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], Monika Fleckenstein[1,2]

* These authors contributed equally to this work.

1. Department of Ophthalmology, University of Bonn, Ernst-Abbe-Str. 2, Bonn, Germany
2. GRADE Reading Center, Ernst-Abbe-Str. 2, Bonn, Germany
3. Sankara Eye Hospital, Varthur Main Road, Kundalahalli Gate, Bangalore, Karnataka 560037, India
4. The Nuffield Laboratory of Ophthalmology, Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

Running head: Green- vs. blue-light autofluorescence in geographic atrophy
Key words: Geographic atrophy, age-related macular degeneration, fundus autofluorescence, green-light autofluorescence
Number of words: 3,341 (excluding title page, abstract, legends, and references)
Number of figures: 6
Number of tables: 3

Correspondence:
PD Dr. Monika Fleckenstein
Department of Ophthalmology
University of Bonn
Ernst-Abbe-Str. 2
53127 Bonn
Germany
Tel.: +49 228 287 16826
Fax: +49 228 287 11470
E-Mail: Monika.Fleckenstein@ukbonn.de

Abstract

Purpose

To compare the inter-modality and inter-reader agreement for geographic atrophy (GA) lesion size quantification in green-light-fundus-autofluorescence- (GAF, excitation=518nm) versus combined blue-light-fundus-autofluorescence (BAF, excitation=488nm) and near infrared reflectance- (NIR, 820nm) 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 age-related macular degeneration 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 a semi-automated 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 0.062mm (95%CI 0.04–0.08mm) larger than BAF+NIR-based measurements and 0.077 mm (95% CI 0.06 – 0.10 mm) larger than BAF-based measurements. Inter-reader agreement was highest for GAF-based analysis ([CR, ICC] 0.196mm, 0.995) followed by BAF+NIR (0.232mm, 0.992) and BAF alone (0.263mm, 0.991). The same was noted for the lesion perimeter and circularity. Post-hoc review revealed that inter-reader differences were associated

69 with media-opacification interfering with lesion-boundary-demarcation to a larger

70 extent in BAF than in GAF.

71 **Conclusions**

72 CSLO-based GAF and combined BAF+NIR imaging with semi-automated lesion

73 delineation allow for an accurate and reproducible quantification of GA. The slightly

74 better inter-reader agreement using cSLO GAF suggests that its use may be

75 preferable in clinical trials examining the change in lesion size as a clinical endpoint.

Introduction

Geographic atrophy (GA) is the non-neovascular late-stage manifestation of age-related macular degeneration (AMD).[1,2] Currently, no approved therapy is available for GA while multiple interventional clinical trials are ongoing.[3] Atrophy of the outer retina and retinal pigment epithelium (RPE) are characteristic for GA and may also develop in presence of the neovascular manifestations (choroidal neovascularization [CNV]) leading to a significant long-term vision loss despite treatment with anti-vascular endothelial growth factor (VEGF) agents.[2–4]

GA size quantification using blue-light autofluorescence (BAF, excitation 488 nm, emission 500-700 nm) confocal scanning laser ophthalmoscopy (cSLO) imaging combined with near-infrared reflectance (NIR, 820 nm) cSLO imaging is the primary outcome measure in various ongoing clinical trials investigating GA (NCT02247531, NCT02247479, NCT02087085, http://clinicaltrials.gov). The loss of RPE and its inherent fluorophores in GA correlates with well-defined areas of decreased autofluorescence .[5,6] Manual, semi-automatic and automatic GA segmentation methods for BAF images have been described.[7–13] The semi-automatic region-growing image analysis approach has been integrated in the RegionFinder™ software (Heidelberg Engineering, Heidelberg Germany).[9–11] Evaluation of the fovea in foveal-sparing GA with BAF imaging may be challenging, since macular pigment (lutein, zeaxanthin, and meso-zeaxanthin) absorbs short-wavelength excitation light.[14,15] Thus, automated registration of BAF and NIR images has been incorporated into the software to allow for semi-automated delineation of the spared fovea in the NIR image and subsequent semi-automated quantification of GA areas.[11] Further, the so-called 'shadow correction' can be used for the assessment of the fovea in BAF images.

101   In contrast to BAF imaging, green-light autofluorescence (GAF, excitation 518 nm)

102   cSLO imaging is not significantly affected by macular pigment due to a lack of

103   absorption.[15] Thus, GAF imaging would probably result in an even more precise

104   assessment of small, central changes including the differentiation between foveal

105   atrophy and foveal sparing. Only one previous study has compared BAF to GAF

106   imaging in GA.[14] However, this study did not compare combined BAF+NIR imaging

107   (as used in currently ongoing clinical trials) to GAF imaging.[14] Furthermore, no

108   manual constraints and no 'shadow correction' were used to exclude regions of

109   foveal sparing from the lesion area measured.[14]Recently, the lesion perimeter (lesion

110   circumference) and lesion circularity (the ratio of area to perimeter squared) have

111   been reported to be prognostic biomarkers for upcoming GA progression (Pfau M, et

112   al. IOVS 2016;57:ARVO E-Abstract 1613).[16] However, no data on the inter-reader

113   agreement of these biomarkers in GAF, BAF and NIR imaging have so far been

114   published.

115

116   The aim of this study was to systematically compare the inter-modality and inter-

117   reader agreement for cSLO GAF, cSLO BAF and cSLO BAF + cSLO NIR,

118   respectively, based on semi-automated delineation of GA in a reading center setting.

119   We tested the hypothesis that there are no differences in lesion size measurements

120   among the assessed modalities. Further, we hypothesized that GAF-based grading

121   exhibits the highest inter-reader agreement for the lesion area, perimeter and

122   circularity, since fewer manual constraints are necessary with regard to measuring

123   the foveal region.

124

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) study (NCT02051998 and NCT02332343, http://clinicaltrials.gov)

The inclusion and exclusion criteria for DSGA have been described previously.[17] 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 ° x 30 ° 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 a 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 – 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° x 30° with a resolution of 768 x 768 pixels and was centered on the fovea. Single BAF, NIR and GAF images were automatically aligned and averaged (up to 100 single frames) in order to maximize the signal-to-noise ratio using the manufacturer's software.

150

151   *Grading*

152   The readers (R1, R2, R3 and R4) were trained according to reading center standard

153   operating procedures (GRADE Reading Center, Bonn, Germany). Measurements of

154   atrophy areas were performed using the RegionFinder^TM software (Heidelberg

155   Engineering, version 2.6.3) as previously described. Briefly, the readers were asked

156   to set at least one seeding point inside of each atrophic region by selecting the pixel

157   with the lowest FAF signal (darkest grey value).[10] Thereafter, the readers had to

158   increase the growth power for each seeding point, which resulted in the inclusion of

159   adjacent pixels depending on the grey value, until the delineation just exceeded the

160   lesion boundaries.[10] Finally, the growth power had to be decreased by one increment

161   below this threshold.[10] The growth limit function was used if the segmentation

162   algorithm included the edges of the image frame. Further, retinal vessels or macula

163   pigment were excluded from the measured lesion area through the automated

164   'vessel detection' and 'shadow correction' or by placing manual constraints.[10] For

165   BAF+NIR-grading in foveal sparing, the readers were asked to delineate the spared

166   fovea in the NIR image semi-automatically prior to semi-automated quantification of

167   GA areas in the corresponding BAF image.[11] Each visit was graded by each reader

168   with (1.) BAF images only, (2.) BAF+NIR images and (3.) GAF images only. The

169   grading task was carried out on separate days and in random order. With the

170   currently available software version combined GAF+NIR-based grading was not

171   possible. The graded annotated images were transferred to ImageJ (Bethesda,

172   Maryland, USA) to measure the (cumulative) lesion circularity and (cumulative) lesion

173   perimeter using a custom-built plug-in (Pfau M, et al. IOVS 2016;57:ARVO E-

174   Abstract 1613).[16] Further, eyes were classified into foveal atrophy, extrafoveal

175 atrophy and foveal sparing according to the extent of GA in/near the fovea. Foveal

176 sparing was defined as an intact, residual foveal island being surrounded by more

177 than 270° of well-demarcated GA-areas .[11]

178

179 *Outcome measures and statistical analyses*

180 Statistical analyses were performed using the software environment R.[18] Area

181 measurements were square-root transformed to obtain normally distributed data. A

182 mixed-effects model considering imaging modality as fixed effect (GAF vs. BAF+NIR

183 vs. BAF) and visit as well as reader as random effects was used to assess whether

184 measured lesion size is dependent on the image modality. For each imaging modality

185 (GAF, BAF+NIR and BAF) the intraclass correlation coefficient (ICC, two-way

186 random, absolute agreement), the 95% coefficient of repeatability (CR) and the

187 coefficient of variation (CV) were determined.[19,20] Moreover, the ICC, CR and CV

188 were also determined for the perimeter and circularity measurements. For

189 visualization, Bland-Altman graphs were plotted. Spearman's rank correlation

190 coefficient (ρ) was calculated between the absolute differences and the mean values

191 to determine whether measurement variability increases with lesion size.[20]

192

193   **Results**

194   *Cohort characteristics*

195   A total of 40 visits of 40 eyes from 29 patients (age [mean ± SD] 79.7 ± 6.2 years, 20

196   female) with GA secondary to AMD were included and graded (Figure 1). Foveal

197   sparing was present in 22 out of 40 (55%) of these eyes. Out of the 40 eyes 31

198   (77%) were pseudophakic (Table 1, Figure 2).

199

200   *Lesion size in dependence of grading modality*

201   A mixed-effects model considering reader and visit as random effects disclosed that

202   the grading modality (GAF vs. BAF+NIR vs. BAF) significantly affected lesion size

203   measurements ($\chi 2(2)=55.257$, $p<0.001$). Hereby, the square-root lesion area was on

204   average 0.062 mm (95% CI 0.04 – 0.08 mm) larger for GAF- based measurements

205   than for BAF+NIR-based measurements. Similarly, GAF-based measurements were

206   on average larger by 0.077 mm (95% CI 0.06 – 0.10 mm) than BAF-based

207   measurements. There were no significant differences in the square-root lesion areas

208   between BAF- and BAF-IR- based measurements (0.01 mm; 95% CI -0.01 – 0.04

209   mm).

210

211   The differences between the measurements were plotted against their respective

212   mean value (Bland-Altman plots) for graphical analysis (Figure 3). To assess whether

213   measurement variability increased with lesion size, the Spearman's rank correlation

214   coefficient (ρ) for absolute differences and mean values was calculated. It indicated

215   for GAF- vs. BAF+NIR- ($\rho= 0.195$, $p=0.23$), for GAF- vs. BAF- ($\rho=-0.172$, $p=0.29$) and

216   for BAF+NIR- vs. BAF- ($\rho=-0.148$, $p=0.36$) based measurements that lesion size did

217   not significantly affect the inter-modality variability (Figure 3). In line with the mixed-

218   effects model, the mean differences of the Bland-Altman plots indicated that GAF-

219 based measurements were larger than BAF+NIR- (0.062 mm) or BAF- (0.077 mm)

220 based measurements (Figure 3).

221

222 Since this difference between GAF- and BAF+NIR- or BAF-based grading was

223 largely caused by 5 visits from 5 eyes (Figure 3), a detailed post hoc analysis of the

224 images was carried out. The eye with the greatest GAF-BAF discrepancy is shown in

225 Figure 4. The 84-year-old, pseudophakic, female patient presented with posterior

226 capsular opacification. The contrast of the lesion as compared to the background

227 signal was higher for the GAF than the BAF image. Especially the temporal lesion

228 boundary was better demarcated in the GAF image. Since readers were instructed to

229 increase the growth power of each seed until the defined area exceeded the lesion

230 boundaries, the measurements tended to be slightly larger for the GAF images. In

231 the BAF grading, the readers had to stop increasing the growth power prematurely

232 due to low contrast segments of the lesion boundary. Further, some foci of

233 questionably decreased autofluorescence (especially at the nasal margin of the

234 lesion) were only visible in the GAF image (Figure 4). Figure 5 shows another eye

235 with very low inter-modality agreement. Both, the GAF and BAF image did not allow

236 for an accurate delineation of the lesion, whereas the lesion boundaries were

237 clearcut in the NIR image. The assessment of foveal atrophy also resulted in some

238 inter-reader differences. As shown in Figure 2, foveal GA foci can have a similar

239 appearance compared to macular pigment in BAF images. This led to omission of

240 foveal GA foci in some BAF based measurements or to an incorrect grading taking

241 macular pigment for atrophy.

242

243 *Inter-reader agreement for lesion size measurements*

244 The CR (i.e. the value below which the difference between two measurements will lie

245 with a probability of 0.95) for the square-root area was 0.196 mm for the GAF-based

246 grading, 0.232 mm for the BAF+NIR-based grading and 0.263 mm for the BAF-based

247 grading. Likewise, the CV (2.87% for GAF, 3.49% for BAF+NIR, 3.98% for BAF) and

248 ICC (0.995 for GAF, 0.992 for BAF+NIR, 0.991 for BAF), which take into account the

249 underlying lesion size, indicated that GAF-based grading has the highest inter-reader

250 agreement (Table 2).

251 The highest inter-reader variability was observed for BAF-based measurements in

252 the subset of eyes with foveal sparing (CR of 0.274 mm) followed by the BAF-based

253 measurements in the subset of eyes without foveal sparing (CR of 0.263 mm). In

254 contrast, in BAF+NIR-based measurements (CR of 0.218 [foveal sparing] and 0.248

255 [non foveal sparing]) and GAF-based measurements (CR of 0.175 [foveal sparing]

256 and 0.211 [non foveal sparing]), the inter-reader variability was lower in eyes with

257 foveal sparing as compared to eyes without foveal sparing.

258

259 *Inter-reader agreement for lesion circularity and perimeter*

260 Despite equal lesion area measurements, the actual underlying delineations may

261 differ, since minor grading differences seem to balance out. Therefore, the lesion

262 perimeter (cumulative circumference) and lesion circularity were analyzed with

263 regard to inter-reader reliability, as these lesion shapedescriptive factors are more

264 susceptible to small differences of the actual underlying delineations. For the

265 perimeter, the GAF-based grading (CR=3.92 mm; CV=6.94%; ICC=0.983) exhibited

266 the highest inter-reader agreement followed by the BAF+NIR-based grading

267 (CR=5.04 mm; CV=9.03%; ICC=0.972) and BAF-based grading (CR=5.25 mm;

268 CV=9.69%; ICC=0.971). Likewise, GAF-based grading exhibited the best inter-reader

269 agreement for lesion circularity (Table 3).

270

271   Major sources of inter-reader disagreement regarding perimeter and circularity were

272   the extent of foveal involvement for BAF-based grading as exemplified in Figure 2.

273   The foveal involvement was difficult to assess using only BAF due to macular

274   pigment interference. Even with the shadow correction tool, which partially allowed

275   for assessment of foveal involvement, the delineation of the spared fovea differed

276   among readers because manual constraints had to be used. In contrast, BAF+NIR

277   imaging allowed for an accurate recognition of foveal sparing - however, (semi-

278   automated) constraints had to be used to delineate the boundary of the spared fovea.

279   In GAF-based grading, the least amount of constraints had to be used as illustrated

280   in Figure 1. Generally, the use of constraints for BAF, BAF+NIR and GAF

281   delineations, which was necessary for GA measurements in some eyes, appeared to

282   be associated with a lower inter-reader agreement. Thus, GAF-based measurements

283   relied mostly on the semi-automatically identified boundaries and were least

284   dependent on manual or semi-automated constraint placements resulting in the

285   highest inter-reader agreement.

286

287   **Discussion**

288   This study demonstrates that GAF and BAF+NIR imaging allow for an accurate and

289   reproducible quantification of GA lesions, and, therefore qualify as measurement

290   tools for clinical trials testing the efficacy of interventions aiming at a slowing down of

291   GA progression. Hereby, GAF based quantification exhibited the best inter-reader

292   agreement. BAF-based measurements also resulted in an excellent inter-reader

293   agreement. Yet, the inter-reader agreement was markedly lower than those obtained

294   from GAF or BAF+NIR imaging – especially when measuring circularity and

295   perimeter. Besides, GAF-based lesion size measurements tended to be minimally

296   larger than BAF+NIR- (or BAF-) based measurements.

297   To date only one study compared BAF vs. GAF imaging in GA.[14] However, this study

298   did not compare GAF-based grading to BAF+NIR-based grading, which serves as

299   primary outcome measure in currently ongoing phase II and III trials (NCT02247531,

300   NCT02247479, NCT02087085).[10,11,14] Further, the authors concluded that lesion

301   sizes in BAF images were larger than in GAF images because of centrally decreased

302   blue-light autofluorescence due to macular pigment.[14] However, it is conceivable that

303   the newer version of the RegionFinder™ software with 'shadow correction' and

304   manual constraints excludes more precisely regions of foveal sparing from the

305   measured lesion area.[14] Indeed, in our study, there was no relevant mean difference

306   between BAF+NIR- and BAF-based grading. The fact that GAF-based grading

307   resulted in minimally larger lesion size measurements than BAF- or BAF+NIR-based

308   grading is most likely attributable to the minimally sharper contrast at lesion

309   boundaries in a subset of GAF images. The readers were asked to set seeding

310   points inside of atrophic regions, then to increase the growth power until the

311   delineation exceeded the lesion boundaries and finally to decrease the growth power

312   by one increment below this threshold.[10] Sharper lesion boundaries allowed for a

313   greater increase of the growth power, while ill-defined lesion boundaries forced

314   readers to restrict the growth power prematurely and to use manual constraints.

315   The underlying reason for the higher contrast in GAF as compared to BAF images

316   could be partially attributed to the aging crystalline lens (23% of the included eyes

317   were phakic) reducing the transmission of short-wavelength light.[21] Other media

318   opacities including posterior capsular opacification and vitreous floaters appeared to

319   affect the quality of BAF images more severely than that of GAF images. In addition,

320   GAF images were more often in perfect focus than BAF images. Usually, the focus is

321   initially adjusted in the NIR mode (for patient comfort) and then quickly re-adjusted

322   for chromatic aberration after switching to the GAF or BAF mode. Since the optimal

323   focus for GAF is closer to the focus of NIR, re-adjustment is easier for GAF imaging.

324   Finally, patients tend to blink less (patient comfort) during GAF than BAF imaging,

325   which facilitates the acquisition of high-quality images.

326   BAF-based grading in eyes with foveal sparing exhibited the highest inter-reader

327   variability in this study. BAF+NIR-based grading, which allowed for a semi-automatic

328   delineation of the residual foveal island in NIR-images, resulted in a better inter-

329   reader agreement as compared to BAF-based grading underscoring the importance

330   of semi-automation of the grading process.[11] The highest inter-reader agreement was

331   observed for GAF images that required the least constraints highlighting the

332   importance of semi-automatic versus manual delineation for the inter-reader

333   agreement. Especially in clinical trials, it is crucial to maximize the inter-reader

334   agreement, since effect sizes are dependent on the underlying measurement

335   variability.[20] Thus, inter-reader agreement affects directly sample size determination.

336 Further studies will be needed to compare GAF imaging to other image modalities.

337 The recently published *Classification of Atrophy Meeting [CAM]*–consensus

338 recommended the use of color fundus photography (CFP), BAF, NIR and optical

339 coherence tomography (OCT) in studies with GA.[22] The inter-reader agreement of

340 these modalities has been assessed previously and reported to be high in a number

341 of studies (ICC values ranging from 0.95 [for CFP] to 0.99 [for BAF and OCT]).[10,23–28]

342 However, the ICC was commonly reported as only outcome measure. It is difficult to

343 compare the ICC across different study cohorts, since it is dependent on the variance

344 of the trait (i.e. lesion size) within the cohort.[19] Addition of a small number of eyes

345 with either very large or very small GA lesions would result in markedly improved ICC

346 values irrespective of the underlying image modality or grading method. The CR as

347 recommended by Bland and Altman was used in our study because it is independent

348 of the average lesion size and may be compared to different study cohorts in a more

349 meaningful manner.[20] With the advent of faster spectral domain and swept source

350 OCT devices, OCT imaging appears to be a potential alternative to BAF or GAF

351 imaging in the setting of GA.[26] Typically, OCT-based segmentation methods rely on

352 *en face* fundus or sub-RPE projection images that depict so-called hyper-

353 transmission into the choroid in regions of GA.[13,26,29] Hereby, large choroidal vessels

354 are typically hyposcattering and may result in segmentation artifacts.[13] Thus, future

355 studies should evaluate automated OCT-based segmentation in comparison to BAF

356 or GAF images as the latter usually depict a higher contrast than OCT images.[13]

357 Noteworthy, one previous study reported that the agreement between automatically-

358 and manually-defined GA regions was better for BAF than OCT images.[13]

359 Finally, GAF imaging tended to be more comfortable for patients than BAF imaging

360 (anecdotal evidence). Furthermore, in *ABCA4*-associated retinopathy, there is some

361 controversy with regard to BAF imaging as it was speculated that it may accelerate

362 accumulation of A2E and, thus, induce apoptosis in RPE cells with particularly high

363 levels of A2E as observed in *Abcr*-knockout mice and cell culture, respectively.[30–34]

364 Although, up to date there is no evidence of phototoxic effects in humans, both, in

365 absence or presence of retinal diseases, Cideciyan and associates have proposed to

366 apply reduced-illuminance BAF imaging to reduce light exposure, and, thus, to

367 reduce a potential risk for adverse effects.[33] Noteworthy, in cultured human RPE cells

368 with internalized A2E, illumination with green-light was shown to result in

369 substantially fewer non-viable cells as compared to illumination with blue-light.[31]

370 Based on these results obtained by ex-vivo analysis and in animal models, it may be

371 speculated that GAF imaging might be safer in patients with RPE atrophy (especially

372 in association with *ABCA4*-associated retinopathy).

373 Limitations of this study must be considered. First, the current version of the

374 RegionFinder™ software does not allow for combined GAF+NIR grading. Potentially,

375 combined GAF+NIR grading would further increase the inter-reader reliability in a

376 subset of patients (cf., Figure 5). Second, the grading of the three modalities was

377 performed by all readers in random order and on separate days. However, it cannot

378 be fully excluded that readers re-recognized eyes. Third, OCT imaging which is the

379 most promising image modality besides BAF and GAF imaging in GA was not

380 included in this study.

381 In summary, this study demonstrated that GAF and combined BAF+NIR imaging

382 allow for reliable assessment of lesion size and shape in GA secondary to AMD, both

383 clinically and particularly in clinical studies. Hereby, GAF based quantification

384 exhibits higher inter-reader agreement. Since media-opacification appears to

385 interfere with lesion-demarcation more strongly in BAF than in GAF, minor

386   differences in lesion size measurements between the different analysis approaches

387   must be considered.

References

1. Schmitz-Valckenberg S, Sadda S, Staurenghi G, et al. Geographic Atrophy. *Retina*. 2016;36(12):2250-2264. doi:10.1097/IAE.0000000000001258.

2. Cheung LK, Eaton A. Age-Related Macular Degeneration. *Pharmacother J Hum Pharmacol Drug Ther*. 2013;33(8):838-855. doi:10.1002/phar.1264.

3. Holz FG, Schmitz-Valckenberg S, Fleckenstein M. Recent developments in the treatment of age-related macular degeneration. *J Clin Invest*. 2014;124(4):1430-1438. doi:10.1172/JCI71029.

4. Rasmussen A, Sander B. Long-term longitudinal study of patients treated with ranibizumab for neovascular age-related macular degeneration. *Curr Opin Ophthalmol*. 2014;25(3):158-163. doi:10.1016/j.ophtha.2013.05.018.

5. von Ruckmann A, Fitzke FW, Bird AC. In vivo fundus autofluorescence in macular dystrophies. *Arch Ophthalmol*. 1997;115(5):609-615.

6. Holz FG, Bellmann C, Margaritidis M, Schütt F, Otto TP, Völcker HE. Patterns of increased in vivo fundus autofluorescence in the junctional zone of geographic atrophy of the retinal pigment epithelium associated with age-related macular degeneration. *Graefe's Arch Clin Exp Ophthalmol*. 1999;237(2):145-152. doi:10.1007/s004170050209.

7. Dreyhaupt J, Mansmann U, Pritsch M, Dolar-Szczasny J, Bindewald a, Holz FG. Modelling the natural history of geographic atrophy in patients with age-related macular degeneration. *Ophthalmic Epidemiol*. 2005;12(6):353-362. doi:10.1080/09286580591005723.

8. Holz FG, Bindewald-Wittich A, Fleckenstein M, Dreyhaupt J, Scholl HPN, Schmitz-Valckenberg S. Progression of Geographic Atrophy and Impact of Fundus Autofluorescence Patterns in Age-related Macular Degeneration. *Am J Ophthalmol*. 2007;143(3):463-472. doi:10.1016/j.ajo.2006.11.041.

418   9.   Deckert A, Schmitz-Valckenberg S, Jorzik J, Bindewald A, Holz FG, Mansmann

419   U. Automated analysis of digital fundus autofluorescence images of geographic

420   atrophy in advanced age-related macular degeneration using confocal

421   scanning laser ophthalmoscopy (cSLO). *BMC Ophthalmol*. 2005;5(1):8.

422   doi:10.1186/1471-2415-5-8.

423   10.   Schmitz-Valckenberg S, Brinkmann CK, Alten F, et al. Semiautomated image

424   processing method for identification and quantification of geographic atrophy in

425   age-related macular degeneration. *Invest Ophthalmol Vis Sci*.

426   2011;52(10):7640-7646. doi:10.1167/iovs.11-7457.

427   11.   Lindner M, Böker A, Mauschitz MM, et al. Directional Kinetics of Geographic

428   Atrophy Progression in Age-Related Macular Degeneration with Foveal

429   Sparing. *Ophthalmology*. 2015;122(7):1356-1365.

430   doi:10.1016/j.ophtha.2015.03.027.

431   12.   Lee N, Laine A, Smith R. A hybrid segmentation approach for geographic

432   atrophy in fundus auto-fluorescence images for diagnosis of age-related

433   macular degeneration. In: *Conf Proc IEEE Eng Med Biol Soc*. IEEE;

434   2007:4965-4968. doi:10.1109/IEMBS.2007.4353455.

435   13.   Hu Z, Medioni GG, Hernandez M, Hariri A, Wu X, Sadda SR. Segmentation of

436   the geographic atrophy in spectral-domain optical coherence tomography and

437   fundus autofluorescence images. *Invest Ophthalmol Vis Sci*.

438   2013;54(13):8375-8383.

439   14.   Wolf-Schnurrbusch UE, Wittwer V V, Ghanem R, et al. Blue Light versus Green

440   Light Autofluorescence: Lesion Size of Areas with Geographic Atrophy. *Invest

441   Ophthalmol Vis Sci*. 2011;52(13):iovs.11-8346-. doi:10.1167/iovs.11-8346.

442   15.   Delori FC, Goger DG, Hammond BR, Snodderly DM, Burns SA. Macular

443   pigment density measured by autofluorescence spectrometry : comparison with

444   reflectometry and heterochromatic flicker photometry. *J Opt Soc Am A*.

445   2001;18(6):1212-1230. doi:10.1364/JOSAA.18.001212.

446   16.   Domalpally A, Danis RP, White J, et al. Circularity index as a risk factor for

447   progression of geographic atrophy. *Ophthalmology*. 2013;120(12):2666-2671.

448   doi:10.1016/j.ophtha.2013.07.047.

449   17.   Lindner M, Bezatis A, Czauderna J, et al. Choroidal Thickness in Geographic

450   Atrophy Secondary to Age-Related Macular Degeneration. *Invest Ophthalmol

451   Vis Sci*. 2015;56(2):875-882. doi:10.1167/iovs.14-14933.

452   18.   R Development Core Team. R: A language and environment for statistical

453   computing. R Foundation for Statistical Computing. 2012. doi:10.1007/978-3-

454   540-74686-7.

455   19.   Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability.

456   *Psychol Bull*. 1979;86(2):420-428. doi:10.1037/0033-2909.86.2.420.

457   20.   G. AD. Statistical methods for assessing agreement between two methods of

458   clinical measurement. *Lancet*. 1986;68(8476):307. doi:10.1016/S0140-

459   6736(86)90837-8.

460   21.   Boettner EA, Wolter JR. Transmission of the Ocular Media. *Invest Ophthalmol

461   Vis Sci*. 1962;1(6):776-783. doi:10.1017/S0025315406014421.

462   22.   Holz FG, Sadda SR, Staurenghi G, et al. Imaging Protocols in Clinical Studies

463   in Advanced Age-Related Macular Degeneration: Recommendations from

464   Classification of Atrophy Consensus Meetings. *Ophthalmology*. 2017.

465   23.   Lujan BJ, Rosenfeld PJ, Gregori G, et al. Spectral domain optical coherence

466   tomographic imaging of geographic atrophy. *Ophthalmic Surg Lasers Imaging*.

467   2009;40(2):96-101. doi:10.3928/15428877-20090301-16.

468   24.   Fleckenstein M, Schmitz-Valckenberg S, Adrion C, et al. Tracking progression

469   with spectral-domain optical coherence tomography in geographic atrophy

470   caused by age-related macular degeneration. *Invest Ophthalmol Vis Sci.*

471   2010;51(8):3846-3852. doi:10.1167/iovs.09-4533.

472 25. Schmitz-Valckenberg S, Fleckenstein M, Göbel AP, Hohman TC, Holz FG.

473   Optical coherence tomography and autofluorescence findings in areas with

474   geographic atrophy due to age-related macular degeneration. *Investig*

475   *Ophthalmol Vis Sci.* 2011;52(1):1-6. doi:10.1167/iovs.10-5619.

476 26. Yehoshua Z, Garcia Filho CA, Penha FM, et al. Comparison of geographic

477   atrophy measurements from the OCT fundus image and the sub-RPE slab

478   image. *Ophthalmic Surg Lasers Imaging Retin.* 2013;44(2):127-132.

479   doi:10.3928/23258160-20130313-05.

480 27. Moussa N Ben, Georges A, Capuano V, Merle B, Souied EH, Querques G.

481   MultiColor imaging in the evaluation of geographic atrophy due to age-related

482   macular degeneration. *Br J Ophthalmol.* 2015;99(6):842-847.

483   doi:10.1136/bjophthalmol-2014-305643.

484 28. Pilotto E, Guidolin F, Convento E, et al. En face optical coherence tomography

485   to detect and measure geographic atrophy. *Investig Ophthalmol Vis Sci.*

486   2015;56(13):8120-8124. doi:10.1167/iovs.15-17366.

487 29. Niu S, de Sisternes L, Chen Q, Leng T, Rubin DL. Automated geographic

488   atrophy segmentation for SD-OCT images using region-based C-V model via

489   local similarity factor. *Biomed Opt Express.* 2016;7(2):581.

490   doi:10.1364/BOE.7.000581.

491 30. Weng J, Mata NL, Azarian SM, Tzekov RT, Birch DG, Travis GH. Insights into

492   the function of rim protein in photoreceptors and etiology of Stargardt's disease

493   from the phenotype in abcr knockout mice. *Cell.* 1999;98(1):13-23.

494   doi:10.1016/S0092-8674(00)80602-9.

495 31. Sparrow JR, Nakanishi K, Parish CA. The Lipofuscin Fluorophore A2E

496   Mediates Blue Light– Induced Damage to Retinal Pigmented Epithelial Cells.

497   *Invest Ophthalmol Vis Sci*. 1981;41(7):1981-1989.

498   32.   Radu RA, Mata NL, Nusinowitz S, Liu X, Sieving PA, Travis GH. Treatment

499   with isotretinoin inhibits lipofuscin accumulation in a mouse model of recessive

500   Stargardt's macular degeneration. *Proc Natl Acad Sci U S A*.

501   2003;100(8):4742-4747. doi:10.1073/pnas.0737855100.

502   33.   Cideciyan A V, Swider M, Aleman TS, et al. Reduced-illuminance

503   autofluorescence imaging in ABCA4-associated retinal degenerations. *J Opt*

504   *Soc Am A Opt Image Sci Vis*. 2007;24(5):1457-1467.

505   doi:10.1364/JOSAA.24.001457.

506   34.   Teussink MM, Lambertus S, de Mul FF, et al. Lipofuscin-associated photo-

507   oxidative stress during fundus autofluorescence imaging. *PLoS One*.

508   2017;12(2):e0172635. doi:10.1371/journal.pone.0172635.

509   35.   Spaide RF. Age-related choroidal atrophy. *Am J Ophthalmol*. 2009;147(5):801-

510   810. doi:10.1016/j.ajo.2008.12.010.

511

512

513 **Figures**

514 **Figure 1. Exemplary grading report**

515 The exemplary annotated images were based on blue-light autofluorescence (**A**) and

516 green-light autofluorescence (**B**) images of a 71-year-old, pseudophakic male

517 patient. While no manual constraints had to be used for the delineation of the GA

518 lesion in the green-light autofluorescence (**B**) image, the reader had to use manual

519 constraints (red lines) towards the fovea in the blue-light autofluorescence (**A**) image

520 due to macular pigment interference.

521

522

523 **Figure 2. Atrophy border in close proximity to the fovea**

524 The upper row (**A, B, C**) shows the blue-light autofluorescence (BAF), the near

525 infrared reflectance (NIR) and the green-light autofluorescence (GAF) images of an

526 81-year-old, female patient with cataract. The contrast of the GA lesion against the

527 background is higher for the GAF image as compared to the BAF image. The lower

528 row (**D, E, F**) shows the images of a 79-year-old, pseudophakic, male patient. Overall

529 the contrast of the GA lesion against the background is similar for the BAF and GAF

530 images. However, based on the BAF image, it is challenging to determine whether

531 the central spot with decreased autofluorescence represents atrophy or macular

532 pigment (**D**, green arrow). The NIR and GAF images facilitate accurate grading and

533 demonstrate that the spot indeed represents atrophy.

534

535

Figure 3. Bland-Altmann graphs for the inter-modality agreement

The Bland-Altmann graphs show the measurement differences for the square-root lesion area of two modalities (y-axis) against their mean (x-axis). The solid line indicates the mean difference and the dashed lines indicate the 95% limits of agreement. Green-light autofluorescence- (GAF) based grading resulted on average in larger measurements as compared to combined blue-light autofluorescence with near infrared reflectance- (BAF+NIR) based grading and/or BAF-based grading (**A, B**) as indicated by the mean differences of -0.062 mm (BAF+NIR vs. GAF) and -0.077 mm (BAF vs. GAF). BAF- compared to combined BAF+NIR- based grading exhibited no relevant mean difference (**C**).

546 **Figure 4. Green- versus blue-light autofluorescence**

547 The blue-light autofluorescence (BAF) and the green-light autofluorescence (GAF)

548 images of this 84-year-old, pseudophakic, female patient with posterior capsular

549 opacification exhibited the greatest discrepancies. At location 1 (green box 1), the

550 GAF images show atrophy while no distinct patch of decreased autofluorescence can

551 seen in the BAF image (**D1**). Overall, the contrast at the temporal lesion boundary is

552 markedly higher in the GAF than the BAF image and the near infrared reflectance

553 (NIR) image. Towards the optic disc, multiple foci of questionably decreased

554 autofluorescence are seen in the GAF image. These are not visible in the BAF image

555 (**D2**).

556

557 **Figure 5. Near infrared reflectance versus autofluorescence imaging**

558 The green-light (GAF) and blue-light (BAF) autofluorescence images of this 92-year-

559 old, female patient with cataract are markedly different when compared to the near

560 infrared reflectance (NIR) image. While the NIR image corresponds to the area of

561 hyper-transmission in optical coherence tomography (**D**), the GAF and BAF images

562 depict larger areas of decreased autofluorescence. The eye exhibited multiple

563 features of age-related macular degeneration (AMD) including hyper-pigmentary

564 changes, reticular pseudodrusen (subretinal drusenoid deposits) and soft drusen.

565 However, the marked tessellated fundus appearance (in conjunction with pronounced

566 choroidal thinning and obliteration of the choroid in proximity to the β-zone of the

567 peripapillary atrophy) could be indicative of so-called age-related choroidal atrophy.[35]

568 It could be argued that the eye should have been excluded from the analysis.

569 However, eyes with GA lesions that are difficult to segment also be considered for

570 real-world clinical trials.

571 **Figure 6. Bland-Altmann graphs for inter-reader agreement**

572 The Bland-Altmann graphs show the measurement differences for the square-root

573 lesion area of two readers (y-axis) against their mean (x-axis). The solid line

574 indicates the mean difference and the dashed lines indicate the 95% limits of

575 agreement. The rows show the pairs of readers. The columns show the image

576 modality (blue-light autofluorescence [BAF], near infrared reflectance [NIR], green-

577 light autofluorescence [GAF]). There were no relevant systematic mean differences

578 between the readers. Please note, the inter-reader variability was lowest for all pairs

579 of readers of GAF-based measurements. Further, the measurement variability did not

580 depend on the measurement value.

581 **Tables**

582

583 **Table 1.** Demographic data of all patients

Mean age ± SD (range) in years	79.73 ± 6.18 (67.7 – 92.2)
Sex, n (%)	
Male	9 (31)
Female	20 (69)
Lens status, n (%)	
Phakic	9 (23)
Pseudophakic	31 (77)
Mean GA area ± SD in mm^2	6.81 ± 5.13
Foveal involvement, n (%)	
Foveal atrophy	6 (15)
Foveal sparing*	22 (55)
Extrafoveal atrophy	12 (30)

584

585 * Foveal sparing was defined as an intact, residual foveal island being surrounded by

586 more than 270° of well-demarcated GA-areas.[11]

587

588   **Table 2.** Inter-reader agreement for square-root lesion area

	Coefficient of repeatability (in mm)	Coefficient of variation (in %)	ICC
GAF	0.196	2.87	0.995 (0.99 – 0.997)
BAF+NIR	0.232	3.49	0.992 (0.988 – 0.996)
BAF	0.263	3.98	0.991 (0.985 – 0.995)

589

590   **Table 3.** Inter-reader agreement for lesion circularity and lesion perimeter

	Grading modality	Coefficient of repeatability	Coefficient of variation (in %)	ICC
Circularity	GAF	0.035	21	0.9 (0.843 – 0.941)
	BAF+NIR	0.040	24	0.87 (0.801 – 0.922)
	BAF	0.044	24.9	0.85 (0.772 – 0.909)
Perimeter	GAF	3.92 mm	6.94	0.983 (0.971 – 0.99)
	BAF+NIR	5.04 mm	9.03	0.972 (0.955 – 0.984)
	BAF	5.25 mm	9.69	0.971 (0.951 – 0.984)

591