Determinants of cone- and rod-function in geographic atrophy: AI-based structure-function correlation

Maximilian Pfau,[1,2,3] Leon von der Emde,[1] Chantal Dysli,[1,4] Philipp T. Möller,[1,2] Sarah Thiele,[1,2] Moritz Lindner,[5,6] Matthias Schmid,[7] Daniel L. Rubin,[3] Monika Fleckenstein,[1,2,8] Frank G. Holz,[1,2] Steffen Schmitz-Valckenberg,[1,2,8]

1. Department of Ophthalmology, University of Bonn, Ernst-Abbe-Str. 2, Bonn, Germany
2. GRADE Reading Center, Ernst-Abbe-Str. 2, Bonn, Germany
3. Department of Biomedical Data Science, Stanford University, Stanford, California, USA
4. Department of Ophthalmology and Department of Clinical Research, Inselspital, Bern University Hospital and University of Bern, Bern, Switzerland
5. The Nuffield Laboratory of Ophthalmology, Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom
6. Institute of Physiology and Pathophysiology, Department of Neurophysiology, Philipps University, Deutschhausstrasse 1-2, Marburg, Germany 35037
7. Institute for Medical Biometry, Informatics and Epidemiology, Faculty of Medicine, University of Bonn, Germany
8. John A. Moran Eye Center, University of Utah, Salt Lake City, UT, USA

Running head: *AI-based structure-function correlation in GA*

Key words: *Age-related macular degeneration; geographic atrophy; microperimetry; scotopic microperimetry; random forest regression; machine learning*

Number of words: *4.304*

Number of figures: *5*

Supplemental figures: *3*

Number of tables: *2*

Correspondence:
Steffen Schmitz-Valckenberg, MD
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: Steffen.schmitz-valckenberg@ukbonn.de

Supplemental Material available at AJO.com

Introduction

Age-related macular degeneration (AMD) constitutes the leading cause of legal blindness among the elderly in industrialized countries.[1] In recent years, remarkable progress has been made for the treatment of macular neovascularization in the context of the exudative late-stage form of the disease.[2] Despite of the identification of multiple environmental and genetic risk factors, no therapy could be established for the underlying degenerative disease itself, i.e. intermediate AMD or the non-exudative late-stage manifestation geographic atrophy (GA).[3]

The selection of appropriate functional outcome measures in AMD is a challenge. Notably, change in best-corrected visual acuity (BCVA) does not adequality capture GA progression, given that GA commonly develops in the parafoveal macula and often spares the fovea initially.[4,5] Despite of good BCVA, patients may suffer from severe reading difficulties due to parafoveal absolute scotomata.[6] Spatially-resolved mapping of macular light sensitivity using mesopic fundus-controlled perimetry (also termed 'microperimetry') has been proposed as a better alternative for the assessment of visual function loss in patients with GA.[7,8] This has been further extended to dark-adapted fundus-controlled perimetry in the context of GA, given that multiple histopathological studies have shown that rod photoreceptor loss typically exceeds cone photoreceptors loss in AMD in general as well as in the junctional-zone of GA.[8,9] However, the test accuracy of fundus-controlled perimetry is limited by the tradeoff between test time (i.e., fatigue), the spatial resolution and spatial extent (i.e., size and number of test-points) as well as the point-wise test accuracy (i.e., threshold strategy). For patients with GA, two optimized strategies to monitor disease progression have been recently proposed: (i.) limiting the spatial extent to relevant test-points by incorporating structural data (patient-tailored perimetry patterns),[8] or (ii.) choosing a fast, qualitative threshold-strategy termed defect-mapping approach.[10] Yet, both of these optimized strategies lead to information loss in terms of the overall covered area or the accuracy of the individual threshold determinations.[8,10]

Artificial intelligence (AI) based prediction of mesopic and dark-adapted retinal sensitivity based on structural imaging (termed "inferred sensitivity" herein) may provide a viable alternative to quantify functional loss in patients with GA. This approach has been previously proposed in the context of macular telangiectasia type 2,[11] Leber congenital amaurosis,[12] and neovascular AMD.[13] A key advantage of this approach would be that structural imaging methods such as infrared reflection (IR) and fundus autofluorescence (FAF) imaging as well as spectral-domain optical coherence tomography imaging (SD-OCT), which would constitute the basis for the inferred sensitivity mapping, provide a much higher *en face* resolution than fundus-controlled perimetry testing (5.7 µm/pixel [e.g. Heidelberg Spectralis OCT 2 device] vs. a 128 µm Goldmann III stimulus).[11,13] However, prior to the application of "inferred-sensitivity" as a quasi-functional surrogate outcome, the prediction accuracy and the relative importance of imaging features must be studied in a disease-specific context.

Accordingly, we aimed to identify the accuracy of AI-based predictions of retinal sensitivity based on retinal microstructure in the presence of GA secondary to AMD. This analysis was based on mesopic and dark-adapted cyan and red fundus-controlled perimetry and multimodal imaging (IR, FAF and SD-OCT). Further, we examined the importance of the included imaging features for the sensitivity predictions as well as the potential influence of behavioral factors ("patient-reliability indices"). Last, we aimed to identify regions with dissociation of rod- and cone-dysfunction.

Methods

Patients and subjects with normal eyes

Patients were recruited in the setting of the non-interventional, prospective natural history study *Directional Spread in Geographic Atrophy* (DSGA, ClinicalTrials.gov: NCT02051998, PI: M. Fleckenstein) at the Department of Ophthalmology, University of Bonn, Germany.[8] For inclusion, patients needed to be older than 55 years at the time of inclusion and had to exhibit GA secondary to AMD in at least one eye.[8] The diagnosis of AMD was hereby based on the

presence of soft drusen and hyperpigmentary changes, and a compatible FAF phenotype, while excluding FAF phenotypes of mimicking diseases such as late-onset Stargardt disease.[14] Further exclusion criteria were any history of retinal surgery, laser photocoagulation, and radiation therapy or other retinal diseases in the study eye such as diabetic retinopathy. Only one eye per patient was included. Functional data for a subgroup of the here reported patients has been previously published.[8] Subjects with normal eyes (to obtain reference values) were recruited from the hospital wards among patients with a normal fellow eye as well as among companions. Both studies followed the Declaration of Helsinki and were approved by the institutional review board (Ethikkommission, Medizinischen Fakultät, Rheinische Friedrich-Wilhelms-Universität Bonn, Germany; approval ID: 197/12 and 191/16). Written informed consent was obtained from all participants prior to the study examinations.

Imaging

All patients and all subjects underwent (following pupil dilatation with 0.5% tropicamide and 2.5% phenylephrine) 30° x 30° fundus autofluorescence imaging (λ excitation 488 nm, λ emission 500 - 700 nm), 30° x 30° infrared reflectance (λ 815 nm) imaging, and 30° x 25° SD-OCT imaging (121 B-scans, ART 25) using a Spectralis HRA+OCT2 (Heidelberg Engineering, Germany).

Psychophysical testing in eyes with GA

Mesopic, dark-adapted cyan and red sensitivity were measured using fundus-controlled perimetry with the S-MAIA device (CenterVue S.p.A., Italy).[15,16] Test-points were placed along iso-hulls (i.e., contour lines) at even distances surrounding the GA boundary using ad-hoc software for the generation of the test patterns.[8] The distances to the GA boundary were 0.43°, 0.86°, 1.29°, 2.15°, and 3.01° to the GA boundary as well as at -0.645° (i.e., within GA). These patient-tailored perimetry patterns maximize the number of test-points in disease relevant regions (i.e., junctional-zone of GA), while limiting the overall number of test-points.[8] The time for dark-adaptation was at least 35 minutes for all patients. The false-positive responses (measured through presentation of suprathreshold stimuli to the optic nerve head [similar to Heijl-Krakau method in standard automated perimetry, but with fundus tracking]), the rate of wrong pressure events outside of the response window of the S-MAIA device as well as the exam duration were considered as “patient-reliability indices”. The 95% bivariate contour ellipse area (BCEA) encompassing 95% of the fixation points was recorded as measure of fixation stability.

Psychophysical testing in normal eyes

For normal eyes, the perimetry grid consisted for all three types of testing (mesopic, DA cyan, DA red) of five rings at the eccentricities of 1°, 3°, 5°, 7°, and 9° with 12 stimuli each, plus additional stimuli along the horizontal meridian at 0°, ±2°, ±4°, ±6°, ±8°, and ±10° (i.e., total of 71 stimuli).

SD-OCT segmentation and image registration

Segmentation of retinal layers was performed using the Heidelberg Eye Explorer (version 1.10.2.0, Heidelberg Engineering, Germany). First, the automated segmentation was applied, followed by two successive rounds of manual correction by two different readers (LvdE, MP). Hereby, the second reader acted as senior reader. The following layer boundary definitions were applied (Figure 1): full retina (internal limiting membrane [ILM] to Bruch's membrane [BrM]); inner retina (ILM to the outer plexiform layer (OPL) / outer nuclear layer (ONL) boundary); outer nuclear layer including Henle's fiber layer (ONL/OPL boundary to the external limiting membrane [ELM]), inner segments (IS, ELM to ellipsoid zone [EZ]), outer segments (OS, EZ to interdigitation zone) and retinal pigment epithelium drusen complex (RPEDC, encompassing subretinal drusenoid deposits [reticular pseudodrusen], the retinal pigment epithelium and drusen).

For all layers, the thickness, as well as the minimum, mean, and maximum intensity projections along the voxels of each A-scan were exported as *en face* images using custom MATLAB (MathWorks, USA) software. All intensity values (FAF, IR and SD-OCT reflectivity data) were scaled to a range of 0 to 1. Thickness maps were not transformed (i.e., each pixel value indicated the thickness in µm). The imaging features (Figure 1, cf. Supplemental Table S1) were then extracted for a circular region-of-interest corresponding to the area of a Goldmann III stimulus (0.43° diameter) using custom-made plugins for the Fiji distribution of ImageJ. Supplemental Figure S1 (Supplemental Material at AJO.com) and previous manuscripts outline these steps in greater detail.[13,17]

Standardization of patient data

Data from healthy eyes were used to standardize all of the patient data (thickness and intensity maps) prior to the prediction analyses in a point-wise (i.e. pixel-wise) manner. The normal data were not used as part of the modeling process itself (i.e., the model accuracy estimates reflect the prediction accuracy for patient data). Point-wise linear regression analysis of data from normal subjects with age as dependent variable was used to obtain normative mean and standard deviation estimates for the retinal sensitivity, layer-thickness, and layer-reflectivity values for a 76-year-old subject (in consideration of the mean patient age). Following spatial-interpolation (as previously described in detail [cf. Supplement of previous manuscript[8]]), these values were then used to standardize the patient data. Imaging features of patients were standardized by computing the z-scores (i.e. the number of normative standard deviations by which a given observation deviates from the normative mean estimate). The sensitivity measurements of patients were standardized by calculating the point-wise sensitivity loss as compared to the spatially corresponding normative value (negative values indicating sensitivity loss). If two sensitivity measurements were available for a given patient (test and retest), then the results of the test and retest were averaged in a point-wise manner. This standardization of patient data was used to ensure that the feature importance values below (cf. Figure 3) reflect disease-specific relationships rather than the generic associations between the topography of retinal sensitivity and disease-unspecific features, which are highly indicative of the retinal topography (i.e., inner retina thickness).

Predictive modeling and statistical analyses

Statistical analyses were performed using the software environment R and the add-on packages lme4, randomForest, caret and forestFloor.[18–21] Visual acuity measurements were converted to the logarithm of the Minimum Angle of Resolution (logMAR).

All features used for the modeling are described in detail in Supplemental Table S1 (Supplemental Material at AJO.com). For the predictive models (random forest models, 1000 trees), we applied nested resampling. Outer resampling was used to estimate the model accuracy and nested inner resampling to tune the random forest parameter *mtry*, which defines the number of variables available for splitting at each tree node.[22] For the first scenario (S1), patient-wise leave-one-out cross-validation (LOO-CV) was applied as outer resampling to obtain an estimate of the model accuracy for an “unknown patient” without prior testing. The parameter *mtry* was tuned using nested inner resampling (5-fold CV with patient-wise splits).

For the second scenario (S2), half of the patient-specific data (i.e., data from every other test-point along each iso-hull) were added to the training data, and only the other half of the patient-specific data served as test data to obtain an estimate of the model accuracy. This reflects a prediction scenario for a patient with “prior perimetry testing data available”. Additionally, the eye ID was added by one-hot encoding as predictor for S2. The parameter *mtry* was tuned using nested inner resampling (5-fold CV).

For both scenarios, three types of feature sets (A, B, C) were evaluated: (S1A [= Scenario 1 – feature set A] and S2A) imaging features only (26 candidate variables), (S1B and S2B) imaging features and “patient reliability indices” (30 candidate variables) and (S1C and S2C) imaging

features, "patient reliability indices" and fixation stability (31 candidate variables, cf. Supplemental Table S1 for detailed outline).

As reference, a null model for S1 was fitted using LOO-CV, representing the mean sensitivity of n-1 patients (training fold) and was compared to the one left out patient. For S2, the null model was based on the patient-specific mean of half of the test-points (every other test-point along each iso-hull), validated against the other half.

The mean absolute error (MAE), defined as the average of the absolute errors between the predictions and psychophysically measured values, served as measure of prediction accuracy. It was compared across the scenarios and feature sets using linear mixed-effects models. The linear mixed-effects models considered the MAE as dependent variable and patient as random effect. A likelihood ratio test was applied to test, whether the scenario and feature set type (i.e., S1A, S1B […] S2C) had an influence as fixed effect on the prediction accuracy followed by post-hoc pair-wise comparisons (Tukey contrasts). The root-mean-square error (RMSE) estimates for all models are also provided for reference (Supplemental Table S2, Supplemental Material at AJO.com).

The permutation importance (increase in mean squared error [% Inc MSE]) was evaluated as measure of feature importance.[22] Plots of the feature contributions were used to visualize the relationship between selected imaging features and the "inferred sensitivity".[21]

Results

Cohort characteristics

Forty-one eyes of 41 patients with a mean age (± SD) of 75.8 ± 8.4 years and with a mean area of GA of 6.0 ± 4.8 mm^2 were included in this study (Table 1). The patient-tailored perimetry patterns encompassed on average (mean ± SD) 68.0 ± 19.2 test-points. Accordingly, a total of 2789 test-points were available for predictive modeling for mesopic testing, and 2785 test-points for dark-adapted cyan and red testing. The missing test-points for the two types of dark-adapted testing are attributable to off-center fixation of the patients during the respective tests, and the limited imaging/testing frame of the S-MAIA device. For a subset of 37 patients, intra-session test-retest data were available (Table 1).

To standardize the data of the patients, data of 40 eyes of 40 controls (55.8 ± 17.4 years [21.8–82.1 years]) with a mean BCVA of 0.03 ± 0.07 [Snellen equivalent approximately 20/20] were included (cf. Methods).

Prediction accuracies of "inferred sensitivity" in an unknown patient (scenario 1)

First, patient-wise leave-one-out cross-validation (LOO-CV) was applied to obtain an estimate for the prediction accuracies of "inferred sensitivity" in an unknown patient (scenario 1). For mesopic sensitivity, the predictions for scenario 1 based on imaging data only (S1A) exhibited a mean absolute error (MAE [95% CI]) of 4.64 dB [4.02, 5.25] (Figure 2). Hereby, neither addition of "patient reliability indices" to the feature set (S1B) nor the addition of the fixation stability (S1C) resulted in an improved prediction accuracy (Table 2 and Supplemental Figure S2, Supplemental Material at AJO.com). For dark-adapted cyan testing (4.89 dB [4.29, 5.49]) and for dark-adapted red testing (4.4 dB [3.89, 4.91]), overall similar prediction accuracies could be achieved based on the imaging features only (S1A). Again, neither addition of "patient reliability indices" to the feature set (S1B) nor the addition of the fixation stability (S1C) improved the prediction accuracies (Table 2 and Figure 2).

Prediction accuracies of "inferred sensitivity" in patients with prior perimetry data (scenario 2)

Given that some influencing factors not related to the retina (e.g., lens opacification), might affect the sensitivity at all test-points, while not markedly altering the SD-OCT signal, we speculated that addition of patient-specific perimetry data could improve the prediction accuracies (S2). For mesopic testing, the prediction accuracy was markedly better for scenario

2 (S2A, 3.14 dB [2.9, 3.39]), and varied significantly in dependence of the feature set (likelihood ratio test, P<0.001). It improved slightly with addition of "patient reliability indices" (S2B, 2.93 dB [2.68, 3.19], P<0.01), and fixation stability (2.89 dB [2.64, 3.13], P<0.01) to the feature set. Likewise, the prediction accuracies for the dark-adapted cyan (S2A, 3.35 dB [3.01, 3.7]) and dark-adapted red (3.24 dB [2.94, 3.53]) sensitivity values were markedly better for scenario 2. For both types of testing, the prediction accuracies improved further significantly (P<0.001) with addition of the "patient reliability indices" (dark-adapted cyan testing: 2.92 [2.6, 3.25]; dark-adapted red testing: 2.86 [2.58, 3.14]), and the fixation stability (dark-adapted cyan testing: 2.86 [2.54, 3.19]; dark-adapted red testing: 2.77 [2.5, 3.04]) to the feature set. Overall, prediction accuracies were comparable across a wide degree of sensitivity losses, with slightly higher predictions errors for test-points with sensitivity losses of approximately -15 dB (Figure 2). For severe sensitivity losses of approx. -20 dB, prediction errors were truncated given the floor of the dynamic range of the S-MAIA device (Figure 2).

Feature importance

For the prediction of mesopic sensitivity, the feature importance of the ONL thickness (35% Inc MSE) was markedly higher as compared to all other predictors. The full retinal thickness (27.2% Inc MSE), and the inner retinal thickness (20.4% Inc MSE) constituted the second and third most important features. For the prediction of dark-adapted cyan testing, the ONL thickness (42.2% Inc MSE) also represented the most important feature, followed by the full retinal thickness (23% Inc MSE), and the RPEDC thickness (21.1% Inc MSE). For dark-adapted red testing, the ONL thickness likewise (53.7% Inc MSE) constituted the most important feature, followed by the full retinal thickness (33.4% Inc MSE), and RPEDC thickness (21.8% Inc MSE). ONL thickness was across all three types of testing an outlier in terms of feature importance (Figure 3). Moreover, features related to SD-OCT based layer thicknesses appeared to exhibit generally a higher feature importance as compared to features related to signal intensity.

Structure-function correlation and cone versus rod dysfunction

Plotting of the feature contribution of the ONL thickness (i.e., the association of ONL thickness with retinal sensitivity after accounting for the effects of all other predictors) revealed, that ONL thinning was associated with a marked sensitivity loss for all three types of testing (Figure 4). However, the degree of functional loss per Z-score of ONL thinning varied among the three types of testing. Comparing the plots against each other highlighted that even slight degrees of ONL thinning are associated with a marked DA cyan sensitivity loss, whereas mesopic sensitivity loss was mostly observed in association with more pronounced degrees of ONL thinning (Figure 4).

Inferred sensitivity mapping

The here devised approach allowed for spatially resolved mapping of mesopic, dark-adapted cyan and dark-adapted red inferred sensitivity based on structural imaging data (Figure 5). Importantly, the random forest model based predictions exhibited all anticipated characteristics of the visual field in patient with GA including: (i.) the scotoma imposed by the main GA lesion for all three types of testing, (ii.) the overall shape of the hill-of-vision for mesopic testing, (iii.) the sensitivity maximum for DA cyan testing close to the region of highest rod photoreceptor density (approx.10° to 16° superior to the fovea), (iv.) the central DA cyan scotoma in relation to the rod-free zone, as well as (v.) a region of a distinctly decreased DA cyan-red sensitivity difference surrounding the GA lesion (Figure 5, Supplemental Figure S3, Supplemental Material at AJO.com).

Discussion

We demonstrated that AI-based prediction of retinal sensitivity ("inferred sensitivity") may provide accurate estimates of retinal cone- and rod-photoreceptor function in eyes with GA

secondary to AMD. While psychophysical tests including light- and dark-adapted visual field testing are established as outcome measure in clinical trials, the limited availability of appropriate devices as well as the burden for elderly patients constitutes a challenge.[23] In this context, imaging-based inferred sensitivity may potentially provide as "quasi-function" surrogate outcome a viable alternative. It could also facilitate disease monitoring in clinical practice and allow for the identification of retinal regions with putative pre-degenerative (treatable) disease-stages as discussed in the following.

Regulatory agencies including the U.S. Food and Drug Administration have expressed a clear preference for functional over structural endpoints for AMD and inherited retinal diseases.[24] However, conventional functional outcome measures such as BCVA are suboptimal in presence of GA, particularly for patients with foveal sparing.[5,6,25,26] Refined functional outcome measures such as reading speed and acuity, or fundus-controlled perimetry testing might be more strongly associated with the disease progression.[5,6,25,26] However, these tests require a significant degree of patient cooperation and training of clinical staff. Given the ubiquitous availability of imaging devices including SD-OCT and swept-source OCT platforms, the relative ease of image acquisition and the high-spatial resolution, AI-based inferred sensitivity mapping may constitute as a "quasi-functional" surrogate endpoint, a viable alternative to extensive psychophysical testing.[11,13]

Criteria to establish inferred sensitivity as "quasi-functional" surrogate endpoint would be the demonstration of the prediction accuracy as well as the strength of surrogacy. In this study, we archived in the absence of patient-specific data prediction accuracies (MAE [mesopic, DA cyan, DA red testing]) of 4.64, 4.89 and 4.40 dB, which was markedly better than the respective null models for scenario 1 (Table 2). With addition of patient-specific data, the prediction accuracies drastically improved to up to 2.89, 2.86, 2.77 dB and were close to the test-retest reliability highlighting the prediction accuracy (cf. retest-reliability in Table 1). The rather marked difference between scenario 1 (without any patient-specific data) and scenario 2 (with partial patient-specific data) may be attributable to factors influencing the sensitivity testing, which are not adequality reflected by retinal imaging such as lens opacification. Accordingly, besides of pure imaging-based inferred sensitivity, a combined strategy with brief fundus-controlled perimetry testing (approx. 30 test-points, 5 minutes) and retinal imaging could distinctly improve the accuracy of inferred sensitivity as a meaningful endpoint. Importantly, such a two-track approach would allow to identify patients with poor prediction accuracies. In therapeutic trials, this may be relevant to identify patients with extramacular adverse effects affecting vision (e.g. cataract, optic neuropathy). For the strength of surrogacy, the International Conference on Harmonization Guidelines on Statistical Principles for Clinical Trials has proposed three criteria,[27] of which this study addressed the first one ("biological plausibility of the relationship") in great detail. As shown in Figure 3, ONL thickness was the most important feature for the prediction of retinal sensitivity. In an idealized model of pure photoreceptor degenerations, one would expect that loss of light sensitivity would be proportional to the number of surviving photoreceptors and the length of their outer segments.[28] Given that both of these parameters were shown to correlate to the ONL thickness in an animal model (for autosomal dominant retinitis pigmentosa),[29] the herein observed feature importance of the ONL thickness is comprehensible and underscores the biological plausibility. The observation that OS thickness did not rank among the five most important features is likely related to the close relationship with the other imaging features ("multicollinearity") as well as the relative susceptibility of thin layers to suboptimal image segmentation ("feature noise").

Beyond the application as surrogate outcome, analysis of the feature contributions and mapping of inferred sensitivity may also provide a refined understanding of the biology of the disease. Previously, the phenomenon of localized ONL thickening has been described in the setting of intermediate AMD and was proposed as marker for photoreceptor stress.[30] However, the functional implications were unclear. Here, we evidenced locations with significant ONL thickening as well, but ONL thickening was not associated with sensitivity loss according to the random forest feature contribution plots (Figure 4), suggesting that slight ONL thickening in the setting of GA may constitute a pre-degenerative disease manifestation with no (or only minor) functional implication. On the other hand, ONL thinning was associated with a marked loss of

retinal function across all three types of tests. Importantly, slight ONL thinning resulted in much greater DA cyan sensitivity loss than mesopic sensitivity loss (Figure 4). This observation is very much compatible with post-mortem histopathological studies reporting selective loss of rod photoreceptors in close proximity to GA and/or AMD-related fibrovascular scars.[31,32] Moreover, a greater degree of rod versus cone dysfunction was evidenced in numerous psychophysical studies across various AMD disease stages.[33,34]

Last, spatial mapping of inferred sensitivity may potentially widen the panel of surrogate markers for clinical trials. In contrast to native SD-OCT biomarkers, inferred sensitivity values provide interpretable, functional information. Previous clinical trials for GA have used structural endpoints such as the *"mean change in GA area"* in conjunction with rather short review periods of one to two years.[35] Accordingly, the endpoint reflects only a minor proportion of the imaged retina (essentially, the thin hull immediately surrounding the GA lesion corresponding to the median lateral spread of GA of 106.90 µm/year).[36] Spatially resolved mapping of mesopic, DA cyan and DA red sensitivity (as demonstrated in Figure 5), could allow to evaluate regions distant to the GA boundary, to subclassify these regions according to the degree of functional loss (e.g., normative function, isolated rod dysfunction, combined rod and cone dysfunction) and to follow these regions over time. Moreover, putative therapeutic effects distant to the GA lesion boundary would be more likely to translate to therapeutic effects in intermediate AMD. Of course, AI-based models to predict retinal sensitivity could also be applied (cautiously) in retrospect to re-evaluated former clinical trials with SD-OCT imaging.

Limitations

In terms of methodology, more complex models such as convolutional neural networks (CNNs) as suggested by another group in the setting of macular telangiectasia type 2 may provide more accurate predictions.[11] However, the here used random forest models provide a very good trade-off between predictive performance and measures of feature importance and contribution (Figure 3 and 4) for the evaluation of the biological plausibility. The discrepancy between scenario 1 (without any patient-specific data) and scenario 2 (with partial patient-specific data), even after compensation of behavioral factors ("patient reliability indices"), is suggestive of relevant mediators of light sensitivity that are not adequately reflected by retinal imaging (e.g., lens opacification). Since currently available measures of lens and media opacification are typically not used in retina clinics, addition of such a measure as feature would have reduced the generalizability of the models shown here. Notable strengths of the study are the rigorous nested cross-validation (separating the assessment of the model performance [outer resampling] from the hyperparameter tuning [nested inner resampling]) to avoid optimization bias, the inclusion of multimodal imaging features beyond SD-OCT, the consideration of "patient reliability indices" as behavioral factors, as well as the relatively large number of test-points for extensive dark-adapted psychophysical examination, especially in consideration of the age of the patients.

In summary, this study underscores that inferred sensitivity from SD-OCT and multimodal imaging data may reliably estimate the outcome of functional tests such as mesopic and dark-adapted fundus-controlled perimetry in patients with GA. Moreover, analysis of measures of feature importance and feature contribution confirmed the biological plausibility of these models. *En face* mapping of mesopic and dark-adapted cyan and red inferred sensitivity across the whole imaged area may provide "quasi functional" surrogate markers for clinical trials, while surpassing the limits of fundus-controlled perimetry testing in terms of area covered and spatial-resolution - even in patients unfit for psychophysical testing.

Acknowledgements

a. Funding/Support: This work was supported by the German Research Foundation [DFG grant PF 950/1-1 to MP; DFG grant 658/4-1 and 658/4-2 to MF; DFG grant 2846/1-1 to ML] and the BONFOR GEROK Program of the Faculty of Medicine, University of Bonn [Grant No O-137.0022 and O-137.0025 to MP] and supported in part by an Unrestricted Grant from Research to Prevent Blindness, New York, NY, to the Department of Ophthalmology & Visual Sciences, University of Utah.

b. Financial Disclosures: CenterVue SpA, Padova, Italy has provided research material (S-MAIA) for the conduct of this study. CenterVue had no role in the design or conduct of the experiments.

MP: Heidelberg Engineering: non financial support, Carl Zeiss MedicTec AG: non financial support, Optos: non-financial support

LvdE: Heidelberg Engineering: non financial support, Carl Zeiss MedicTec AG: non financial support, Optos: non-financial support

CD: Heidelberg Engineering: non financial support, Carl Zeiss MedicTec AG: non financial support, Optos: non-financial support

PTM: Heidelberg Engineering: non financial support, Carl Zeiss MedicTec AG: non financial support, Optos: non-financial support

ST: Heidelberg Engineering: non financial support, Carl Zeiss MedicTec AG: non financial support, Optos: non-financial support

ML: Heidelberg Engineering: non financial support, Carl Zeiss MedicTec AG: non financial support, Optos: non-financial support

MS: No financial disclosures.

DLR: No financial disclosures.

MF: Heidelberg Engineering: Grant, personal fees, non financial support, Carl Zeiss MediTec AG: non financial support, Optos: Grant, non financial support, Novartis: Grand, personal fees, Bayer: Grant, personal fees, Alimera: Personal fees, Pending patent: US20140303013 A1

FGH: Heidelberg Engineering: Grant, Personal fees, non financial support, Novartis: Grant, personal fees, Bayer: Grant, personal fees, Genetech: Grant, personal fees, Acucela: Grant, personal fees, Boehringer Ingelheim: Personal fees, Alcon: Grant, personal fees, Allergan: Grant, personal fees, Optos: Grant, personal fees, non financial support, Carl Zeiss MediTec AG: non financial support

SSV: Centervue, Inc: non-financial support; Heidelberg Engineering: Grant, non financial support, Bayer: Grant, personal fees, Genentech/Roche: Grant, personal fees; Alcon/Novartis: Grant, personal fees, Bioeq/Formycon: Grant, Optos: non-financial support, Allergan: Grant, personal fees, Carl Zeiss MediTec AG: non financial support

c. The authors thank Joanna Czauderna, Verena Bonn and Ruth Hassenrik for patient coordination and data acquisition.

References

1. Lim LS, Mitchell P, Seddon JM, Holz FG, Wong TY. Age-related macular degeneration. *Lancet (London, England)*. 2012;379(9827):1728-1738.

2. Miller JW. Beyond VEGF-The Weisenfeld Lecture. *Invest Ophthalmol Vis Sci*. 2016;57(15):6911-6918.

3. Fleckenstein M, Mitchell P, Freund KB, et al. The Progression of Geographic Atrophy Secondary to Age-Related Macular Degeneration. *Ophthalmology*. 2018;125(3):369-390.

4. Sarks JP, Sarks SH, Killingsworth MC. Evolution of geographic atrophy of the retinal pigment epithelium. *Eye*. 1988;2(5):552-577.

5. Lindner M, Nadal J, Mauschitz MM, et al. Combined Fundus Autofluorescence and Near Infrared Reflectance as Prognostic Biomarkers for Visual Acuity in Foveal-Sparing Geographic Atrophy. *Invest Ophthalmol Vis Sci*. 2017;58(6):BIO61-BIO67.

6. Lindner M, Pfau M, Czauderna J, et al. Determinants of Reading Performance in Eyes with Foveal-Sparing Geographic Atrophy. *Ophthalmol Retin*. 2019;3(3):201-210.

7. Takahashi A, Ooto S, Yamashiro K, et al. Photoreceptor Damage and Reduction of Retinal Sensitivity Surrounding Geographic Atrophy in Age-Related Macular Degeneration. *Am J Ophthalmol*. 2016;168:260-268.

8. Pfau M, Müller PL, von der Emde L, et al. Mesopic and Dark-Adapted Two-Color Fundus-Controlled Perimetry in Geographic Atrophy Secondary To Age-Related Macular Degeneration. *Retina*. 2020;40(1):169-180.

9. Curcio CA, Owsley C, Jackson GR. Spare the rods, save the cones in aging and age-related maculopathy. *Invest Ophthalmol Vis Sci*. 2000;41(8):2015-2018.

10. Wu Z, Cimetta R, Caruso E, Guymer RH. Performance of a Defect-Mapping Microperimetry Approach for Characterizing Progressive Changes in Deep Scotomas. *Transl Vis Sci Technol*. 2019;8(4):16.

11. Kihara Y, Heeren TFC, Lee CS, et al. Estimating Retinal Sensitivity Using Optical Coherence Tomography With Deep-Learning Algorithms in Macular Telangiectasia Type 2. *JAMA Netw open*. 2019;2(2):e188029.

12. Sumaroka A, Garafalo A V, Semenov EP, et al. Treatment Potential for Macular Cone Vision in Leber Congenital Amaurosis Due to CEP290 or NPHP5 Mutations: Predictions From Artificial Intelligence. *Invest Ophthalmol Vis Sci*. 2019;60(7):2551-2562.

13. von der Emde L, Pfau M, Dysli C, et al. Artificial intelligence for morphology-based function prediction in neovascular age-related macular degeneration. *Sci Rep*. 2019;9(1):11132.

14. Muller PL, Pfau M, Moller PT, et al. Choroidal Flow Signal in Late-Onset Stargardt Disease and Age-Related Macular Degeneration: An OCT-Angiography Study. *Invest Ophthalmol Vis Sci*. 2018;59(4):AMD122-AMD131.

15. Pfau M, Lindner M, Fleckenstein M, et al. Test-Retest Reliability of Scotopic and Mesopic Fundus-Controlled Perimetry Using a Modified MAIA (Macular Integrity Assessment) in Normal Eyes. *Ophthalmologica*. 2017;237(1):42-54.

16. Pfau M, Lindner M, Muller PL, et al. Effective Dynamic Range and Retest Reliability of

Dark-Adapted Two-Color Fundus-Controlled Perimetry in Patients With Macular Diseases. *Invest Ophthalmol Vis Sci*. 2017;58(6):BIO158-BIO167.

17. Pfau M, Lindner M, Gliem M, et al. Mesopic and dark-adapted two-color fundus-controlled perimetry in patients with cuticular, reticular, and soft drusen. *Eye*. 2018;32(12):1819-1830.

18. R Core Team 3.0.1. A Language and Environment for Statistical Computing. *R Found Stat Comput*. 2013;2:https://www.R-project.org.

19. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Software; Vol 1, Issue 1* . October 2015.

20. Kuhn M. Caret package. *J Stat Softw*. 2008;28(5):1-26.

21. Welling SH, Refsgaard HHF, Brockhoff PB, Clemmensen LH. Forest floor visualizations of random forests. *arXiv Prepr arXiv160509196*. 2016.

22. Breiman L. Random Forests. *Mach Learn*. 2001;45(1):5-32.

23. McGuigan DB, Roman AJ, Cideciyan A V., et al. Automated light- And dark-adapted perimetry for evaluating retinitis pigmentosa: Filling a need to accommodate multicenter clinical trials. *Investig Ophthalmol Vis Sci*. 2016;57(7):3118-3128.

24. Csaky K, Ferris F 3rd, Chew EY, Nair P, Cheetham JK, Duncan JL. Report From the NEI/FDA Endpoints Workshop on Age-Related Macular Degeneration and Inherited Retinal Diseases. *Invest Ophthalmol Vis Sci*. 2017;58(9):3456-3463.

25. Sayegh RG, Sacu S, Dunavolgyi R, et al. Geographic Atrophy and Foveal-Sparing Changes Related to Visual Acuity in Patients With Dry Age-Related Macular Degeneration Over Time. *Am J Ophthalmol*. 2017;179:118-128.

26. Sunness JS, Rubin GS, Zuckerbrod A, Applegate CA. Foveal-Sparing Scotomas in Advanced Dry Age-Related Macular Degeneration. *J Vis Impair Blind*. 2008;102(10):600-610.

27. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. *Stat Med*. 1999;18(15):1905-1942.

28. Jacobson SG, Aleman TS, Cideciyan A V, et al. Identifying photoreceptors in blind eyes caused by RPE65 mutations: Prerequisite for human gene therapy success. *Proc Natl Acad Sci U S A*. 2005;102(17):6177-6182.

29. Machida S, Kondo M, Jamison JA, et al. P23H rhodopsin transgenic rat: Correlation of retinal function with histopathology. *Investig Ophthalmol Vis Sci*. 2000;41(10):3200-3209.

30. Sadigh S, Cideciyan A V., Sumaroka A, et al. Abnormal thickening as well as thinning of the photoreceptor layer in intermediate age-related macular degeneration. *Investig Ophthalmol Vis Sci*. 2013;54(3):1603-1612.

31. Curcio CA, Medeiros NE, Millican CL. Photoreceptor loss in age-related macular degeneration. *Investig Ophthalmol Vis Sci*. 1996;37(7):1236-1249.

32. Shelley EJ, Madigan MC, Natoli R, Penfold PL PJ, Shelley EJ, Madigan MC, Natoli R, Penfold PL, Provis JM. Cone degeneration in aging and age-related macular degeneration. *Arch Ophthalmol*. 2009;127(4):483-492.

33. Owsley C, Jackson GR, Cideciyan A V., et al. Psychophysical evidence for rod

vulnerability in age-related macular degeneration. *Investig Ophthalmol Vis Sci*. 2000;41(1):267-273.

34. Flynn OJ, Cukras CA, Jeffrey BG. Characterization of Rod Function Phenotypes Across a Range of Age-Related Macular Degeneration Severities and Subretinal Drusenoid Deposits. *Invest Ophthalmol Vis Sci*. 2018;59(6):2411-2421.

35. Holz FG, Sadda SR, Busbee B, et al. Efficacy and Safety of Lampalizumab for Geographic Atrophy Due to Age-Related Macular Degeneration: Chroma and Spectri Phase 3 Randomized Clinical Trials. *JAMA Ophthalmol*. 2018;136(6):666-677.

36. Fleckenstein M, Schmitz-Valckenberg S, Adrion C, et al. Tracking progression with spectral-domain optical coherence tomography in geographic atrophy caused by age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2010;51(8):3846-3852.

Figure captions

Figure 1. Segmentation of spectral-domain optical coherence tomography (SD-OCT) data and registration of functional fundus-controlled perimetry and imaging data. Panel A shows the layer definitions used in this study. The full retina (azure overlay) ranged from the internal limiting membrane (ILM) to Bruch's membrane (BrM), the inner retina (purple overlay) from the ILM to the outer plexiform layer (OPL) / outer nuclear layer (ONL) boundary. The here used definition of the ONL (navy blue overlay) included Henle's fiber layer (HFL) and ranged from the OPL/ONL boundary to the external limiting membrane (ELM). The inner segments (IS, orange overlay) ranged from the ELM to the ellipsoid zone (EZ). The outer segments (OS, green overlay) from the EZ to the interdigitation zone. The retinal pigment epithelium drusen complex (RPEDC, red overlay) encompassed subretinal drusenoid deposits (reticular pseudodrusen), the retinal pigment epithelium, drusen with BrM as outer boundary. Panel B shows the multimodal registration of the functional fundus-controlled perimetry data to the structural imaging data including SD-OCT as well as normalized fundus-autofluorescence (FAF) and infrared reflectance (IR) images.

Figure 2. Prediction accuracies. The upper row (**A**) dot-plots show the patient-wise prediction accuracies (mean absolute error [MAE]) for mesopic testing, dark-adapted (DA) cyan testing and DA red testing. The prediction accuracies were evaluated for two scenarios: (S1) with patient-wise leave-one-out cross validation ("prediction accuracy for an unknown patient"), and (S2) with partial addition of perimetry data ("prediction accuracy for a patient with prior perimetry data available"). For clarity, results for feature set A (imaging data only) are shown (cf. Supplemental Figure S2 for all feature sets). The individual points were slightly jittered along the x-axis and plotted semi-transparent to avoid over-plotting. The red dot indicates the mean estimate for the MAE and the red lines the 95% confidence interval for estimate. Notably, across all three types of testing, addition of patient-specific sensitivity data markedly improved the prediction accuracies. The lower row (**B**) shows the Bland-Altman plots for the point-wise prediction accuracies for scenario 2A (red solid lines indicate the mean difference, red dashed denote the limits of agreement [LOA]).

Figure 3. Feature importance plots. The feature importance plots show the permutation-based importance in terms of the percentage increase in mean squared error (% Inc MSE). The colors indicate the respective retinal structure represented by the respective imaging feature. Please note, overall spectral-domain optical coherence tomography layer thicknesses exhibited higher feature importance values compared to intensity related features. Notably, the outer nuclear layer thickness constituted the most important imaging feature across all three types of testing.

Abbreviations: full retina (FRET), inner retina (IRET), outer nuclear layer (ONL), inner segments (IS), outer segments (OS), retinal pigment epithelium drusen complex (RPEDC), fundus-autofluorescence (FAF) and infrared reflectance (IR) imaging

Figure 4. Structure-function relationship. The feature contribution plots (i.e., plots of the association of a given predictor with retinal sensitivity after accounting for the effects of all other predictors) show the relationship of the most important feature (outer nuclear layer thickness [ONL]) with the inferred sensitivity loss for each type of testing (mesopic, dark-adapted [DA] cyan and DA red testing). The small semi-transparent circles represent all patient data. Hereby, the x-axis depicts the ONL thickness (cf. Figure 1) in terms of the z-score (number of normative standard deviations by which a given value deviates from the normative mean). The dashed vertical lines denote the ONL thickness normal range. Notably, across all three types of testing, ONL thickening (even outside the reference range) was not associated with functional loss, while thinning was associated with a marked loss of retinal sensitivity. While the magnitudes of the DA cyan and DA red sensitivity losses as function of ONL thickness were similar, the association of ONL thickness with DA cyan sensitivity loss appeared to be slightly steeper. The goodness-of-visualization R^2 values were 0.88, 0.91 and 0.90 for these plots.[21] Therefore, the residual variance of feature contributions not explained by these plots (due to possible interaction effects) was low.

Figure 5. Mapping of "inferred sensitivity". In this exemplary patient, mesopic, dark-adapted (DA) cyan and DA red sensitivity as well as the DA cyan-red sensitivity difference were predicted based on structural imaging data. The white dashed lines in the color fundus photography image indicate the spectral-domain optical coherence tomography (SD-OCT) B-scan positions. Notably, the random forest models adequatly recognized the scotoma related to the main geographic atrophy (GA) lesion. Further, the two small foci of GA labeled by the two white arrowheads in the plot for inferred mesopic sensitivity (cf. color fundus photo and upper B-scan)", were recognized as scotomata. For DA cyan testing, the respective random forest model accounted adequately for the central rod-free area (grey arrowhead). Last, the inferred cyan-red sensitivity difference plot revealed a region with a markedly reduced DA cyan-red sensitivity difference sourrounding the area of atrophy. The senstivity maps were minimally smoothed (median filtering, window: 3 px x 3 px).