# How "Micro" Is Microperimetry? Characterizing the Effect of Fundus Tracking on the Psychometric Function

Theresa Lipsky[1,2] Clara Ehrenzeller[1,3] Georg Ansari[1] Kristina Pfau[1,2] Wolf Harmening[2] Zhichao Wu[4,5] and Maximilian Pfau[1,2]

[1]Department of Ophthalmology, University Hospital Basel, Basel, Switzerland
[2]Department of Ophthalmology, University Hospital Bonn, Bonn, Germany
[3]Nuffield Laboratory of Ophthalmology, Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom
[4]Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia
[5]Department of Ophthalmology and Surgery, The University of Melbourne, Melbourne, Australia

Correspondence: Maximilian Pfau, Department of Ophthalmology, University Hospital Bonn, Venusberg-Campus 1, Bonn 53127, Germany; maximilian.pfau@ukbonn.de.

TL and CE contributed equally to this work.

**Received:** March 25, 2026
**Accepted:** June 23, 2026
**Published:** August 24, 2026

Citation: Lipsky T, Ehrenzeller C, Ansari G, et al. How "Micro" is microperimetry? Characterizing the effect of fundus tracking on the psychometric function. *Invest Ophthalmol Vis Sci.* 2026;67(10):55. https://doi.org/10.1167/iovs.67.10.55

**Purpose.** The purpose of this study was to quantify whether fundus tracking in microperimetry improves psychometric parameter estimation (as an in vivo demonstration of improved stimulus-delivery precision), and to derive a psychometrically grounded criterion intensity for suprathreshold (defect-mapping) microperimetry.

**Methods.** Twenty-five healthy volunteers underwent Macular Integrity Assessment (MAIA)2 microperimetry at five loci referenced to Bruch's membrane opening (BMO): three outside and two inside the blind spot. At each location, frequency-of-seeing (FoS) functions were measured with a method of constant stimuli in four blocks (2 had tracking on and 2 had tracking off). FoS data were fit with a cumulative-Gaussian psychometric function estimating sensitivity, spread, guess, and lapse rate. Tracking effects on m, denoting the sensitivity corresponding to the midpoint of the psychometric function, and s, the spread parameter of the underlying Gaussian, were tested per locus using mixed-effects models. The optimal criterion intensity to separate "non-seeing" from "seeing" loci was derived from FoS fits using posterior predictive simulations.

**Results.** Tracking had little effect on sensitivity estimates at loci outside the blind spot, but produced lower sensitivity estimates within the blind spot — indicating fewer spurious detections at non-seeing loci (posterior median difference, 95% credible intervals [CrIs] of −1.78 decibel [dB] 95% CrI = −2.58 to −0.94 at locus 4, and −0.75 dB, 95% CrI = −1.71 to 0.15 at locus 5). Tracking was also associated with a smaller spread (steeper FoS curves) at loci 1 to 3, with posterior median differences of −0.13 dB, 95% CrIs = −0.28 to 0.02, −0.27 dB, 95% CrI = −0.43 to −0.12, and −0.25 dB, 95% CrI = −0.42 to −0.07, respectively. Without tracking, false-positive responses were more frequent when fixation shifts displaced stimuli toward the "seeing" retina. Simulation-based analysis identified 13 dB as the nominally optimal criterion intensity for suprathreshold microperimetry (Youden index = 0.74, 95% CrI = 0.70–0.77); a 10 dB criterion yielded comparable discrimination (Youden index = 0.72, 95% CrI = 0.69–0.75).

**Conclusions.** Even in healthy volunteers with stable fixation, fundus tracking measurably lowered apparent sensitivity at "non-seeing" loci (i.e., reduced spurious responses inside the blind spot) and steepened FoS curves in the "seeing" retina. A criterion intensity in the range of 10 to 16 dB is a defensible choice for separating "seeing" and "non-seeing" retina in suprathreshold (defect-mapping) perimetry paradigms.

Keywords: microperimetry, frequency-of-seeing (FoS) curves, fundus tracking, defect-mapping

Microperimetry (fundus-controlled perimetry) has matured from a research technique into a standard clinical trial tool for macular visual field testing.[1] By visualizing the fundus with a scanning laser ophthalmoscope or a fundus camera in real time, stimuli can be presented at predefined retinal locations, even in patients with unstable fixation. In principle, simultaneous imaging of the retina at the moment of stimulus delivery permits verification of each stimulus's retinal location to within the system's resolution; in clinical practice, however, the implementation of this link between fundus tracking and stimulus delivery is not directly accessible to the operator.[1] Especially in geographic atrophy (GA) secondary to age-related macular degeneration (AMD), microperimetry is becoming increasingly important as it enables tracking disease progression in eyes showing no change in best-corrected

visual acuity due to foveal sparing or early foveal involvement.[2,3]

To date, it is unclear how accurately microperimetry devices can track and present stimuli at the intended retinal location, given the complexity of fixational eye movements, including microsaccades, drift, and torsional motion.[4–6] Clinical observations provide face validity, because test points within areas of GA, especially centrally located within lesions, typically show no response, whereas points just outside the GA boundary exhibit a steep functional increase within the outer approximately 500 µm junctional zone. This mirrors the pattern of photoreceptor loss seen in histopathology.[3,7,8] Moreover, localized optical coherence tomography data can be used to predict function with high accuracy, indicating that stimuli are, on average, presented where intended.[9–11]

However, these demonstrations of face validity rely on aggregate data across stimulus presentations and do not directly address a key psychophysical question: how precisely is each individual stimulus positioned at the intended retinal location? This question becomes critical in junctional zones with steep sensitivity gradients, where even sub-degree placement errors or small fixation excursions can flip a response from "seen" to "not seen."[12] Although stimulus-placement accuracy has been examined previously for a modified MP1 microperimeter, those findings are difficult to interpret in the present context owing to the device-specific modifications used and the obsolescence of the MP1 platform.[13]

Although GA provides a clinically relevant setting for understanding the effect of tracking on the psychometric function, it is not an ideal experimental model for isolating stimulus-delivery precision. Areas of atrophy do not uniformly represent absolute scotomata, but can still yield residual detections, particularly at loci with a residual outer nuclear layer.[14] The physiologic blind spot provides a useful complementary model for studying stimulus-delivery precision. The optic nerve head constitutes an anatomically defined deep scotoma with a comparatively sharp boundary and, unlike GA, is not confounded by variable residual photoreceptor structure.[15]

Knowledge of the psychometric function close to scotomata boundaries is particularly important for suprathreshold testing paradigms that aim to separate "seeing" from "non-seeing" retina with a single presentation.[10,15] Ideally, the fixed criterion intensity for suprathreshold perimetry should be justifiable based on both clinical heuristics,[16] as well as the psychometric function.

Fixed-stimulus testing, analyzed through frequency-of-seeing (FoS) modeling, provides the natural framework for defining psychometric cutoffs between "seeing" and "non-seeing" retina.[17] The shape of the FoS curve has been extensively characterized in photopic standard automated perimetry.[18–22] Recent mesopic FoS data have been reported in AMD, but were obtained without active fundus tracking, leaving the influence of stimulus-placement variability on the psychometric function unresolved; mesopic FoS data are also absent for the Macular Integrity Assessment (MAIA)2 device itself.[17]

These gaps motivate a closer examination of how "micro" microperimetry truly is, using prospectively acquired FoS data with the MAIA device, with and without tracking, in healthy volunteers at three loci outside the physiologic blind spot and two loci within it, thereby leveraging the optic nerve head as a natural scotoma. Specifically, we asked: (i) Does fundus tracking provide stimulus-placement precision beyond natural fixation stability in young observers, and is this reflected in measurable changes in FoS sensitivity or in the spread of the psychometric function? (ii) Does fixation-related stimulus misplacement in the absence of tracking systematically give rise to false-positive and false-negative responses at "non-seeing" and "seeing" loci, respectively? (iii) Do these psychometric data support the current criterion intensity for defect-mapping microperimetry?

## Methods

### Participants

The study was approved by the Ethics Committee for Northwestern and Central Switzerland (EKNZ) and conducted in accordance with the Declaration of Helsinki. All participants received written and verbal information about the procedures and provided written informed consent prior to participation.

Eligible participants were ≥18 years old and had no history of ocular surgery that, in the investigator's judgment, could affect visual function outcomes. Surgeries permitted for inclusion were uncomplicated cataract extraction, YAG laser capsulotomy, and laser retinopexy.

### Clinical Examination

Participants underwent autorefraction, followed by best-corrected visual acuity testing. Retinal structure was assessed with spectral-domain optical coherence tomography (SD-OCT; Heidelberg Spectralis OCT2, Heidelberg Engineering, Heidelberg, Germany), including (1) a macular volume scan (30 degrees × 25 degrees, 121 B-scans, and automatic real-time averaging = 25) and (2) an optic nerve head scan using the device's preset Bruch's membrane opening (BMO) protocol.

### Test Grid Design

The experiment was designed to use the optic nerve head as a natural deep scotoma with a sharp boundary.[15] Using the OCT volume, an experienced grader identified the central horizontal B-scan passing through the optic disc (oriented along the disc's horizontal meridian) and manually labeled the temporal BMO on that B-scan and on the co-acquired Spectralis infrared (IR) reference image.

A MAIA microperimetry examination was then initiated via the Open Perimetry Interface (OPI).[23,24] At test start, OPI provides a "MAIA template IR" image, which served as the retinal tracking reference for the entire examination. The labeled Spectralis IR image was registered to the MAIA template IR image in ImageJ software using a landmark-based transform (Landmark Correspondences). A custom ImageJ software[25] plugin then generated MAIA stimulus coordinates for five test locations positioned relative to the BMO border: three locations outside the BMO (+0.25 degrees, +0.75 degrees, and +1.25 degrees) and two locations inside the BMO (−0.25 degrees and −0.75 degrees). Offsets were chosen so that the Goldmann III target (0.43 degrees diameter and 0.215 degrees radius) did not fall exactly on the anatomic boundary.

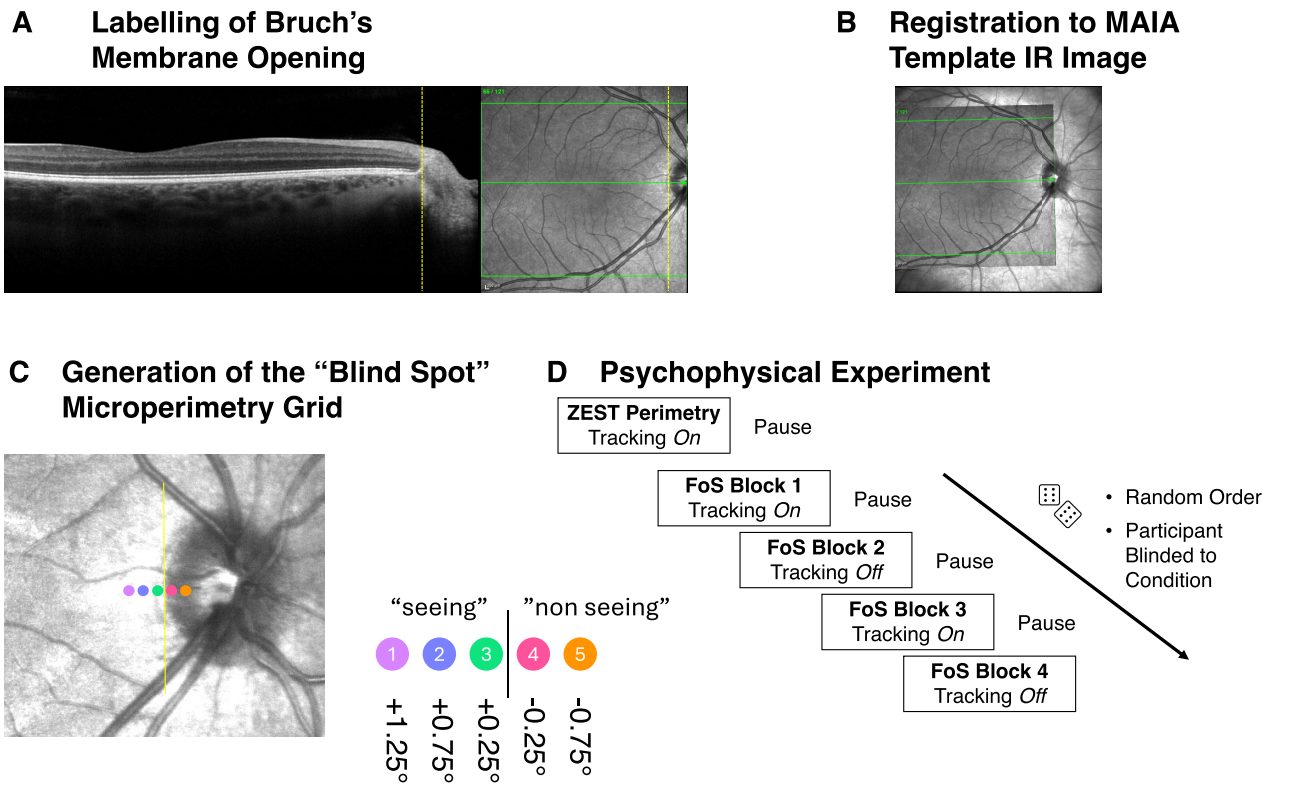


**Figure 1. Experimental Design.** The optic disc was used as a natural scotoma. (**A**) First, an optical coherence tomography scan was used to identify the temporal Bruch's membrane opening of the optic disc, and the label was transferred to the co-acquired infrared reflectance (IR) image (*yellow dasbed line*). **(B)** The IR image and label were then registered to the MAIA device's IR image from the start of the examination. (**C**) The testing grid comprised five locations with three locations outside of Bruch's membrane opening (+1.25 degrees, +0.75 degrees, and +0.25 degrees) and two locations within Bruch's membrane opening (−0.25 degrees and −0.75 degrees). The marker size approximates the Goldmann III stimulus size. (**D**) The actual psychophysical experiment consisted of an initial perimetry test to define the range of sensitivity for each locus, followed by four blocks with frequency-of-seeing (FoS) testing. These were ordered randomly, and the participants were blinded to the condition (tracking enabled or disabled).

## Psychophysical Experiment

Sensitivity at each location was first estimated using Zippy Estimation by Sequential Testing (ZEST).[26] This estimate defined the stimulus range for FoS measurements: FoS experiments spanned a total of seven intensity levels centered on the ZEST estimate and spanning ±3 decibels (dB; i.e., 1 dB steps). If the targeted range included steps below 0 dB, the FoS range was truncated to sensitivities of 0 dB or more.

FoS functions were then measured using a method of constant stimuli with a yes/no detection paradigm — participants pressed a response button whenever a stimulus was perceived — in a total of four blocks: two with retinal tracking enabled (blocks 1 and 3) and two with tracking disabled (blocks 2 and 4). The block order was randomized, and participants were masked to the tracking condition (Fig. 1). For each location, 15 presentations per intensity set were delivered per block (i.e., 15 × 7 = 105 presentations per location and block; slightly fewer for locations within BMO due to the floor of the sensitivity range at 0 dB).

## Statistical Analysis

Participant and trial-level data are publicly available (Zenodo repository: 10.5281/zenodo.21627620). All analyses were performed in R software. The FoS data were modeled for each subject, locus, and condition using a cumulative Gaussian psychometric function. The probability of reporting a stimulus as seen was expressed as:

$$P(\text{seen}) = \text{guess} + (1 - \text{guess} - \text{lapse})\ \Phi\left(-\frac{dB - m}{s}\right) \quad (1)$$

Equation 1, where m denotes the sensitivity (in dB) corresponding to the midpoint of the function, s is the spread parameter of the underlying Gaussian (a smaller s corresponds to a steeper FoS curve), guess represents the lower asymptote (false-positive rate), lapse represents the upper asymptote (false-negative rate), and Φ is the standard normal cumulative distribution function. The negative sign in −(dB − m)/s reflects the dB convention used on the MAIA scale: dB denotes attenuation, so higher dB corresponds to a dimmer stimulus (i.e., one of lower physical intensity); the cumulative distribution is therefore reflected along the dB axis such that the probability of "seen" increases as dB decreases. Model fitting was performed within a Bayesian framework to allow direct propagation of uncertainty in parameter estimates (Fig. 2).[27]

Test–retest reliability was assessed separately for tracking-on and tracking-off conditions by comparing

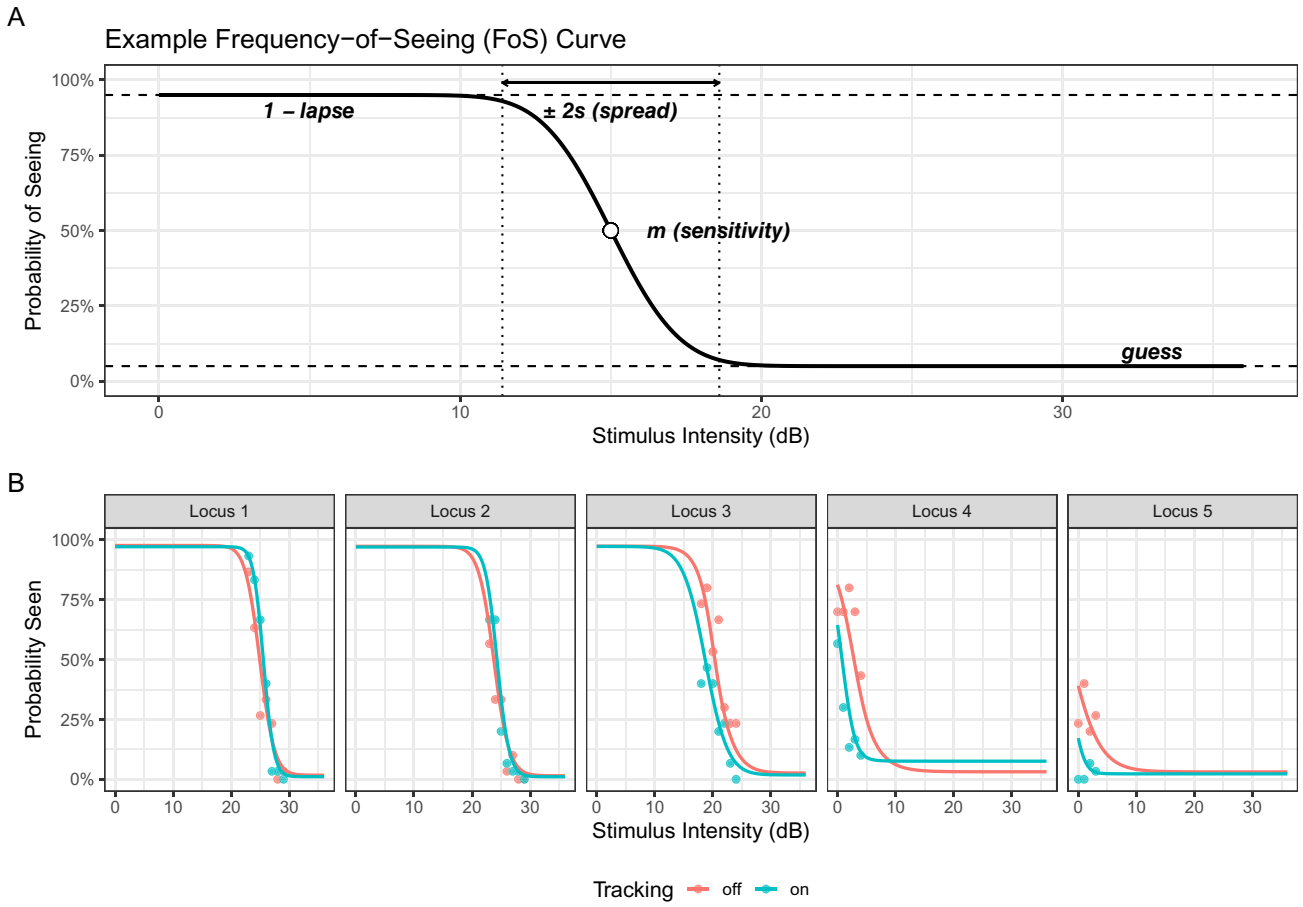


**Figure 2. Frequency-of-seeing (FoS) psychometric function and example individual data.** The *top panel* (**A**) shows a schematic frequency-of-seeing curve illustrating the main psychometric parameters. The probability of reporting a stimulus as seen is plotted as a function of stimulus intensity. The sensitivity (m) corresponds to the stimulus level at which the response probability reaches 50% after accounting for false-positive (guess) and false-negative (lapse) response rates. The spread parameter (s) describes the steepness of the transition from "non-seeing" to "seeing" around the sensitivity estimate. *Dashed horizontal lines* indicate the lower and upper asymptotes defined by guess and lapse rates. The *bottom panel* (**B**) shows example data from a single study participant showing fitted FoS curves at five retinal loci (*columns*). *Points* represent observed response probabilities at each stimulus intensity, and *solid lines* show the fitted psychometric functions. Data are shown separately for tracking-off (*red*) and tracking-on (*blue*) conditions. For loci 4 and 5 (targeting the "non-seeing" Bruch's membrane opening), tracking results in a reduced rate of responses as anticipated.

repeated blocks within each condition (tracking-on: FoS1 versus FoS3 and tracking-off: FoS2 versus FoS4). Reliability was evaluated for sensitivity, spread, guess, and lapse using Bland–Altman analyses implemented with SimplyAgree.[28]

The primary analysis of tracking effects on psychometric parameters was performed using a joint Bayesian multivariate model with sensitivity and spread as correlated outcomes. Sensitivity and spread were modeled jointly as Gaussian outcomes with fixed effects for tracking, locus, and their interaction, and with shared subject-level random intercepts and random slopes for tracking. Outcome-specific inverse-variance weights, defined as the inverse squared standard errors from the first-stage psychometric fits, were used to account for the differing precision of parameter estimates. Because different weights were used for sensitivity and spread, residual correlations between outcomes were not estimated; however, cross-outcome associations at the subject level were retained through shared random effects. Posterior summaries were reported as posterior medians, 95% credible intervals (CrIs), and probabilities of direction (pd). Population-level marginal means under tracking-off and locus-specific tracking effects (tracking-on minus tracking-off) were derived from posterior predictions.

As sensitivity analyses, sensitivity and spread were also modeled separately using analogous Bayesian mixed-effects models fitted across all loci. In addition, for direct comparison with more conventional approaches, frequentist linear mixed-effects models were fitted separately for each locus using lme4, with tracking as a fixed effect and subject-specific random intercepts and slopes. Further, we tested whether block administration order influenced the estimated sensitivity or spread parameters using Bayesian linear mixed-effects models with block position as a fixed effect and subject-specific random intercepts.

To examine fixation-related sources of likely false-positive and false-negative responses, fixation measurements were mapped to the timing of stimulus presentation. The internal fundus-tracking rate of the MAIA2 is 25 hertz (Hz; per manufacturer specifications); however, the OPI exposes the fixation log at a substantially lower cadence

(one sample per 642 ms), and a per-stimulus fixation readout is not directly available via OPI on the MAIA2. We therefore used a fuzzy time match between each stimulus presentation and the nearest OPI fixation sample. The median absolute time offset between matched pairs was 157 ms (maximum = 341 ms). These constraints place a floor on the spatial precision with which we can attribute individual false responses to specific fixational events, but should bias against — rather than in favor of — detecting fixation-related effects.

Note that fixation data were also recorded under the tracking-on condition. These data reflect where the participant was fixating during the examination. Stimulus onset under tracking-on should occur at the intended retinal locus (the stimulus does not, though, move during the 200-ms presentation window).

Bayesian logistic mixed-effects models were fitted jointly across loci. For likely false-positive responses, the model included negative horizontal displacement (corresponding to displacement of the stimulus toward the "seeing" retina), tracking condition, locus, and all interactions among these terms, with subject-specific random intercepts. For likely false-negative responses, an analogous model was fitted using positive horizontal displacement (corresponding to displacement of the stimulus toward the optic nerve head), again including tracking, locus, and their interactions, with subject-specific random intercepts. Locus-specific odds ratios, 95% CrIs, and probabilities of direction were derived from posterior linear predictors. As sensitivity analyses, frequentist per-locus logistic mixed-effects models were also fitted using glmer.

Optimal suprathreshold stimulus intensity for defect-mapping microperimetry was determined from operating characteristics derived from psychometric functions fitted under tracked conditions only, reflecting the clinically relevant examination setting. Posterior predictive simulations were used to propagate uncertainty arising from psychometric parameter estimation and binomial response variability at the trial level. True-positive and false-positive rates were estimated across the full range of stimulus intensities, and the criterion maximizing Youden's index — defined as sensitivity + specificity − 1, equivalent to (true-positive rate) − (false-positive rate), with a range of 0 (no discrimination) to 1 (perfect classification) — was identified, with uncertainty summarized using posterior CrIs. As a sensitivity analysis, the same procedure was repeated using only locus 1 as the "seeing" locus and locus 5 as the "non-seeing" locus, to minimize potential confounding from peripapillary atrophy.

## Results

### Demographics

A total of 25 healthy volunteers (median age [Q1, Q3]: 27.0 years [25.1, 30.1]; 12 women and 13 men) were enrolled. All participants exhibited normal ocular findings on slit-lamp examination and SD-OCT and had best-corrected visual acuity of 20/16 Snellen equivalent or better (median [Q1, Q3] of −0.1 logMAR [−0.1, −0.1]).

The dataset comprised 500 independently measured FoS functions.

The acquisition of a block took, on average, (median [Q1, Q3]) 21.3 minutes [19.7, 23.5] with tracking off and 23.3 minutes [20.9, 25.3] with tracking on.

### Locus-Specific Psychometric Parameters and Test–Retest Reliability

Absolute psychometric parameters differed systematically across loci, confirming that the experimental paradigm correctly captured the expected sensitivity gradient across the optic-disc border (see the Table). Under tracking-off conditions, sensitivity was highest at loci 1 to 3, with posterior medians of 25.26 dB (95% CrI = 23.87–26.64), 23.43 dB (95% CrIs = 22.06–24.83), and 20.01 dB (95% CrIs = 18.61–21.41), respectively. Sensitivity declined sharply at loci 4 and 5, with posterior medians of 6.28 dB (95% CrIs = 4.87–7.73) at locus 4 and 0.42 dB (95% CrIs = −1.03 to 1.90) at locus 5, approaching the measurement floor. Spread estimates likewise varied across loci. Under tracking-off conditions, posterior median spread estimates were 1.57 (95% CrIs = 1.39–1.74) at locus 1, 1.86 (95% CrIs = 1.69–2.03) at locus 2, 2.33 (95% CrIs = 2.15–2.52) at locus 3, 2.49 (95% CrIs = 2.29–2.69) at locus 4, and 1.94 (95% CrIs = 1.75–2.14) at locus 5.

Test–retest reliability was assessed separately for tracking-on and tracking-off conditions by comparing repeated blocks within each condition (blocks 1 vs. 3 and blocks 2 vs. 4, respectively). Bland–Altman analyses showed comparable repeatability with and without tracking across all psychometric parameters (Supplementary Fig. S1; Supplementary Table S1). Sensitivity showed small biases and limits of agreement of up to ±2.6 dB in both conditions, whereas spread estimates (s) showed minimal bias and limits of agreement of up to ±2 dB. Guess and lapse rates were highly stable, with near-zero biases and narrow limits of agreement irrespective of tracking status.

### Effect of Fundus Tracking on the Psychometric Function

Bayesian mixed-effects models were used to estimate the effect of retinal tracking on sensitivity and spread at each test locus (see the Table; Fig. 3). Across loci 1 to 3, located outside the blind spot, the effect of tracking on sensitivity was small and uncertain. Posterior median differences were −0.23 dB (95% CrI = −0.81 to 0.36, pd = 0.782) at locus 1, 0.20 dB (95% CrI = −0.39 to 0.81, pd = 0.75) at locus 2, and −0.28 dB (95% CrI = −0.95 to 0.40, pd = 0.783) at locus 3. Thus, for these loci, the posterior distributions were centered close to zero and the CrIs overlapped zero.

In contrast, at the two loci located within the blind spot, sensitivity estimates (in dB on the MAIA scale) were lower when tracking was enabled — indicating reduced spurious detections. At locus 4, the posterior median difference was −1.78 dB (95% CrI = −2.58 to −0.94, pd = 1), and at locus 5 it was −0.75 dB (95% CrI = −1.71 to 0.15, pd = 0.945). These results indicate strong directional evidence for a tracking-related reduction of apparent sensitivity at both blind-spot loci.

For the spread parameter s (a smaller s corresponds to a steeper FoS curve), tracking effects were more spatially heterogeneous. At loci 1 to 3, tracking was associated with a smaller s (i.e., steeper FoS), with posterior median differences of −0.13 (95% CrI = −0.28 to 0.02, pd = 0.953) at locus 1, −0.27 (95% CrI = −0.43 to −0.12, pd = 1) at locus 2, and −0.25 (95% CrI = −0.42 to −0.07, pd = 0.996) at locus 3. By contrast, within the blind spot, there was little evidence that tracking altered s. Posterior median differences were 0.03

**Table.** Estimates of the Joint Bayesian Model for Sensitivity and Spread Across Loci for Tracking On and Off

| | Sensitivity, dB | | | Spread, dB | | |
|---|---|---|---|---|---|---|
| Locus | Intercept (Tracking-Off) Posterior Median [95% CrI] | Effect of Tracking On Posterior Median [95% CrI] | *pd* | Intercept (Tracking-Off) Posterior Median [95% CrI] | Effect of Tracking On Posterior Median [95% CrI] | *pd* |
| 1 | 25.26 [23.87 to 26.64] | −0.23 [−0.81 to 0.36] | 0.782 | 1.57 [1.39 to 1.74] | −0.13 [−0.28 to 0.02] | 0.953 |
| 2 | 23.43 [22.06 to 24.83] | 0.20 [−0.39 to 0.81] | 0.75 | 1.86 [1.69 to 2.03] | **−0.27 [−0.43 to −0.12]** | **1** |
| 3 | 20.01 [18.61 to 21.41] | −0.28 [−0.95 to 0.40] | 0.783 | 2.33 [2.15 to 2.52] | **−0.25 [−0.42 to −0.07]** | **0.996** |
| 4 | 6.28 [4.87 to 7.73] | **−1.78 [−2.58 to −0.94]** | **1** | 2.49 [2.29 to 2.69] | 0.03 [−0.19 to 0.24] | 0.6 |
| 5 | 0.42 [−1.03 to 1.90] | −0.75 [−1.71 to 0.15] | 0.945 | 1.94 [1.75 to 2.14] | −0.18 [−0.37 to 0.02] | 0.96 |

CrI, credible interval; pd, probability of direction.

Estimates are reported as posterior medians with 95% CrIs. Values of pd ≥ 0.975 are highlighted in bold, indicating strong evidence for a directional effect.

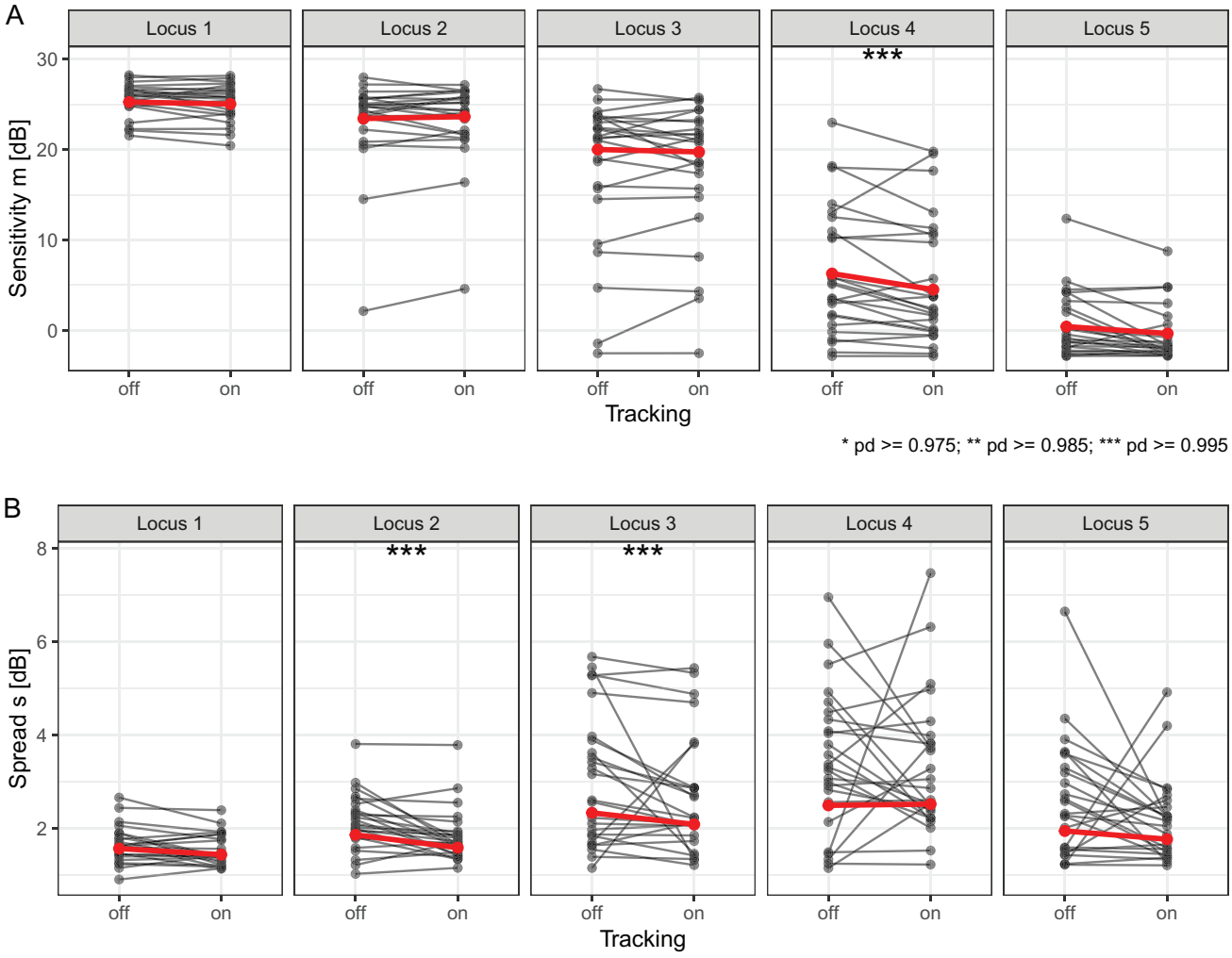


**Figure 3. Effect of tracking on the psychometric function.** Sensitivity m (*top row*) and the spread parameter s of the cumulative-Gaussian psychometric function (*bottom row*) are shown for the five retinal loci (*columns*). A smaller spread parameter s corresponds to a steeper FoS curve. *Gray points* and *connecting lines* represent within-subject averages for tracking-off and tracking-on conditions. *Red lines* indicate the fixed-effect estimates from Bayesian models. Across loci, tracking-related differences were generally small and comparable to test–retest variability. *Asterisks* above each locus mark posterior probabilities of direction (* pd ≥ 0.975; ** pd ≥ 0.985; and *** pd ≥ 0.995). Modest reductions in sensitivity with tracking were observed at loci closer to the blind spot; the spread parameter s decreased (FoS sharpened) with tracking at loci 1 to 3.

(95% CrI = −0.19 to 0.24, pd = 0.6) at locus 4 and −0.18 (95% CrI = −0.37 to 0.02, pd = 0.96) at locus 5, indicating no clear effect at locus 4 and only weak-to-moderate directional evidence at locus 5.

Consistent results – both for the effect of tracking on sensitivity and spread – were obtained across all model specifications, including separate Bayesian models for sensitivity and spread fitted across all loci as well as frequentist linear mixed-effects models fitted per locus (Supplementary Tables S2–S4).

Importantly, fatigue evaluated as block order (first versus second block) did not influence the estimated sensitivity

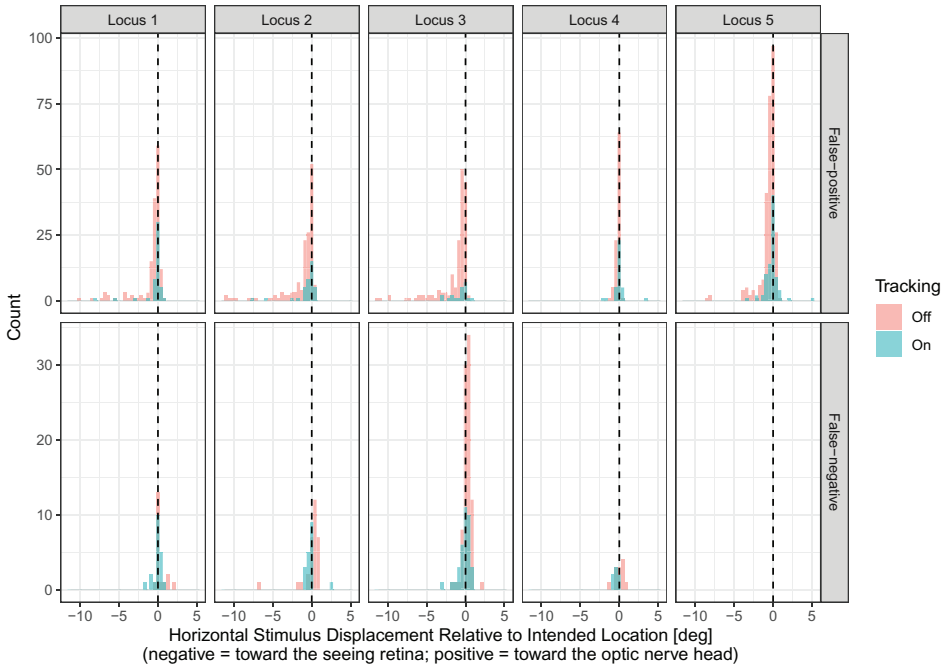


**Figure 4. Spatial distribution of fixation at the time of likely false-positive and false-negative responses.** Histograms show the horizontal stimulus displacement, evaluated at or near the time of stimulus presentation. Data are shown separately for five test loci (*columns*) and response type (*rows*: false-positive responses, *top*; and false-negative responses, *bottom*). False-positive responses were defined as responses with a low predicted probability of being seen (<0.1), whereas false-negative responses were defined as missed stimuli despite a high predicted probability of being seen (>0.9). Within each panel, distributions are shown for tracking-off and tracking-on conditions. Under the tracking-off (*red* histograms), false-positive responses were associated with systematic horizontal fixation shifts resulting in a foveal-ward displacement of the stimulus (especially evident for locus 5 within the blind spot). Conversely, false-negative responses—particularly at locus 3 (just outside the blind spot)—were associated with fixation shifts in the opposite direction, leading to stimulus presentations displaced toward the optic nerve head. Under the tracking-on condition (*blue* histograms)—where the stimulus is presented at the intended location regardless of the subject's current fixation—these patterns were attenuated.

or spread parameters for the tracking on or off condition (Supplementary Tables S5, S6).

## Identifying the Causes of False-Negative and -Positive Responses in the Absence of Tracking

To determine whether fixation position at or near stimulus presentation contributed to likely false-positive responses, we analyzed participant responses for stimuli with a low predicted probability of being seen (<0.1, based on psychometric curve fits under tracked conditions). Across loci, responses obtained without tracking were consistently more likely to be reported as seen, with odds ratios (ORs) ranging from 2.04 to 3.28 and strong directional evidence (pd $\geq$ 0.99 at all loci; Fig. 4; Supplementary Tables S7, S8).

A negative horizontal displacement (i.e., stimulus displacement toward the "seeing" retina) showed modest and heterogeneous effects across loci under tracking-on conditions: The largest point estimate was observed at locus 2 (OR = 1.70, 95% CrI = 0.88–3.54, pd = 0.941), although the CrI included 1; effects at other loci were smaller and more uncertain.

But the interaction between a negative shift and tracking off showed strong evidence at selected loci, most notably at locus 3 (OR = 3.80, 95% CrI = 1.43–9.98, pd = 0.997) and locus 5 (OR = 3.19, 95% CrI = 1.85–5.50, pd = 1.000), indicating that fixation displacement substantially increased the likelihood of false-positive responses in the absence of tracking.

Conversely, for stimuli expected to be seen with high probability (>0.9), fixation shifts displacing the stimuli toward the optic nerve head under tracking-off conditions were associated with a reduced probability of reporting the stimulus as seen, corresponding to an increased occurrence of false-negative responses (see Fig. 4; Supplementary Tables S9, S10). Although the main effects of horizontal displacement and tracking alone were generally small and uncertain, strong evidence for interaction effects was observed at loci 2 and 3. Specifically, the odds of reporting a stimulus as seen were markedly reduced when positive displacement occurred with tracking off, with ORs of 0.12 (95% CrI = 0.03–0.41, pd = 1.000) at locus 2 and 0.24 (95% CrI = 0.10–0.54, pd = 1.000) at locus 3.

## Optimal Criterion for Suprathreshold (Defect-Mapping) Microperimetry

Analysis of the operating characteristics derived from the fitted FoS functions identified a criterion intensity of 13 dB as nominally optimal for distinguishing "seeing" from "non-seeing" retinal loci in single-presentation suprathreshold testing (Fig. 5). At this intensity, the Youden index reached 0.74 (95% CrI = 0.70–0.77). Restricting the analysis to locus 1 ("seeing") and locus 5 ("non-seeing") — thereby excluding peripapillary regions — yielded a nominally optimal criterion intensity of 15 dB (Youden index = 0.91, 95% CrI = 0.88–0.93); at 13 dB, the Youden index in this restricted

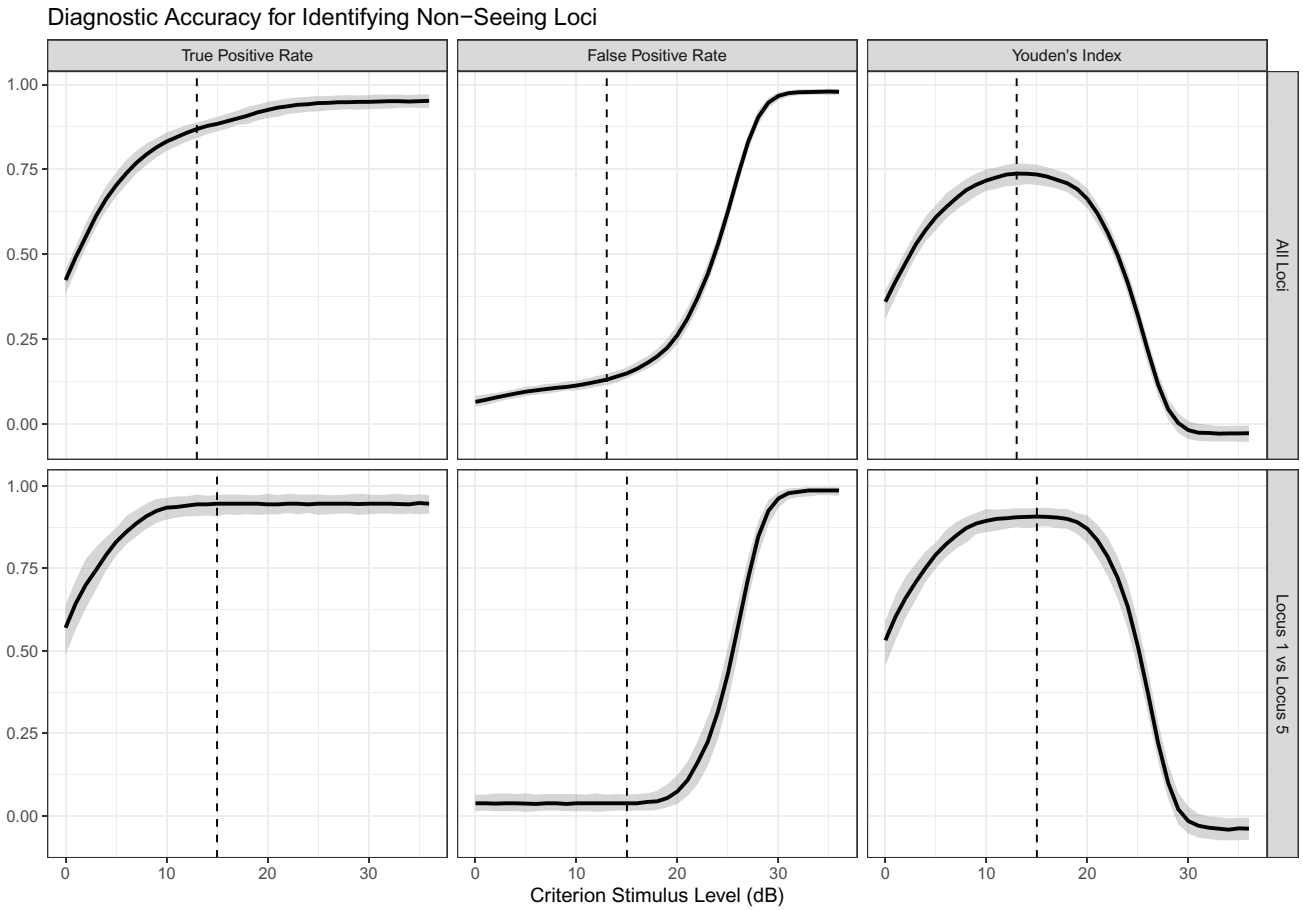


**Figure 5. Determination of an optimal criterion for suprathreshold (defect-mapping) microperimetry.** Using the fitted frequency-of-seeing curves, the figure illustrates how the choice of stimulus intensity affects the ability to identify "non-seeing" retinal loci with a single presentation. Curves depict the median (*solid line*) and 95% credible intervals (*shaded bands*) of the true-positive rate, false-positive rate, and the Youden index (*right*), plotted as a function of stimulus intensity. The *vertical dashed line* indicates the criterion intensity that maximizes the Youden index, representing the optimal trade-off between sensitivity and specificity. The *lower row* shows the same analysis, including only locus 1 as "seeing" and locus 5 as "non-seeing" to minimize artifacts related to peripapillary atrophy, which could alter the psychometric function, especially at locus 3. The 95% credible intervals are wider in the *bottom row* than in the *top row* because averaging over fewer per-subject curves leaves more posterior variance at each criterion intensity.

analysis was essentially identical (0.91, 95% CrI = 0.88–0.93; see Fig. 5).

The 95% CrI for the Youden index at 13 dB overlapped with that at 10 dB (Youden index 0.72, 95% CrI = 0.69–0.75), suggesting similar overall discriminative performance across this range of criterion intensities. Criterion levels below approximately 10 dB were associated with a progressive decline in the Youden index, driven primarily by a reduction in the true-positive rate for identifying "non-seeing" loci. In contrast, criterion levels above approximately 16 dB showed a marked decrease in the Youden index due to an increasing false-positive rate, reflecting misclassification of "seeing" loci as "non-seeing." Together, these findings indicate a relatively broad plateau of near-optimal performance between 10 and 16 dB, with diminishing diagnostic utility outside this range.

## Discussion

Microperimetry has become a practical clinical-trial tool promising retinal-contingent stimulus delivery, especially for patients with fixation instability. In routine interpretation, non-responses within established scotomata (e.g., within GA) and steep sensitivity recovery just outside lesion borders are often taken as indirect validation that stimuli are delivered where intended.[10] Yet, such "face validity" is an aggregate phenomenon; it does not resolve whether each individual presentation lands at its intended retinal location, particularly in regions with steep sensitivity gradients where sub-degree placement errors can change an outcome from "seen" to "not seen." Here, we addressed this knowledge gap directly by prospectively sampling FoS functions with the MAIA device under controlled conditions, and by explicitly manipulating the fundus tracking state (tracking enabled versus disabled). The optic nerve head provided a natural, anatomically defined deep scotoma with a relatively sharp boundary,[15] allowing us to evaluate stimulus delivery and psychometric behavior near the scotoma edge – precisely the context where tracking accuracy matters most for clinical monitoring.

Using the OPI and a constant-stimuli FoS paradigm, we quantified how tracking changes the psychometric function as a surrogate for the tracking quality. Two patterns were evident. First, sensitivity estimates in intact

retina were largely unchanged by tracking, consistent with the expectation that small placement variability has a limited effect when sensitivity is high. Second, tracking sharpened the psychometric transition (i.e., reduced the spread), most importantly at loci closest to the scotoma boundary. Overall, the spread parameter s was compatible with the typical point-wise retest reliability observed in microperimetry.[29–31]

These observations are the psychophysical signature expected if tracking reduces trial-to-trial spatial jitter: the sensitivity estimates – especially in the "seeing retina" – as an aggregate phenomenon of many stimulus presentations change little, but the FoS steepens, meaning that sensitivity estimates become less uncertain. If tracking can further reduce effective stimulus-placement variability in this best-case scenario, the benefit in clinical cohorts, where fixation stability is frequently degraded by disease, is likely to be larger, particularly for endpoints that depend on classifying loci in junctional zones or at lesion borders.

The present data are directly relevant to suprathreshold microperimetry, also known as defect-mapping microperimetry. The defect-mapping paradigm, advanced by Wu and colleagues in the context of retinal diseases as an alternative to conventional staircase thresholding, replaces full threshold determination with a binary classification of function—"seeing" versus "non-seeing"—derived from responses to a single fixed-intensity stimulus delivered at each locus.[10,15] By avoiding threshold estimation, test time can be reallocated to higher spatial sampling density across the macula, thereby improving the ability to capture and follow complex scotoma topographies.[10,15]

A central methodological issue for defect-mapping microperimetry is the choice of the criterion stimulus intensity. Two candidate criteria have been used in practice, 0 dB and 10 dB, but the 2 are not equivalent psychophysically. To date, 3 convergent lines of evidence have suggested that a criterion near 10 dB may be better suited for separating the "seeing" from the "non-seeing" retina than the maximum-intensity stimulus of 0 dB (equal to 318 cd/m$^2$ on the mesopic MAIA scale). First, 10 dB has been conceptualized as the lower bound of the "effective dynamic range," recognizing that measurements at the physical floor are not stably reproduced on retest and that the probability of observing <0 dB is nontrivial even when true sensitivity is near the floor.[31] Second, in a GA cohort, defect mapping with a 10 dB stimulus has demonstrated a closer association with imaging-derived atrophy metrics than defect mapping using 0 dB.[10] Notably, responses to the 0 dB stimulus within areas of atrophy have been documented – either due to residual cone cells or stray light.[14] Third, simulation analyses anchored in empirical microperimetry datasets indicate that repeatable <0 dB outcomes occur in only approximately 60% of test–retest pairs even when a truly nonresponding location exists, whereas repeatable deep defects defined by ≤10 dB on both tests exceed 90% sensitivity while maintaining low false-positive rates.[16] Together, these observations argue that 0 dB is an overly stringent operational definition of "non-seeing," whereas 10 dB more closely tracks the reliably classifiable portion of deep sensitivity loss.

The present study adds a fourth and independent line of evidence by providing a psychometrically grounded criterion derived from FoS modeling. Specifically, FoS curves allow estimation of the stimulus intensity corresponding to a low detection probability (e.g., 5%), thereby defining a suprathreshold cutoff anchored to a measured response behavior rather than a heuristic convention. On this basis, a criterion stimulus level of 13 dB emerged as the nominally optimal separator between "seeing" and "non-seeing" loci. However, this estimate is anchored to normative sensitivity at "seeing" locations in healthy observers. Eyes with GA secondary to AMD show considerable retina-wide sensitivity loss, whereas sensitivity rises sharply across the junctional zone (usually >10 dB directly outside of GA).[3,32] These opposing effects mean the 10 to 16 dB plateau we observed in healthy eyes for discriminating "seeing" from "non-seeing" retina likely narrows in the context of GA, with the 10 dB criterion being among the most defensible choices for defect-mapping microperimetry.

## Limitations

First, the data were obtained from young healthy volunteers (median age 27 years), and the optic nerve head was used as a surrogate scotoma. Older adults — and especially patients with AMD or GA — exhibit reduced fixation stability, slower reaction times, and greater media opacity, all of which are expected to amplify the tracking-related effects we report. Because eyes with GA show distinct features (macula-wide sensitivity loss, border complexity with various degrees of dysfunction in the junctional zone, and fixation instability), disease-specific data could yield a more definite criterion choice for suprathreshold perimetry.

Second, the MAIA2 cannot deliver a stimulus brighter than 0 dB, so the left tail of the FoS curve at loci 4 and 5 is truncated at the device floor; the sensitivity estimates at these loci should be interpreted as the apparent midpoint of the measurable portion of the psychometric function. In addition, the OPI fixation log is output at a lower rate than the MAIA2's internal tracking (i.e., only a fuzzy match between fixation and stimulus responses was possible). A custom adaptive-optics microperimetry setup with full access to fixation data could enable a more precise evaluation of how variability in stimulus placement shapes psychometric behavior.[33–36] Likewise, the availability of newer hardware (e.g., MAIA3[37] with improved tracking according to the manufacturer) may alter the magnitude of effects observed here.

Third, we assumed a causal link between horizontal stimulus displacement and false-positive and false-negative responses; however, attentional drift over time (i.e., changes in lapse and guess rate) – which is itself likely associated with fixational drift – might also contribute directly to such responses.

In summary, fundus tracking confers a modest yet consistent gain in measurement reliability, even under stable fixation, and is associated with steeper psychometric functions near scotoma boundaries. These data also provide a psychometric foundation for selecting the suprathreshold criterion used in defect-mapping paradigms. A cutoff near 13 dB (on the respective mesopic microperimetry scale) is supported as a reasonable and defensible criterion for separating "seeing" from "non-seeing" retina in normative young retina, with the more conservative 10 dB criterion preferred for clinical use in AMD/GA cohorts. Together, these findings strengthen the scientific rationale of current defect-mapping microperimetry protocols as an efficient outcome measure in macular diseases.

Investigative Ophthalmology & Visual Science

### *Acknowledgments*

Supported by the BrightFocus Foundation (M2024009N to Maximilian Pfau).

Disclosure: **T. Lipsky**, None; **C. Ehrenzeller**, None; **G. Ansari**, None; **K. Pfau**, Daiichi Sankyo (C), Inozyme (F), Heidelberg Engineering (R), Bayer (R), Roche (R); **W. Harmening**, None; **Z. Wu**, None; **M. Pfau**, iCare (F), Inozyme (F), Roche (F)

### *References*

1. Pfau M, Jolly JK, Wu Z, et al. Fundus-controlled perimetry (microperimetry): application as outcome measure in clinical trials. *Prog Retin Eye Res*. 2021;82:100907.
2. Csaky KG, Patel PJ, Sepah YJ, et al. Microperimetry for geographic atrophy secondary to age-related macular degeneration. *Surv Ophthalmol*. 2019;64(3):353–364.
3. 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.
4. Hofmann J, Domdei L, Jainta S, Harmening WM. Assessment of binocular fixational eye movements including cyclotorsion with split-field binocular scanning laser ophthalmoscopy. *J Vis*. 2022;22(10):5.
5. Reiniger JL, Domdei N, Holz FG, Harmening WM. Human gaze is systematically offset from the center of cone topography. *Curr Biol*. 2021;31(18):4188–4193. e3.
6. Morales MU, Saker S, Wilde C, et al. Reference clinical database for fixation stability metrics in normal subjects measured with the MAIA microperimeter. *Transl Vis Sci Technol*. 2016;5(6):6.
7. Schmitz-Valckenberg S, Bültmann S, Dreyhaupt J, Bindewald A, Holz FG, Rohrschneider K. Fundus autofluorescence and fundus perimetry in the junctional zone of geographic atrophy in patients with age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2004;45(12):4470–4476.
8. Li M, Huisingh C, Messinger J, et al. Histology of geographic atrophy secondary to age-related macular degeneration: a multilayer approach. *Retina*. 2018;38(10):1937–1953.
9. Pfau M, von der Emde L, Dysli C, et al. Determinants of cone- and rod-function in geographic atrophy: AI-based structure-function correlation. *Am J Ophthalmol*. 2020;217:162–173.
10. Wu Z, Hadoux X, Jannaud M, et al. Visual sensitivity loss in geographic atrophy: structure-function evaluation using defect-mapping microperimetry. *Invest Ophthalmol Vis Sci*. 2024;65(1):36.
11. Saeed A, Guymer RH, Hadoux X, et al. Longitudinal changes of visual sensitivity with geographic atrophy progression assessed by defect-mapping microperimetry. *Invest Ophthalmol Vis Sci*. 2024;65(14):13.
12. Maddess T. The influence of sampling errors on test-retest variability in perimetry. *Invest Ophthalmol Vis Sci*. 2011;52(2):1014–1022.
13. Cideciyan AV, Roman AJ, Jacobson SG, et al. Developing an outcome measure with high luminance for optogenetics treatment of severe retinal degenerations and for gene therapy of cone diseases. *Invest Ophthalmol Vis Sci*. 2016;57(7):3211–3221.
14. Pfau M, von der Emde L, Dysli C, et al. Light sensitivity within areas of geographic atrophy secondary to age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2019;60(12):3992–4001.
15. 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.
16. Wu Z, Pfau M, Fleckenstein M, Guymer RH. Microperimetry characteristics of regions with a truly nonresponding location: implications for atrophic age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2024;65(8):44.
17. Denniss J, Baggaley HC, Astle AT. Frequency-of-seeing curves (psychometric functions) for perimetric stimuli in age-related macular degeneration. *Ophthalmic Physiol Opt*. 2025;45(1):301–307.
18. Chauhan BC, Tompkins JD, LeBlanc RP, McCormick TA. Characteristics of frequency-of-seeing curves in normal subjects, patients with suspected glaucoma, and patients with glaucoma. *Invest Ophthalmol Vis Sci*. 1993;34(13):3534–3540.
19. Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci*. 2000;41(2):417–421.
20. Spry PG, Johnson CA, McKendrick AM, Turpin A. Variability components of standard automated perimetry and frequency-doubling technology perimetry. *Invest Ophthalmol Vis Sci*. 2001;42(6):1404–1410.
21. Rubinstein NJ, Turpin A, Denniss J, McKendrick AM. Effects of criterion bias on perimetric sensitivity and response variability in glaucoma. *Transl Vis Sci Technol*. 2021;10(1):18.
22. Wall M, Maw RJ, Stanek KE, Chauhan BC. The psychometric function and reaction times of automated perimetry in normal and abnormal areas of the visual field in patients with glaucoma. *Invest Ophthalmol Vis Sci*. 1996;37(5):878–885.
23. Turpin A, Artes PH, McKendrick AM. The Open Perimetry Interface: an enabling tool for clinical visual psychophysics. *J Vis*. 2012;12(11):22.
24. Turpin A. OPI: Open Perimetry Interface. 2022. Accessed July 28, 2026. https://cran.r-project.org/package=OPI.
25. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9(7):671–675.
26. King-Smith PE, Grigsby SS, Vingrys AJ, Benes SC, Supowit A. Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Res*. 1994;34(7):885–912.
27. Bürkner PC. brms: an R package for Bayesian multilevel models using Stan. *J Stat Softw*. 2017;80(1):1–28.
28. Caldwell AR. SimplyAgree: an R package and jamovi module for simplifying agreement and reliability analyses. *J Open Source Softw*. 2022;7(71):4148.
29. Wong EN, De Soyza JDA, Mackey DA, Constable IJ, Chen FK. Intersession test-retest variability of microperimetry in type 2 macular telangiectasia. *Transl Vis Sci Technol*. 2017;6(6):7.
30. Schweighofer J, Birner K, Mohamed H, et al. Benchmarking test-retest variability in microperimetry for intermediate age-related macular degeneration using MP-3 and MAIA. *Can J Ophthalmol*. 2025;60(6):e879–e887.
31. 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(6):BIO158–BIO167.
32. Pfau M, von der Emde L, de Sisternes L, et al. Progression of photoreceptor degeneration in geographic atrophy secondary to age-related macular degeneration. *JAMA Ophthalmol*. 2020;138(10):1026–1034.
33. Tuten WS, Tiruveedhula P, Roorda A. Adaptive optics scanning laser ophthalmoscope-based microperimetry. *Optom Vis Sci*. 2012;89(5):563–574.
34. Harmening WM, Tuten WS, Roorda A, Sincich LC. Mapping the perceptual grain of the human retina. *J Neurosci*. 2014;34(16):5667–5677.

35. Bruce KS, Harmening WM, Langston BR, Tuten WS, Roorda A, Sincich LC. Normal perceptual sensitivity arising from weakly reflective cone photoreceptors. *Invest Ophthalmol Vis Sci*. 2015;56(8):4431–4438.
36. Tuten WS, Vergilio GK, Young GJ, et al. Visual function at the atrophic border in choroideremia assessed with adaptive optics microperimetry. *Ophthalmol Retina*. 2019;3(10):888–899.
37. Ansari G, Giudici NL, Cancian G, et al. Validation and repeatability of differential light sensitivity measurements with the novel MAIA microperimetry device. *Ophthalmol Sci*. 2025;5(6):100886.

Investigative Ophthalmology & Visual Science