Special Issue

Effective Dynamic Range and Retest Reliability of Dark-Adapted Two-Color Fundus-Controlled Perimetry in Patients With Macular Diseases

Maximilian Pfau, Moritz Lindner, Philipp L. Müller, Johannes Birtel, Robert P. Finger, Wolf M. Harmening, Monika Fleckenstein, Frank G. Holz, and Steffen Schmitz-Valckenberg

Department of Ophthalmology, University of Bonn, Bonn, Germany

Correspondence: Steffen Schmitz-Valckenberg, Department of Ophthalmology, University of Bonn, Ernst-Abbe-Str. 2, 53127 Bonn, Germany; steffen.schmitz-valckenberg@ukb.uni-bonn.de.

Submitted: January 9, 2017
Accepted: April 18, 2017

Citation: Pfau M, Lindner M, Müller 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:BIO158–BIO167. DOI:10.1167/iovs.17-21454

Purpose. To determine the effective dynamic range (EDR), retest reliability, and number of discriminable steps (DS) for mesopic and dark-adapted two-color fundus-controlled perimetry (FCP) using the S-MAIA (Scotopic-Macular Integrity Assessment) "micro-perimeter."

Methods. In this prospective cross-sectional study, each of the 52 eyes of 52 subjects with various macular diseases (mean age 62.0 ± 16.9 years; range, 19.1–90.1 years) underwent duplicate mesopic (achromatic stimuli, 400–800 nm), dark-adapted cyan (505 nm), and dark-adapted red (627 nm) FCP using a grid of 61 stimuli covering 18° of the central retina. The EDR, the number of DS, and the retest reliability for point-wise sensitivity (PWS) were analyzed. The effects of fixation stability, sensitivity, and age on retest reliability were examined using mixed-effects models.

Results. The EDR was 10 to 30 dB with five DS for mesopic and 4 to 17 dB with four DS for dark-adapted cyan and red testing. PWS retest reliability was good among all three types of retinal sensitivity assessments (coefficient of repeatability ±5.79, ±4.72, and ±4.77 dB, respectively) and did not depend on fixation stability or age. PWS had no effect on retest variability in dark-adapted cyan and dark-adapted red testing but had a minor effect in mesopic testing.

Conclusions. Combined mesopic and dark-adapted two-color FCP allows for reliable topographic testing of cone and rod function in patients with various macular diseases with and without foveal fixation. Retest reliability is homogeneous across eccentricities and various degrees of scotoma depth, including zones at risk for disease progression. These reliability estimates can serve for the design of future clinical trials.

Keywords: microperimetry, dark-adapted two-color perimetry, scotopic sensitivity, macular disease

The physical dynamic range of a perimetry device, which is defined by the range of the dimmest to the brightest presentable stimulus, is not necessarily representative of its effective dynamic range (EDR).[1] While the physical dynamic range is determined by arbitrary factors, including technical limitations, the upper limit of the EDR is dependent on both the physical dynamic range and the differential (mesopic testing) or absolute (dark-adapted testing) light sensitivity of the normal human eye.[1] The lower limit of the EDR was shown to be mainly limited by the retest variability in standard automated perimetry, which was shown to increase substantially at low sensitivities.[1–4] For sensitivity values reported as being 15 dB or lower, the 95% confidence intervals for the retest sensitivities at the same location were shown to include 0 dB (the floor of the physical dynamic range).[1–4] Thus, results of 15 dB or less are statistically indistinguishable from the physical floor and do not contribute to the EDR of the device.[1–4] This phenomenon has been attributed to inaccurate stimulus placement within sparse, irregular receptive fields.[1,3] Wall and associates[1] have proposed a framework to assess the EDR. Using this approach, the Humphrey field analyzer (Carl Zeiss Meditec, Dublin, CA, USA) has been shown to have an EDR of only about 2 log units in patients with glaucoma despite its physical dynamic range of 5 log units.[1] Moreover, the authors introduced a measure to determine the number of discriminable steps (DS) for a given perimetry device that is scale-independent and that takes into account the variability of retest reliability in dependence of the retinal sensitivity.[1]

Dark-adapted two-color perimetry has been used in various macular diseases to differentiate between rod- and cone-mediated function.[5–13] As of late, it has been also applied in form of standard automated perimetry as an outcome measure in human gene therapy trials.[14–16] A recent pilot study in patients with AMD—the most common cause of legal blindness in developed countries—highlighted the potential applicability of dark-adapted two-color perimetry in common diseases.[13,17]

So called fundus-controlled perimetry (FCP) that uses retina tracking for accurate stimulus placement allows for reliable sensitivity testing even in patients with extrafoveal or unstable fixation.[18–23] Longitudinal studies in patients with AMD demonstrated that mesopic FCP allows for precise structure–function correlation even over time.[24,25] Recently, the S-MAIA (Scotopic-Macular Integrity Assessment; CenterVue SpA, Padova, Italy), a FCP device that allows for mesopic (achromatic

stimuli), dark-adapted cyan and dark-adapted red testing with confocal scanning laser ophthalmoscopy–based (cSLO-based) fundus tracking has become available.[23] It has been demonstrated that this device allows for a reliable assessment of both mesopic and dark-adapted retinal function in normal subjects.[23] The normative data of the S-MAIA device must be considered for the interpretation of dark-adapted two-color FCP test results.[23] In short, a cyan-red difference of approximately 0 dB would be indicative of normal rod function (as observed at an eccentricity of 7°), whereas a decrease of the cyan-red difference (i.e., more negative values) would indicate that rod-dysfunction exceeds cone-dysfunction.[26,27] A decrease of the cyan-red difference value would indicate that rod dysfunction exceeds cone dysfunction.[26,27]

The aim of this study was to assess the EDR, retest reliability, and number of DS for mesopic, dark-adapted cyan, and dark-adapted red FCP, and for the cyan–red difference in patients with various macular diseases using the recently introduced S-MAIA device. We hypothesized that the retest variability is homogenous across the dynamic range (homoscedastic) since retina tracking should allow for accurate stimulus placement. Furthermore, the influence of eccentricity, fixation stability, and retinal sensitivity on the point-wise retest reliability was explored.

Methods

Subjects

To be included, at least one eye of the subjects needed to present with a known macular disease. All patients were recruited at the Department of Ophthalmology, University of Bonn, Germany. Exclusion criteria included refractive errors $\geq$5.00 diopters of spherical equivalent and >1.50 diopters of astigmatism assessed by autorefraction (ARK-560A; Nidek, Gamagori, Japan), as well as a history of glaucoma or relevant anterior segment diseases with media opacities. If both eyes met the inclusion criteria, the eye with better visual acuity was included. Apart from taking the medical history, all subjects underwent routine ophthalmological examinations, including best corrected visual acuity, slit-lamp, and funduscopic examinations. Spectral-domain optical coherence tomography (SD-OCT) raster scanning was performed using a 30° × 25° scan field (121 B-scans, automated real time (ART) mode 20 frames, centered on the fovea) (Spectralis OCT2; Heidelberg Engineering, Heidelberg, Germany).

The study was prospectively approved by the Institutional Review Board of the University of Bonn (ethics approval ID: 191/16). After explanation of the nature and possible consequences of the study, informed written consent was obtained from all subjects. The protocol followed the tenets of the Declaration of Helsinki.

Fundus-Controlled Perimetry

Prior to imaging and FCP testing, pupil dilation was performed using 1.0% tropicamide to facilitate imaging and fundus tracking. A short mesopic practice examination was performed in patients with no prior perimetry or FCP experience. All patients underwent duplicate mesopic (achromatic stimuli, 400–800 nm) FCP, followed by 30 minutes of dark adaptation (light level < 0.1 lux), then duplicate dark-adapted cyan (505 nm) FCP, and finally duplicate dark-adapted red (627 nm) FCP (Supplementary Fig. S1). All tests were carried out using a grid of 61 stimuli covering 18° of the central retina with the S-MAIA (Fig. 1). The basic design of the MAIA platform has been described previously.[23,28] Briefly, it employs cSLO with a central wavelength of 850 nm (36.5° × 36.58°, 25 frames per second) for fundus tracking.

The stimuli grid distribution contained five concentric rings with 12 evenly distributed stimuli (angular distance of 30° to each other), separated at 1°, 3°, 5°, 7°, and 9° from a central stimulus (Fig. 1). A 645-nm red ring with 1° diameter size was used as target of fixation. Initially, the fixation target was presented to each patient with the preset luminance for mesopic (2.6 ± 0.5 apostilb [asb]) or the preset luminance for dark-adapted testing (0.01 asb). In dark-adapted testing, the luminance was increased if patients reported that the fixation target was not visible (range, 0.01–2.6 asb). Goldmann III-sized stimuli were applied, directly projected on the retina by means of LED projectors, and each presented for 200 ms. For mesopic testing, achromatic stimuli (400–800 nm, dynamic range of 36 dB) were presented using a 4-2 staircase threshold strategy, while patients observed the fixation target against a background of 1.27 cd/m^2 (4 asb).[23,29,30] The minimum and maximum luminance of the stimuli were 0.08 cd/m^2 (0.25 asb) and 318 cd/m^2 (1000 asb), respectively. For dark-adapted testing, two different LED types were used—cyan (505 nm) and red (627 nm)—and projected with a 2-1 staircase threshold strategy (dynamic range of 20 dB), while patients observed the fixation target against a background of <0.0001 cd/m^2.[29,30] The minimum and maximum luminance of the cyan and red stimuli were 0.0025 scotopic (scot.) cd/m^2 and 0.25 scot. cd/m^2. The conversion of radiant energy to luminous energy was based on the scotopic luminosity function $V'(\lambda)$ as adopted by the CIE (Commission Internationale de l'Éclairage) in 1951.[26,31]

Irrespective of the type of testing (mesopic, dark-adapted cyan, dark-adapted red), the MAIA started examinations with one paracentral test stimulus in each quadrant (eccentricity of 3°, angular position of 0°, 90°, 180°, and 270°, respectively). These four threshold values were then used to adjust the initial brightness levels for measuring the remaining test loci in each of the corresponding quadrants. For mesopic function, the testing sensitivity started with a level of 2 dB brighter of the respective initial threshold value using a 4-2 dB full-threshold strategy.[29,30] For dark-adapted testing, the testing sensitivity started with a level of 2 dB dimmer than the respective initial threshold, particularly to avoid bleaching, and used a 2-1 dB full-threshold strategy.[23,29,30] These differences between mesopic and dark-adapted FCP were not modifiable as they were included in the presetting of the device. After the first test and for each of the three types of functional testing, subjects were asked to sit back and rest until they were ready to proceed to the second test. The second test was performed using the follow-up mode. For mesopic function, the starting brightness for each stimulus was automatically set to 2 dB brighter than the previously determined threshold, while the threshold sensitivity for dark-adapted testing was preset as the initially determined brightness.

The frequency of false-positive responses was measured by presentation of suprathreshold stimuli to the optic nerve head. Test runs with false-positive responses >33% were not excluded from the analysis, since the predictive value of fixation losses with regard to retest variability was shown to be insignificant.[32] Overall, only four tests from three different patients out of 300 tests exhibited false-positive rates >33%.

Outcome Measures and Statistical Analyses

Statistical analyses were performed using the software environment R. Data were evaluated for possible learning curve and bleaching effects using mixed-effects-model analysis. The primary outcome measure, the EDR, and the number of DS were analyzed as proposed by Michael Wall and associates.[1] The upper EDR boundary was defined as the 99.5th percentile,

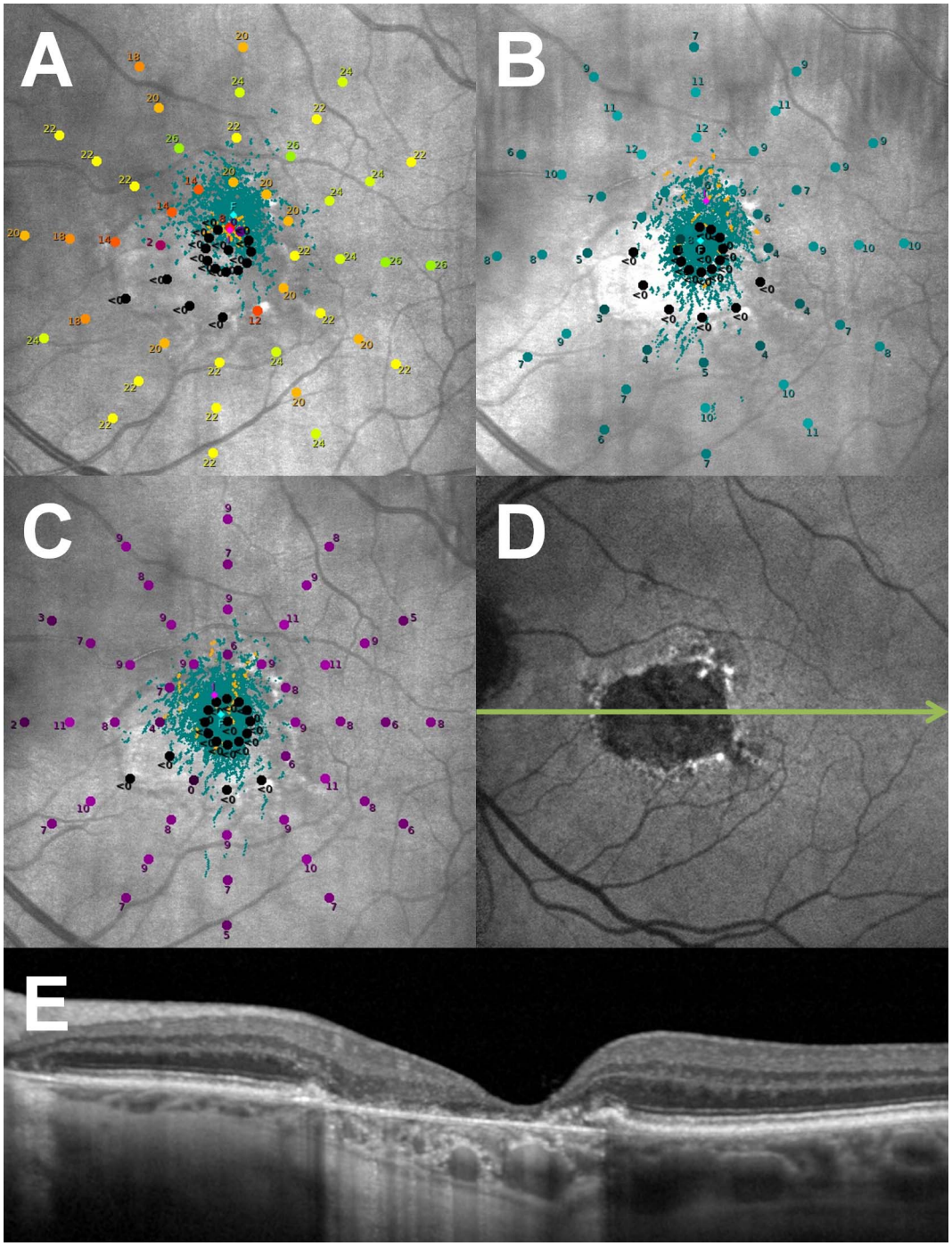

Figure 1. Exemplary report from the S-MAIA of a patient with geographic atrophy secondary to AMD for mesopic (**A**), dark-adapted cyan (**B**), and dark-adapted red (**C**) FCP. The *colored dots* show stimulus positions, and dot saturation and hue denote measured sensitivity values. The numbers indicate sensitivity in decibels. Smaller *turquoise dots* are tracked retinal locations during the measurement, and their spatial extent represents fixation stability. (**D**) Shown are the near infrared autofluorescence image and (**E**) the corresponding SD-OCT (B-scan location indicated by the *arrow* in **D**).

considering values above this threshold as noise due to false-positive responses.[1] The lower EDR boundary was defined as the highest value where 5% of the retest values are 0 dB, since this would indicate that the reported value is statistically indifferent from the floor of the physical dynamic range.[1] To determine the number of DS, Loess-smoothed curves were used to denote the fifth and 95th percentile retest values for the 52 subjects. The first DS was defined as the vertical distance from where the fifth percentile boundary crosses the line of unity (45° line) to where the line intersects the 95th percentile. It indicates how low the retest value would have to be to be considered statistically worse as compared to the first test value (considering a one-sided 95% confidence interval). The next DS commences at this value and is again indicated by an arrow drawn to the 95th percentile. This staircase process starts in the top (right) end of the decibel range and continues until 0 dB (left) is reached and the number of DS is counted (Fig. 2).

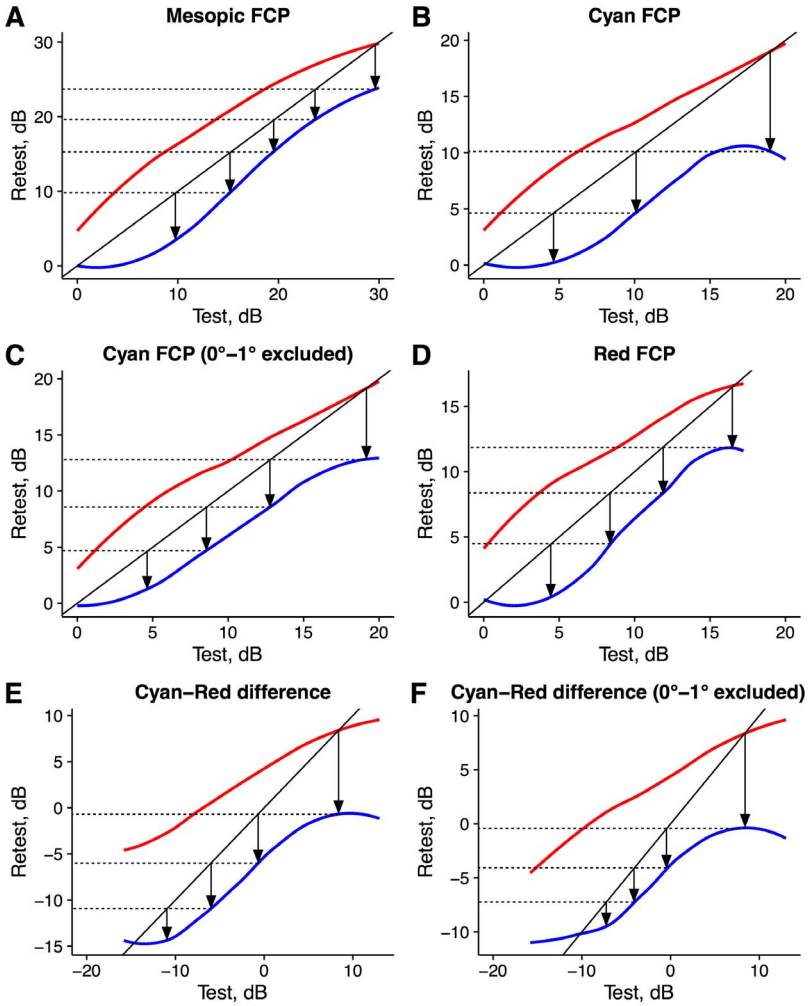

Figure 2. Graphical analysis of the number of DS (*vertical arrows*) for mesopic (**A**), dark-adapted cyan (**B**), dark-adapted cyan excluding data from 0 to 1° (**C**), dark-adapted red (**D**), the cyan–red difference (**E**), and the cyan–red difference excluding data from 0 to 1° (**F**). The Loess-smoothed lines were used to indicate the fifth (*red*) and 95th (*blue*) percentile of retest values including the data of all patients. The 45° line denotes the line of equal test-retest values. The first discriminable step commences at the intersection of fifth percentile and the line of equal test-retest values and ends with the intersection with the 95th percentile. The next structure of the staircase is a horizontal line drawn from the 95th percentile to the line of equal test-retest values. This staircase process starts in the top ("*right*") end of the decibel range and continues until 0 dB ("*left*") is reached and the number of DS is counted.

As secondary outcome measure, the point-wise sensitivity (PWS) intrasession retest reliability was assessed using the 95% coefficient of repeatability (CoR) for the three types of testing.[33] Last, the effects of global factors (type of test, fixation stability, frequency of false-positive responses), and locus-specific factors (eccentricity, retinal sensitivity) were analyzed with regard to their influence on retest reliability using mixed-effects models. Hereby, the point-wise retest variance was modeled as outcome variable. The 0.95 bivariate contour ellipse area (0.95 bivariate contour ellipse area [BCEA], the area that covers 95% of the fixation positions) as measure of fixation stability was log-transformed to obtain normally distributed data.[34] P values < 0.05 were considered statistically significant. Additionally, effect estimates with respective standard errors were provided for all effects with $P < 0.10$.

Results

A total of 52 eyes of 52 patients were recruited between August 2016 and November 2016. All patients (mean ± SD age 62.0 ± 16.9 years, range 19.1–90.1 years, 29 female) underwent the complete standardized protocol, including duplicate mesopic, dark-adapted cyan, and dark-adapted red FCP, and were included in the following analysis (Supplementary Table S1 provides a detailed overview of the included macular diseases).

Predictors of PWS

Linear mixed-effects models considering the hierarchical nature of the data (test locations nested within eyes) were used to examine the influence of test number (second versus first test) and eccentricity on PWS for all three types of testing with regard to learning curve or bleaching effects (Table 1). The test number had no significant effect on PWS in mesopic testing. In dark-adapted cyan testing, there was a decrease in sensitivity between the first and second test that exhibited a significant interaction with the eccentricity ($\chi 2(5) = 49.71$, $P < 0.001$), that is, the test-retest difference was most pronounced at 0 and 1° (–1.79 and –0.88 dB) and almost absent at 3° (–0.43 dB), 5° (–0.29 dB), 7° (–0.17 dB),

Table 1. Mixed-Effects Models for PWS (Considering Stimulus Location Nested in Eye as Random Effect)

	Mesopic		Scotopic Cyan		Scotopic Red		
Variable	**Estimate, dB**	**95% CI**	**Estimate, dB**	**95% CI**	**Estimate, dB**	**95% CI**	***P* Value†**
Intercept*	18.75	16.94 to 20.56	2.27	1.05 to 3.49	7.56	6.38 to 8.74	
Eccentricity*							<0.001
1°	0.71	−0.67 to 2.09	0.06	0.89 to 1.02	0.58	0.36 to 1.52	
3°	0.96	−0.42 to 2.34	3.95	3 to 4.9	0.86	−0.08 to 1.8	
5°	1.51	0.13 to 2.89	5.92	4.97 to 6.87	1.3	0.36 to 2.24	
7°	1.23	−0.15 to 2.61	6.62	5.67 to 7.58	0.72	−0.22 to 1.66	
9°	0.71	−0.67 to 2.09	6.81	5.86 to 7.77	−0.09	−1.03 to 0.85	
Angular position*							<0.001
30°	0.6	−0.18 to 1.37	0.98	0.48 to 1.48	0.17	−0.32 to 0.67	
60°	0.61	−0.17 to 1.38	1.86	1.36 to 2.36	0.38	−0.12 to 0.87	
90° (superior)	0.73	−0.05 to 1.5	2.13	1.62 to 2.63	0.58	0.09 to 1.07	
120°	0.38	−0.39 to 1.16	2.27	1.76 to 2.77	0.44	−0.06 to 0.93	
150°	0.58	−0.2 to 1.35	1.65	1.15 to 2.15	0.37	−0.12 to 0.86	
180° (temporal)	1.18	0.41 to 1.96	1.2	0.7 to 1.71	1.05	0.56 to 1.54	
210°	0.65	−0.12 to 1.42	0.74	0.23 to 1.24	0.56	0.07 to 1.05	
240°	0.13	−0.64 to 0.91	0.45	−0.05 to 0.95	−0.05	−0.54 to 0.44	
270° (inferior)	−0.02	−0.79 to 0.76	0.2	−0.3 to 0.71	−0.25	−0.74 to 0.24	
300°	0.03	−0.74 to 0.81	0.03	−0.47 to 0.53	−0.38	−0.87 to 0.12	
330°	−0.18	−0.96 to 0.59	−0.2	−0.7 to 0.3	−0.23	−0.72 to 0.27	
Test number (second vs. first test)*	0.03	−0.07 to 0.14	−1.79	−2.43 to −1.15	0.13	−0.52 to 0.79	<0.001
Test number × eccentricity‡							<0.001
Second test – 1°			0.91	0.24 to 1.57	−0.16	−0.84 to 0.53	
Second test – 3°			1.36	0.69 to 2.02	−0.39	−1.07 to 0.29	
Second test – 5°			1.5	0.84 to 2.17	−0.7	−1.38 to −0.02	
Second test – 7°			1.62	0.95 to 2.28	−0.44	−1.12 to 0.24	
Second test – 9°			1.31	0.64 to 1.97	−0.58	−1.26 to 0.11	

The average test value for the mean measured sensitivity for the eccentricity of 0° and the angular position of 0° (nasal) of the first test is indicated by the intercept. Please note that no learning effect was observable for mesopic testing. For dark-adapted cyan and red testing, a significant effect of the test number (second vs. first test) that interacted with the eccentricity was observable. For example, the sensitivity for cyan testing in the second test at an eccentricity of 1° was reduced on average by −0.88 dB (i.e., −1.79 + 0.91 dB) as compared to the first test.v

* The intercept indicates the mean measured sensitivity for the eccentricity of 0° and the angular position of 0° (nasal) of the first test run.

† *P* values were obtained from the likelihood ratio tests.

‡ A likelihood ratio test indicated a significant interaction between test number and eccentricity for scotopic cyan and red testing. No significant interaction was observed for mesopic testing.

and 9° (–0.48 dB). In dark-adapted red testing, there was also a minor decrease of sensitivity that exhibited significant interaction with the eccentricity ($\chi 2(5) = 19.811$, $P < 0.01$); however, the effect sizes were small. At 5° (–0.57 dB), 7° (–0.31 dB), and 9° (–0.45 dB), the sensitivity values were slightly lower in the second test run, whereas the sensitivity at 0° (+0.13 dB), 1° (–0.03 dB), and 3° (–0.26 dB) did not appear to change (Table 1).

Since the systematic threshold shift for dark-adapted cyan testing at the fovea (0 and 1°) was most likely attributable to the fixation target (compare discussion), all of the following analyses for dark-adapted cyan testing were performed not only with the data of all 61 stimuli but also with exclusion of data from 0 to 1° (data of 48 eccentric stimuli only).

Effective Dynamic Range

The ceiling of the EDR (defined as the 99.5th percentile) was 30 dB for mesopic, 17 dB for dark-adapted cyan, 17 dB for dark-adapted red testing, and 11 dB for the cyan-red difference (Table 2). The floor of the EDR was 10 dB for mesopic, 9 dB for dark-adapted cyan, 4 dB for dark-adapted cyan, excluding data from 0 to 1°, 4 dB for dark-adapted red testing, and −10 dB for the cyan-red difference (Table 2).

Number of Discriminable Steps

Mesopic testing had the greatest number of DS with five steps (Fig. 2A). Both dark-adapted cyan (with exclusion of data at 0–

Table 2. Analysis of the Floor and Ceiling of the EDR and Number of DS

	EDR Floor	EDR Ceiling	
	5% of Retest Values 0 dB†	**Fewer Than 0.5% of Values**	**No. of DS**
Mesopic	10 dB	30 dB	5
Scotopic cyan	9 dB	17 dB	3
Scotopic cyan*	4 dB	17 dB	4
Scotopic red	4 dB	17 dB	4
Cyan–red difference	−10 dB	+11 dB	4
Cyan–red difference*	−10 dB	+11 dB	4

* Results for scotopic cyan testing with exclusion of data from 0 to 1°. The analysis was conducted since data at 0 to 1° was affected by bleaching effects due to the fixation target.

† For the cyan–red difference, the floor was defined as (1) the highest value where 5% of the retest values are −20 dB (i.e., maximal possible cyan–red difference) or (2) the highest value with −20 dB as the most frequent retest value.

Table 3. The CoRs Indicate the Value Below Which the Absolute Differences Between Two Measurements Would Lie With 0.95 Probability

	CoR, dB					
	Mesopic	Cyan	Cyan*	Red	Cyan–Red	Cyan–Red*
Overall	5.79	4.72	4.3	4.77	6.7	6.3
0–1°	5.92	6.02	NA	4.93	8.01	NA
3°	5.06	4.31	4.31	4.47	6.3	6.3
5°	6.07	4.28	4.28	4.88	6.25	6.25
7°	6.12	4.09	4.09	4.57	6.14	6.14
9°	5.71	4.51	4.51	4.96	6.5	6.5

* Results for scotopic cyan testing with exclusion of data from 0 to 1°. The analysis was conducted since data at 0 to 1° was affected by bleaching effects due to the fixation target.

1°) and dark-adapted red testing had four DS (Figs. 2C, 2D). Without exclusion of data from 0 to 1°, a bulging of the fifth percentile toward the lower end of the dynamic range (<5 dB) was observed, resulting in only three DS. The cyan-red difference showed four DS.

CoR and Frequency of Change in PWS

The CoR, that is, the value below which the absolute differences between two measurements would lie with 0.95 probability, was ±5.79 dB for mesopic, ±4.72 dB for dark-adapted cyan, ±4.30 dB for dark-adapted cyan with exclusion of data from 0 to 1°, and ±4.77 dB for dark-adapted red testing (Table 3). The CoR for the cyan-red difference was ±6.70 dB and ±6.30 dB (excluding data from 0 to 1°). Table 4 lists the frequency of change in PWS. For all three types of testing, the cumulative percentage of stimulus points within ±2 dB repeatability was >70%. Bland-Altman plots for all three types of testing are provided in Figure 3.

Influence of Patient-Specific (Global) Factors on Retest Reliability

Next, we analyzed the effect of patient-specific (global) factors (i.e., factors that would affect the retest variability of every point) on retest reliability. Mixed-effects model analyses considering test type as fixed effect and the stimulus location nested within the eye as random effect disclosed that the pointwise retest variance was higher in mesopic testing as compared to dark-adapted cyan and dark-adapted red testing (estimate ± standard error of 4.37 ± 0.23 dB^2 for mesopic, 2.90 ± 0.22 dB^2 for dark-adapted cyan, and 2.96 ± 0.22 dB^2 for dark-adapted red testing; $\chi2(2) = 58.50$, $P < 0.001$). Age had no significant effect on retest variance (0.006 ± 0.01 dB^2/year; $\chi2(1) = 0.238$, $P = 0.626$). Neither the 0.95 bivariate contour ellipse area (0.95 BCEA) as measure of fixation instability of the first nor of the second test run exhibited a significant effect on retest variance (0.56 ± 0.32 dB^2/log10(deg^2); $\chi2(1) = 3.0989$, $P = 0.078$, and 0.56 ± 0.34 dB^2/log10(deg^2); $\chi2(1) = 2.7163$, $P = 0.099$). Also, the false-positive response rates (to suprathreshold stimuli to the optic nerve head) of the first and second test run exhibited no significant effect on the retest variance ($\chi2(1) = 2.4073$, $P = 0.121$, and $\chi2(1) = 2.1069$, $P = 0.147$). Graphical analysis of the test-retest sensitivity differences plotted against the above-mentioned factors confirmed additionally that no nonlinear relationships were present (Supplementary Figs. S2–S4).

Table 4. Frequency of Change in PWS

Type of Testing	Change, dB	Frequency, Count	Cumulative Percentage
Mesopic	No change	1304	41.1
	±1 to ±2	1057	74.4
	±3 to ±4	559	92
	±5 to ±6	156	96.9
	>±6	96	100
Scotopic cyan	No change	1021	32.2
	±1 to ±2	1489	79.1
	±3 to ±4	464	93.7
	±5 to ±6	121	97.5
	>±6	77	100
Scotopic red	No change	787	24.8
	±1 to ±2	1706	78.6
	±3 to ±4	489	94
	±5 to ±6	111	97.5
	>±6	79	100
Cyan–red difference	No change	635	20
	±1 to ±2	1365	63
	±3 to ±4	722	85.8
	±5 to ±6	257	93.9
	>±6	193	100

Influence of Locus-Specific Factors on Retest Reliability

Next, we evaluated whether the overall repeatability is representative of the repeatability in a locus-specific (eccentricity, retinal sensitivity) context (excluding data of 0–1° for dark-adapted cyan testing) using mixed-effects model analyses. For mesopic testing, higher retinal sensitivity was associated with minimally less retest variance (slope estimate ± standard error –0.19 ± 0.030 dB^2/dB; $\chi2(1) = 36.57$, $P < 0.001$). Eccentricities from the fovea had neither as linear covariate nor as fixed effect influence on the retest variance ($\chi2(1) = 0.4871$, $P = 0.48$). In dark-adapted cyan testing (excluding data from 0 to 1°), neither the sensitivity ($\chi2(1) = 0.1734$, $P = 0.68$) nor the eccentricity ($\chi2(1) = 1.0439$, $P = 0.31$) had a significant effect on retest variance. Likewise, in dark-adapted red testing, both the test sensitivity (–0.38 ± 0.22 dB^2/dB; $\chi2(1) = 3.41$, $P = 0.065$) and eccentricity had no effect ($\chi2(1) = 0.0371$, $P = 0.85$) on retest variance.

Examination Time

The retest strategy (staircase starting close to the previously determined threshold) resulted in a significantly shorter test duration in the second mesopic test as compared to the first test (8.41 ± 0.8 vs. 8.00 ± 1.06 minutes; paired *t*-test; $P = 0.021$). In both dark-adapted cyan (10.25 ± 1.29 vs. 8.20 ± 1.00 minutes; $P < 0.001$) and dark-adapted red (10.14 ± 1.47 vs. 8.60 ± 1.17 minutes; $P < 0.001$) testing, the retest proceeded likewise faster (Table 5).

Discussion

These results indicate that the recently introduced S-MAIA device allows for a reliable, practicable, and robust assessment of mesopic, dark-adapted cyan and dark-adapted red sensitivity in patients with macular diseases. Detailed functional assessment with separation of rod- and cone-mediated retinal sensitivity and structure–function correlation is of importance for a better understanding of manifestation and progression in many retinal diseases. FCP may serve as an important tool as a functional endpoint in clinical trials in the future. It may allow for detection of early therapeutic effects, including threshold changes, for example, in relation to drusen regression or

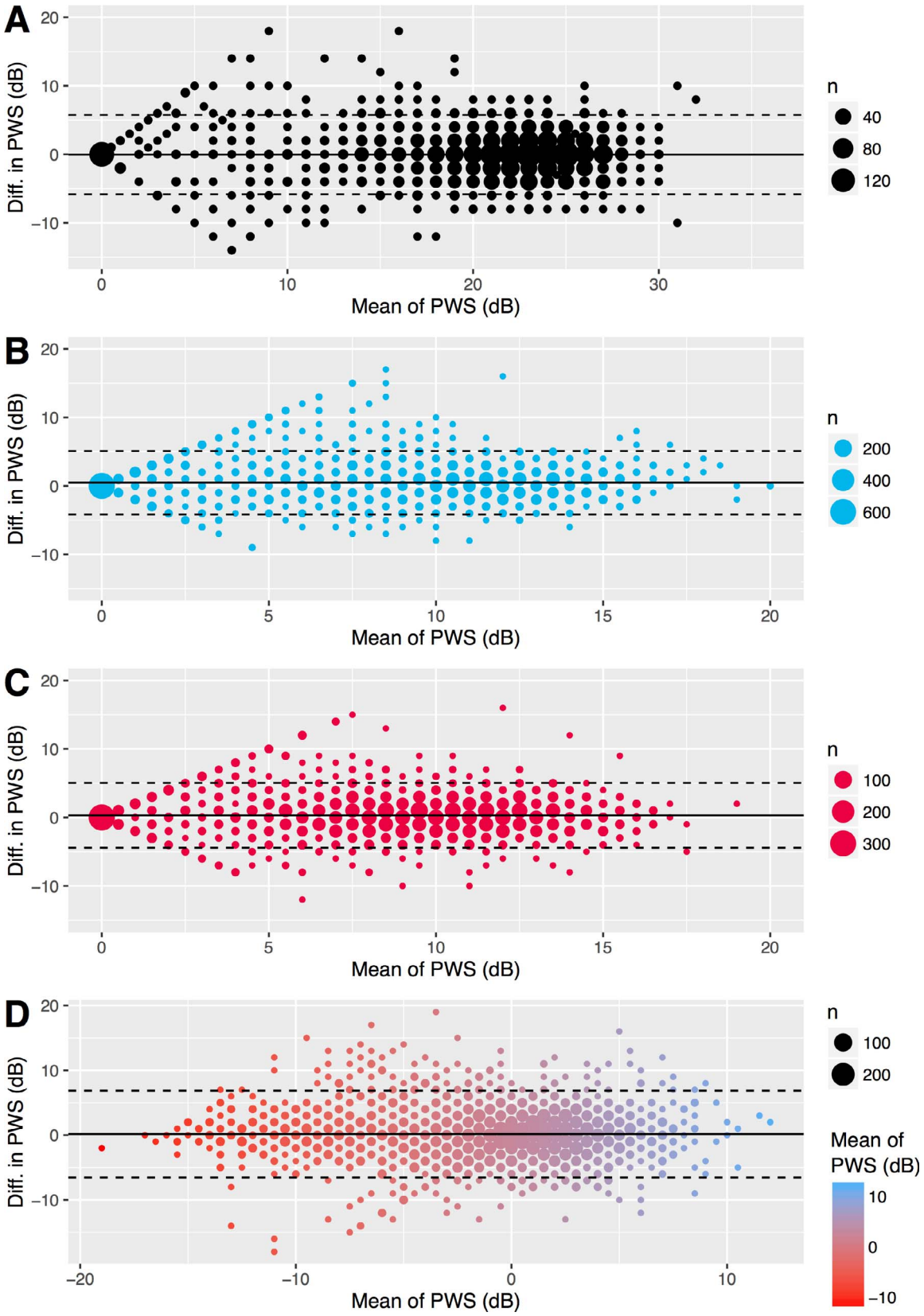

FIGURE 3. Bland-Altman plots for mesopic (**A**), dark-adapted cyan (**B**), dark-adapted red (**C**) testing, and the cyan–red difference (**D**).The x-axis shows the mean PWS for each pair of repeated tests. The PWS difference between two tests (first minus second test) is indicated on the y-axis. The overall mean difference is illustrated by the *solid line*; the 95% limits of agreement are marked by the two *dashed lines*. The size of individual *circles* illustrates the count of overlapping data points. Please note instrument limitations in maximum and minimum stimulus intensity that manifest as converging point spread at both ends. For mesopic testing, no convergence is observable at the upper end of the dynamic range, indicating that no ceiling effect is present. For dark-adapted cyan and red testing, the ceiling effect is also less prominent as compared to the floor effect.

Investigative Ophthalmology & Visual Science

Table 5. Differences in Test Duration

Type of Testing	Duration of First Test in Minutes, Mean ± SD	Duration of Second Test in Minutes, Mean ± SD	Paired *t*-Test
Mesopic*	8.41 ± 0.81	8.00 ± 1.06	$P = 0.021$
Scotopic cyan*	10.25 ± 1.29	8.2 ± 1	$P < 0.001$
Scotopic red*	10.14 ± 1.47	8.6 ± 1.17	$P < 0.001$

* The retest strategy with regard to the initial stimulus intensity was different for the mesopic as compared to the scotopic tests. The mesopic retest started at each location with an intensity of 2 dB above (brighter) the previously determined threshold. The scotopic retests started with the intensity of the previously determined threshold.

within the junctional zone of geographic atrophy. In this context, the current study provides relevant data to differentiate between test variability and true disease-related threshold changes that would be a prerequisite for a meaningful interpretation of test results in future studies. Of note, data from recent interventional studies in inherited retinal diseases have shown that changes in visual function were not associated with detectable changes in structural outcome measures, highlighting the importance of refined functional tests such as FCP.[14,19,35]

Based on the key findings, we propose that PWS changes greater than 5.79 dB for mesopic, 4.72 dB for dark-adapted cyan, and 4.77 dB for dark-adapted red testing would exceed test variability and would be indicative of disease progression using the S-MAIA. In this context, it also appears important to note that no obvious training effects were observed. Further, we provide data—given the current dynamic range of test stimuli—demonstrating that the device allows for detection of five DS for mesopic function, while both dark-adapted cyan (with exclusion of data within 1°) and red had four discernible steps that would permit a robust assessment of functional difference. Since test variability was not dependent on sensitivity (with the exception of mesopic testing), fixation, or age, these results appear to be applicable to a wide range of patients, including patients with extrafoveal or unstable fixation. In addition, the results are comparable to the test-rest variability for the mesopic MAIA with a CoR of 3.74 dB and the MP-1 (Nidek Technologies, Padua, Italy) with CoRs of 4.21 dB for red-on-red testing and 5.56 dB for achromatic testing.[18,19,36,37] Further, it was demonstrated previously for the S-MAIA device that in normal eyes the cyan (505 nm) stimulus reflects outside of the fovea largely rod function based on the (1) threshold distribution, (2) reaction time, and (3) the sensitivity to light adaptation.[23] However, it must be noted that the observed difference between the cyan-red differences (0° vs. 7°, 11 dB) of the device in normal eyes was lower than the predicted difference (approximately 2 log units) for isolated cone mediation at 0° and isolated rod mediation at 7° based on the CIE luminosity functions.[23,26,31] This is most likely a result of the limited dynamic range for the cyan stimulus (floor effects). The maximum measurable cyan-red difference with the current version is in the range of −11 dB impeding the quantification of severe rod dysfunction.[23,26,31] Nevertheless, the cyan-red sensitivity difference appears to be an interesting candidate as functional biomarker—especially for retinal diseases that affect either predominantly the cone or predominantly the rod pathway.[5–16]

While some studies have described the retest reliability for mesopic or dark-adapted FCP in terms of the CoR (assuming homoscedasticity of retest variability), to the best of our knowledge none have examined the EDR or number of DS so far.[18,19,23,36] The floor of the EDR for standard automated perimetry in the setting of glaucoma has been described directly or indirectly using several inherently different approaches. Yet, Wall et al.[1] (using the here applied definition for floor of the EDR), Gardiner et al.[3] (examining frequency-of-seeing curves in dependence of retinal sensitivity), and Saunders et al.[4] (analyzing the residuals of regression analysis of longitudinal clinical data) reported consistently a pronounced increase in test variability in the sensitivity range of 15–20 dB (i.e., nonstationary measurement variability) for standard automated perimetry.[1–4] In contrast, analysis of our data (Fig. 5) showed (1) no significant widening between the fifth and 95th retest percentile towards the lower limit of the EDR for all three types of testing, (2) floors of the EDRs close to 0 dB, and (3) high numbers of DS in relation to the respective EDR, especially when compared to the previously reported 4 DS within a log 3 range for standard automated perimetry.[1] Taken together, this indicated that retest reliability was almost independent of retinal sensitivity in our study (homoscedastic) as confirmed by the mixed-models analysis (with the exception of a slight effect of sensitivity on retest reliability in mesopic testing). This is likely attributable to the efficiency of the fundus tracking in FCP that ensures accurate stimulus placement. Indeed, the fixation stability had no (or only a weak) effect on retest reliability (considering both the *P* value and the effect estimates). The fact that retinal sensitivity had no (or only a weak) effect on the retest reliability is of central importance, since testing of transitional zones and zones at risk for disease progression (which are commonly associated with reduced sensitivity) is of special interest in the setting of clinical studies.[19,36,37]

Nevertheless, direct comparison of test variability with standard automated perimetry and glaucoma patients must be regarded with caution due to differences in background luminance, stimulus color, and test locations. Ganglion cell response saturation for bright stimuli in photopic conditions has been suggested to contribute to the test variability in the setting of glaucoma.[3,38] With the here used mesopic conditions, this potential source of retest variability would not be observable. Further, it is still unclear to what extent test variability in glaucoma is attributable to inaccurate stimulus placement within sparse, irregular receptive fields using standard automated perimetry versus "real" short-term threshold changes.[1,29,37–46] Moreover, Wu et al.[39] observed (using the MAIA) that steep local sensitivity gradients are associated with an increase in retest variability, suggesting that even with fundus tracking incorrect sampling of neighboring regions might occur to some extent.

Several limitations of the study herein must be considered. Firstly, the rest between the first and second dark-adapted cyan test was short, which might have caused the observed systematic elevations of the threshold at eccentricities of 0 to 1°. Since this region is rod free, the measured thresholds at 0 to 1° most likely represent M-cone (and possibly L-cone) function. The spatial proximity to the fixation target suggests that light adaptation might be responsible for the threshold shift.[23] To estimate the retest reliability at 0 to 1° accurately, a second complete period of dark adaptation appears to be necessary. Secondly, the radial test grid does not allow for feasible analysis of the effect of local slope on retest variability due to unequal spacing between the test points. Thirdly, the difference in retest reliability between mesopic and dark-adapted testing is likely attributable to the preset (nonalterable) staircase procedures (4-2 dB for mesopic, 2-1 dB for dark-adapted testing).[29] Adding another reversal (i.e., 4-2-1 dB staircase) would presumably improve the retest reliability for mesopic testing, however, increasing test duration and possibly fatigue. Further, dark adaptometry studies suggest that the difference between the rod/cone break and the lower rod plateau

amounts to 2.5–3.0 log units for blue stimuli.[27,47] Since no relevant ceiling effects could be observed, the preset ceiling of the physical dynamic range appears justified. However, the lower end of the dynamic range should be increased (i.e., allowing brighter stimuli) for dark-adapted cyan to cover the complete range between the rod/cone break and the lower rod plateau (log 3 range). As a further consequence, the reported retest variability was potentially slightly underestimated due to floor effects.

In summary, the current study reports a systematic assessment of the EDR and retest reliability of a recently introduced combined mesopic and two-color dark-adapted fundus-controlled ("micro-") perimetry device. For all three types of testing, there was no relevant effect of fixation stability, retinal sensitivity, or eccentricity on retest reliability. In conjunction, this constitutes the basis for accurate structure–function correlation and for the application of mesopic and dark-adapted FCP as functional outcome measure in clinical studies.

Acknowledgments

The authors thank Carlo Pellizzari and Marco U. Morales (CenterVue SpA, Padova, Italy) for technical support.

Supported by the BONFOR GEROK Program of the Faculty of Medicine, University of Bonn, Grant O-137.0022 (MP), Grant O-137.0020 (ML), and Grant O-137.0023 (PLM) and by the German Research Foundation Grant 5323/5-1 (WMH) and Grant 658/4-1 and 658/4-2 (MF). 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 this research.

Disclosure: **M. Pfau**, Heidelberg Engineering (F), Optos (F), Zeiss (F), CenterVue (F); **M. Lindner**, Heidelberg Engineering (F), Optos (F), Zeiss (F), CenterVue (F), Genentech (F); **P.L. Müller**, Heidelberg Engineering (F), Optos (F), Zeiss (F), CenterVue (F); **J. Birtel**, Heidelberg Engineering (F), Optos (F), Zeiss (F), CenterVue (F); **R.P. Finger**, Heidelberg Engineering (F), Optos (F), Zeiss (F), CenterVue (F); **W.M. Harmening**, Heidelberg Engineering (F), Optos (F), Zeiss (F), CenterVue (F); **M. Fleckenstein**, Heidelberg Engineering (F, R), Optos (F), Zeiss (F), CenterVue (F); **F.G. Holz**, Heidelberg Engineering (C, F, R), Optos (F), Zeiss (C, F), CenterVue (F); **S. Schmitz-Valckenberg**, Heidelberg Engineering (F, R), Optos (F), Zeiss (F), CenterVue (F)

References

1. Wall M, Woodward KR, Doyle CK, Zamba G. The effective dynamic ranges of standard automated perimetry sizes III and V and motion and matrix perimetry. *Arch Ophthalmol*. 2010; 128:570–576.
2. Russell RA, Crabb DP, Malik R, Garway-Heath DF. The relationship between variability and sensitivity in large-scale longitudinal visual field data. *Invest Ophthalmol Vis Sci*. 2012; 53:5985–5990.
3. Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. *Ophthalmology*. 2014;121:1359–1369.
4. Saunders LJ, Russell RA, Crabb DP. Measurement precision in a series of visual fields acquired by the standard and fast versions of the Swedish interactive thresholding algorithm: Analysis of large-scale data from clinics. *JAMA Ophthalmol*. 2015;133:74–80.
5. Gunkel RD. Retinal profiles: a psychophysical test of rod and cone sensitivity. *Arch Ophthalmol*. 1967;77:22–25.
6. Massof RW, Finkelstein D. Rod sensitivity relative to cone sensitivity in retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 1979;18:263–272.
7. Arden GB, Carter RM, Hogg CR, et al. Rod and cone activity in patients with dominantly inherited retinitis pigmentosa: comparisons between psychophysical and electroretinographic measurements. *Br J Ophthalmol*. 1983;67:405–418.
8. Lyness AL, Ernst W, Quinlan MP, et al. A clinical, psychophysical, and electroretinographic survey of patients with autosomal dominant retinitis pigmentosa. *Br J Ophthalmol*. 1985; 69:326–339.
9. Jacobson SG, Voigt WJ, Parel J-M, et al. Automated light- and dark-adapted perimetry for evaluating retinitis pigmentosa. *Ophthalmology*. 1986;93:1604–1611.
10. Jacobson SG, Kemp CM, Sung CH, Nathans J. Retinal function and rhodopsin levels in autosomal dominant retinitis pigmentosa with rhodopsin mutations. *Am J Ophthalmol*. 1991;112: 256–271.
11. Jacobson SG, Cideciyan AV, Iannaccone A, et al. Disease expression of RP1 mutations causing autosomal dominant retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 2000;41: 1898–1908.
12. Birch DG, Herman WK, DeFaller JM, Disbrow DT. The relationship between rod perimetric thresholds and full-field rod ERGs in retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 1987;28:954–965.
13. Fraser RG, Tan R, Ayton LN, Caruso E, Guymer RH, Luu CD. Assessment of retinotopic rod photoreceptor function using a dark-adapted chromatic perimeter in intermediate age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2016;57: 5436–5442.
14. Cideciyan AV, Aleman TS, Boye SL, et al. Human gene therapy for RPE65 isomerase deficiency activates the retinoid cycle of vision but with slow rod kinetics. *Proc Natl Acad Sci U S A*. 2008;105:15112–15117.
15. Bainbridge J, Smith A. Effect of gene therapy on visual function in Leber's congenital amaurosis. *N Engl J Med*. 2008; 358:2231–2239.
16. McGuigan DB, Roman AJ, Cideciyan AV, et al. Automated light- and dark-adapted perimetry for evaluating retinitis pigmentosa: filling a need to accommodate multicenter clinical trials. *Invest Ophthalmol Vis Sci*. 2016;57:3118–3128.
17. Cheung LK, Eaton A. Age-related macular degeneration. *Pharmacotherapy*. 2013;33:838–855.
18. Chen FK, Patel PJ, Xing W, et al. Test-retest variability of microperimetry using the Nidek MP1 in patients with macular disease. *Invest Ophthalmol Vis Sci*. 2009;50:3464–3472.
19. Cideciyan AV, Swider M, Aleman TS, et al. Macular function in macular degenerations: Repeatability of microperimetry as a potential outcome measure for ABCA4-associated retinopathy trials. *Invest Ophthalmol Vis Sci*. 2012;53:841–852.
20. Rohrschneider K, Bültmann S, Springer C. Use of fundus perimetry (microperimetry) to quantify macular sensitivity. *Prog Retin Eye Res*. 2008;27:536–548.
21. Springer C, Rohrschneider K. Statische Fundus perimetrie bei Probanden. *Ophthalmologe*. 2006;103:214–220.
22. Sergouniotis PI, McKibbin M, Robson AG, et al. Disease expression in autosomal recessive retinal dystrophy associated with mutations in the *DRAM2* gene. *Invest Ophthalmol Vis Sci*. 2015;56:8083–8090.
23. 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;237:42–54.
24. Vujosevic S, Pucci P, Casciano M, et al. Long-term longitudinal modifications in mesopic microperimetry in early and intermediate age-related macular degeneration. *Graefes Arch Clin Exp Ophthalmol*. 2016:1–9.
25. Wu Z, Cunefare D, Chiu E, et al. Longitudinal associations between microstructural changes and microperimetry in the

early stages of age-related macular degeneration. *Invest Opthalmol Vis Sci*. 2016;57:3714–3722.

26. Wald G. Human vision and the spectrum. *Science*. 1945;101: 653–658.
27. Zeavin BH, Wald G. Rod and cone vision in retinitis pigmentosa. *Am J Ophthalmol*. 1956;42:253–269.
28. Midena E, ed. *Microperimetry and Multimodal Retinal Imaging*. Berlin: Springer; 2014.
29. Johnson CA, Chauhan BC, Shapiro LR. Properties of staircase procedures for estimating thresholds in automated perimetry. *Invest Ophthalmol Vis Sci*. 1992;33:2966–2974.
30. von Békésy G. A new audiometer. *Acta Otolaryngol*. 1947;35: 411–422.
31. Crawford BH. The scotopic visibility function. *Proc Phys Soc B*. 1948;62:321–334.
32. Bengtsson B. Reliability of computerized perimetric threshold tests as assessed by reliability indices and threshold reproducibility in patients with suspect and manifest glaucoma. *Acta Ophthalmol Scand*. 2000;78:519–522.
33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (London, England)*. 1986;1:307–310.
34. Crossland MD, Dunbar HMP, Rubin GS. Fixation stability measurement using the MP1 microperimeter. *Retina*. 2009; 29:651–656.
35. Cideciyan AV, Hauswirth WW, Aleman TS, et al. Vision 1 year after gene therapy for Leber's congenital amaurosis. *N Engl J Med*. 2009;361:725–727.
36. Wu Z, Ayton LN, Guymer RH, Luu CD. Intrasession test–retest variability of microperimetry in age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2013;54:7378–7385.
37. Wu Z, Jung CJ, Ayton LN, Luu CD, Guymer RH. Test-retest repeatability of microperimetry at the border of deep scotomas. *Invest Ophthalmol Vis Sci*. 2015;56:2606–2611.
38. Anderson AJ, McKendrick AM, Turpin A. Do intense perimetric stimuli saturate the healthy visual system? Visual system saturation and perimetry. *Invest Ophthalmol Vis Sci*. 2016;57:6397–6404.
39. Wu Z, McKendrick AM, Hadoux X, et al. Test-retest variability of fundus-tracked perimetry at the peripapillary region in open angle glaucoma. *Invest Ophthalmol Vis Sci*. 2016;57: 3619–3625.
40. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol*. 1989;108:130–135.
41. Iii S, Perimetry M, Wall M, Woodward KR, Doyle CK, Artes PH. Repeatability of automated perimetry: a comparison between standard automated perimetry with stimulus. *Invest Ophthalmol Vis Sci*. 2009;50:974–979.
42. Henson DB, Evans J, Chauhan BC, Lane C. Influence of fixation accuracy on threshold variability in patients with open angle glaucoma. *Invest Ophthalmol Vis Sci*. 1996;37: 444–450.
43. Demirel S, Vingrys AJ. Eye movements during perimetry and the effect that fixational instability has on perimetric outcomes. *J Glaucoma*. 1994;3:28–35.
44. Wyatt HJ, Dul MW, Swanson WH. Variability of visual field measurements is correlated with the gradient of visual sensitivity. *Vision Res*. 2008;47:925–936.
45. Wyatt HJ. Automated perimetry: using gaze-direction data to improve the estimate of scotoma edges. *Invest Ophthalmol Vis Sci*. 2011;52:5818–5823.
46. Springer C, Bültmann S, Völcker HE, Rohrschneider K. Fundus perimetry with the Micro Perimeter 1 in normal individuals: comparison with conventional threshold perimetry. *Ophthalmology*. 2005;112:848–854.
47. Sloan LL. The threshold gradients of the rods and the cones: in the dark-adapted and in the partially light-adapted eye. *Am J Ophthalmol*. 1950;33:1077–1089.

Investigative Ophthalmology & Visual Science