset of GAF images. The readers were asked to set seeding points inside of atrophic regions, then to increase the growth power.
until the delineation exceeded the lesion boundaries and finally to decrease the growth power by one increment below this threshold.10 Sharper lesion boundaries allowed for a greater increase of the growth power, whereas ill-defined lesion boundaries forced readers to restrict the growth power prematurely and to use manual constraints.

The underlying reason for the higher contrast in GAF as compared with BAF images could be partially attributed to the aging crystalline lens (23% of the included eyes were phakic) reducing the transmission of short-wavelength light.21 Other media opacities, including posterior capsular opacification and vitreous floaters, appeared to affect the quality of BAF images more severely than that of GAF images. In addition, GAF images were more often in perfect focus than BAF images. Usually, the focus is initially adjusted in the NIR mode (for patient comfort) and then quickly re-adjusted for chromatic aberration after switching to the GAF or BAF mode. Because the optimal focus for GAF is closer to the focus of NIR, re-adjustment is easier for GAF imaging. Finally, patients tend to blink less (patient comfort) during GAF than BAF imaging, which facilitates the acquisition of high-quality images.

BAF-based grading in eyes with foveal sparing exhibited the highest interreader variability in this study. BAF+NIR-based grading, which allowed for a semiautomatic delineation of the residual foveal island in NIR images, resulted in a better interreader agreement as compared with BAF-based grading underscoring the importance of semiautomation of the grading process.11 The highest interreader agreement was observed for GAF images that required the least constraints, highlighting the importance of semiautomatic versus manual delineation for the interreader agreement. Especially in clinical trials, it is crucial to maximize the interreader agreement, because effect sizes are dependent on the underlying measurement variability.20 Thus, interreader agreement affects directly sample size determination.

Further studies will be needed to compare GAF imaging with other image modalities. The recently published Classification of Atrophy Meeting consensus recommended the use of color fundus photography (CFP), BAF, NIR, and optical coherence tomography (OCT) in studies with GA.22 The interreader agreement of these modalities has been assessed previously and reported to be high in a number of studies (ICC values ranging from 0.95 [for CFP] to 0.99 [for BAF and OCT])10,23–28; however, the ICC was commonly reported as the only outcome measure. It is difficult to compare the ICC across different study cohorts, because it is dependent on the variance of the trait (i.e., lesion size) within the cohort.19 Addition of a small number of eyes with either very large or very small GA lesions would result in markedly improved ICC values irrespective of the underlying image modality or grading method. The CR as recommended by Bland and Altman was used in our study because it is independent of the average lesion size and may be compared with different study cohorts in a more meaningful manner.20 With the advent of faster spectral-domain and swept-source OCT devices, OCT imaging appears to be a potential alternative to BAF or GAF imaging in

**Figure 4.** GAF versus BAF. The BAF (A) and the GAF (B) images of this 84-year-old, pseudophakic, female patient with posterior capsular opacification exhibited the greatest discrepancies. At location 1 (green box 1), the GAF images show atrophy, whereas no distinct patch of decreased autofluorescence can be seen in the BAF image (D1). Overall, the contrast at the temporal lesion boundary is markedly higher in the GAF than the BAF image and the NIR image (C). Toward the optic disc, multiple foci of questionably decreased autofluorescence are seen in the GAF image. These are not visible in the BAF image (D2).
Typically, OCT-based segmentation methods rely on en face fundus or sub-RPE projection images that depict so-called hypertransmission into the choroid in regions of GA. Hereby, large choroidal vessels are typically hyposcattering and may result in segmentation artifacts. Thus, future studies should evaluate automated OCT-based segmentation in comparison with BAF or GAF images, as the latter usually depict a higher contrast than OCT images. Noteworthy, one previous study reported that the agreement between automatically and manually defined GA regions was better for BAF than OCT images.

Finally, GAF imaging tended to be more comfortable for patients than BAF imaging (anecdotal evidence). Furthermore, in ABCA4-associated retinopathy, there is some controversy with regard to BAF imaging, as it was speculated that it may accelerate accumulation of A2E and, thus, induce apoptosis in RPE cells with particularly high levels of A2E as observed in Abcr knockout mice and cell culture, respectively. Although to date there is no evidence of phototoxic effects in humans, both in the absence or presence of retinal diseases. Cideciyan and associates proposed to apply reduced-illumination BAF imaging to reduce light exposure, and, thus, to reduce a potential risk for adverse effects. Noteworthy, in cultured human RPE cells with internalized A2E, illumination with green light was shown to result in substantially fewer nonviable cells as compared with illumination with blue light. Based on these results obtained by ex vivo analysis and in animal models, it may be speculated that GAF imaging...
might be safer in patients with RPE atrophy (especially in association with ABCA4-associated retinopathy).

Limitations of this study must be considered. First, the current version of the RegionFinder software does not allow for combined GAF+NIR grading. Potentially, combined GAF+NIR grading would further increase the interreader reliability in a subset of patients (cf., Fig. 5). Second, the grading of the three modalities was performed by all readers in random order and on separate days. However, it cannot be fully excluded that readers re-recognized eyes. Third, OCT imaging, which is the most promising image modality besides BAF and GAF imaging in GA, was not included in this study.

In summary, this study demonstrated that GAF and combined BAF+NIR imaging allow for reliable assessment of lesion size and shape in GA secondary to AMD, both clinically and particularly in clinical studies. Hereby, GAF-based quantification exhibits higher interreader agreement. Because media opacification appears to interfere with lesion-demarcation more strongly in BAF than in GAF, minor differences in lesion size measurements between the different analysis approaches must be considered.

Acknowledgments

The authors thank Jan Dechent and Martin Diller (Heidelberg Engineering GmbH, Heidelberg, Germany) for the technical support. The authors also thank Suzan E. Hunt for language editing and proofreading.

Supported by BONFOR GEROK Program, Faculty of Medicine, University of Bonn, Grants O-137.0020 (ML) and O-137.0022 (MP);
References


