Supplemental Material

Glaucoma home-monitoring using a tablet-based visual field test (Eyecatcher): An assessment of accuracy and adherence over six months

Pete R. Jones, Peter Campbell, Tamsin Callaghan, Lee Jones, Daniel S. Asfaw, David F. Edgar, David P. Crabb

1. Additional *Methods*

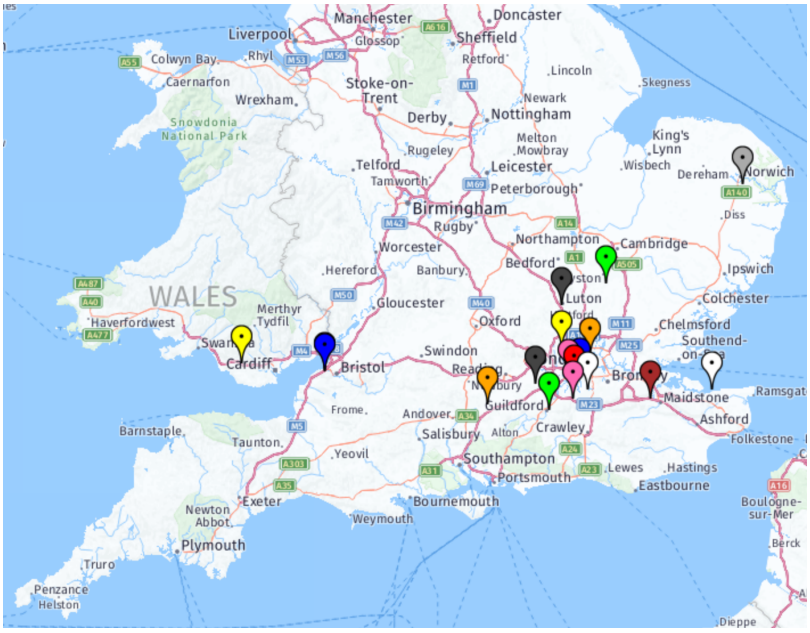

Fig S1. Geospatial distribution of participants.

1.1. Screen (luminance) calibration

As shown in Figure S2, input-output functions (i.e., mapping command levels, in bits, to display luminance, in cd/m^2) were measured at 48 uniformly-spaced screen locations. At each location, measurements were made for 64 uniformly spaced input levels (0 . . . 255, with bitstealing for sub-unit steps). Two-dimensional tensor-product linear-interpolation was then used to compute the appropriate calibration function for each and every pixel location. Measurements were made a ColorCal MK II colorimeter (Cambridge Research Systems, Cambridge, UK), and validated with a CS-100 chromameter (Minolta Camera Co., Osaka, JP). All calibration measurements were performed using localized targets, presented against an approximately 10 cd/m^2 background.

The CS-100 spot photometer was also used to examine the effects of viewing angle. Even at the most eccentric test-angles ($\pm$ 15°), luminance drop-off was minimal. Therefore, while viewing-angle effects could be corrected for in future iterations of the test, not doing so is unlikely to have affected the present results substantively.

LCD screens are notoriously non-uniform in their luminance output, with input-output functions varying markedly as a function of screen location. This was certainly the case with the inexpensive HP Pavilion x360 15.6″ device used in the present study. Thus, median maximum luminance (across all displays/screen-locations) was 217 cd/m^2. However, within the 20 individual displays,

luminance at each screen location varied by up to 27 – 63 cd/m^2 (Median: 46 cd/m^2). A similar level of luminance variability was observed between displays. When comparing corresponding screen-locations, luminance varied by 27 – 59 cd/m^2 (Median: 41 cd/m^2). Further one of the twenty displays had to be replaced as it was unable to present the requisite range of stimulus values. The practical corollary is that, at least with the present hardware, extensive photometric calibration is required to ensure precise stimulus control, and such calibrations must be performed independently for each individual display (i.e., it would not have been appropriate to generalize from the calibration of a single device, or to use the average measurements from multiple devices).

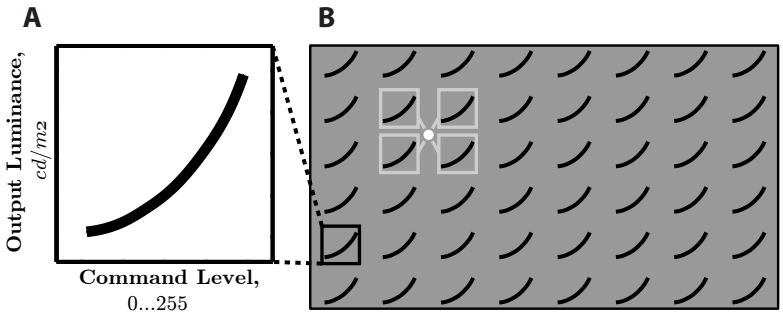

Fig S2. Screen (luminance) calibration schematic. Empirical measurements made using a uniform grid of test locations were interpolated to determine the input-output function for any given pixel location.

1.2. Computing MD

Following Heijl (1987), Mean Deviation (MD) was computed as the weighted-mean difference between estimated sensitivities at each location, and age-corrected normative values, weighted to take into account the relative reliability (in normal eyes) of measurements at each location, thus:

$$MD = \frac{\sum_{i=n}^{n} \frac{(\bar{x_i} - z_i)}{s_i^2}}{\sum_{i=n}^{n} \frac{1}{s_i^2}}, \tag{1}$$

where n is the number of test locations (here fixed at $n = 24$), $\bar{x_i}$ is the estimated sensitivity at location i, z_i is the corresponding age-corrected normal value at that location, and s_i^2 is the variance of the age-corrected normal values. MD becomes more negative with increased field loss. Note that normative values were taken from measurements made previously using standard automated perimetry (Heijl, 1987), not Eyecatcher.

2. Additional *Results*

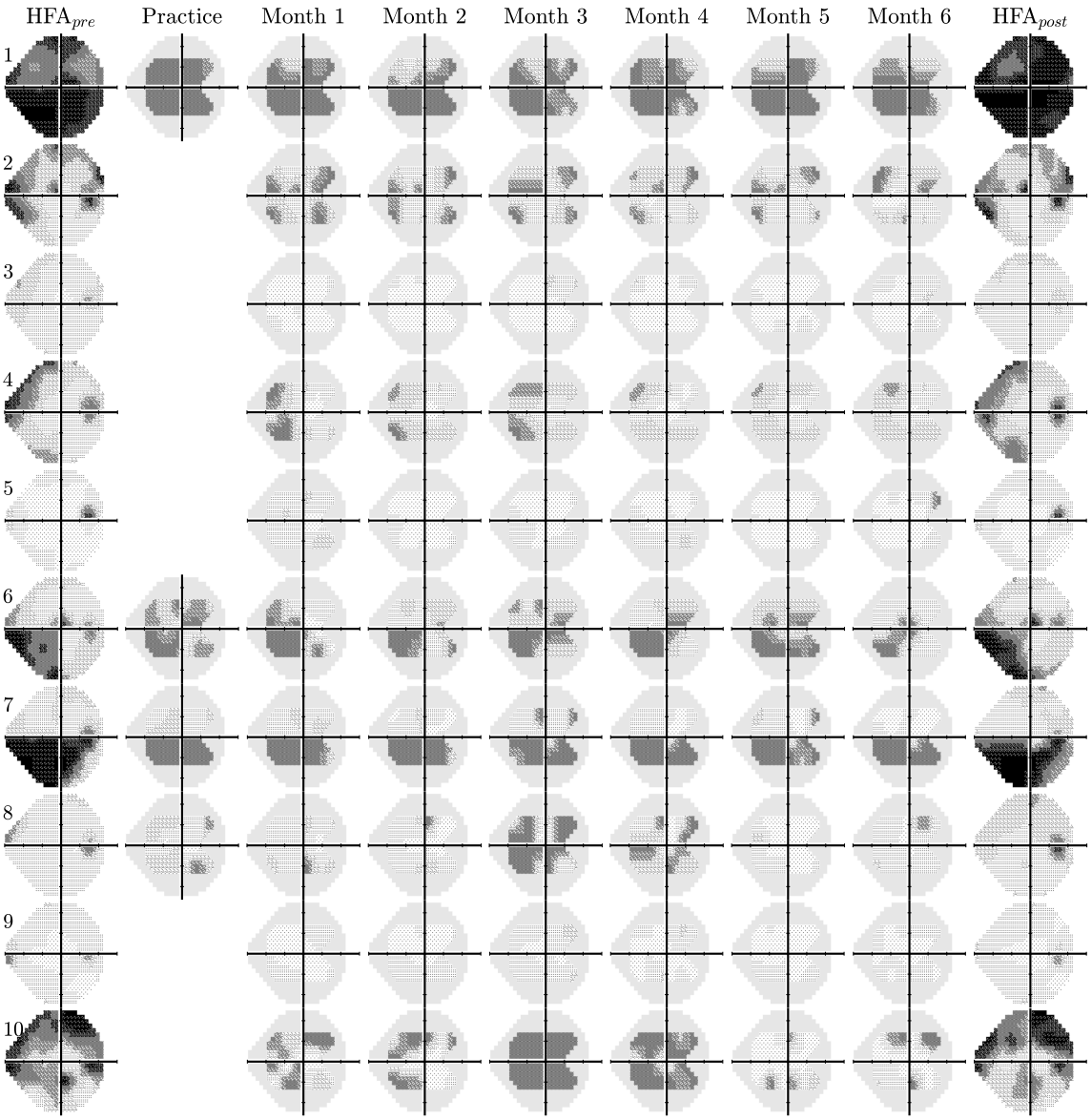

Fig S3. Raw right eye VF data for participants 1-10 (same format as left-eye data, shown in Figure 4 of *Main Manuscript*).

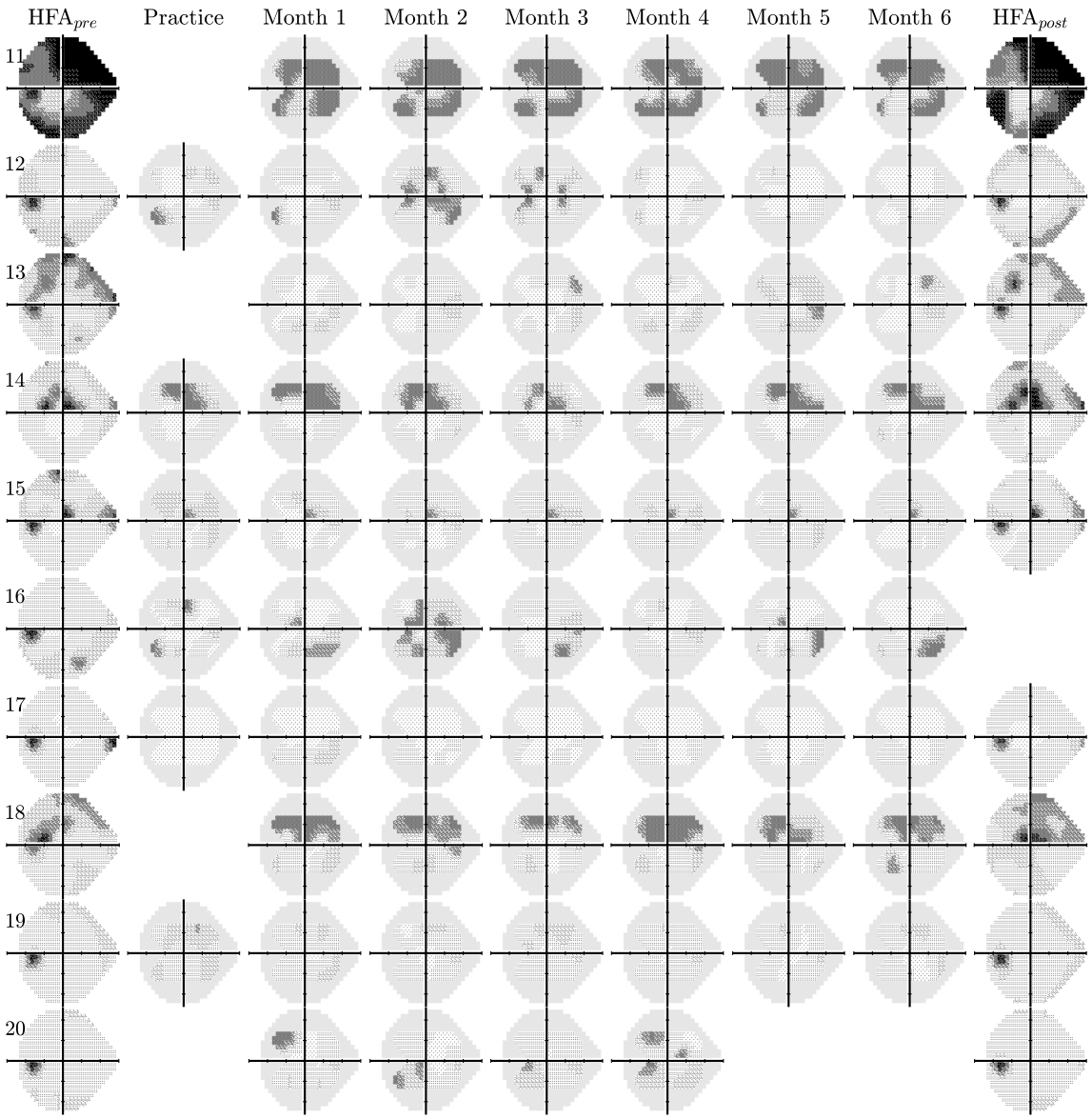

Fig S4. Raw left eye VF data for participants 11-20 (same format as Figure 4 of *Main Manuscript*).

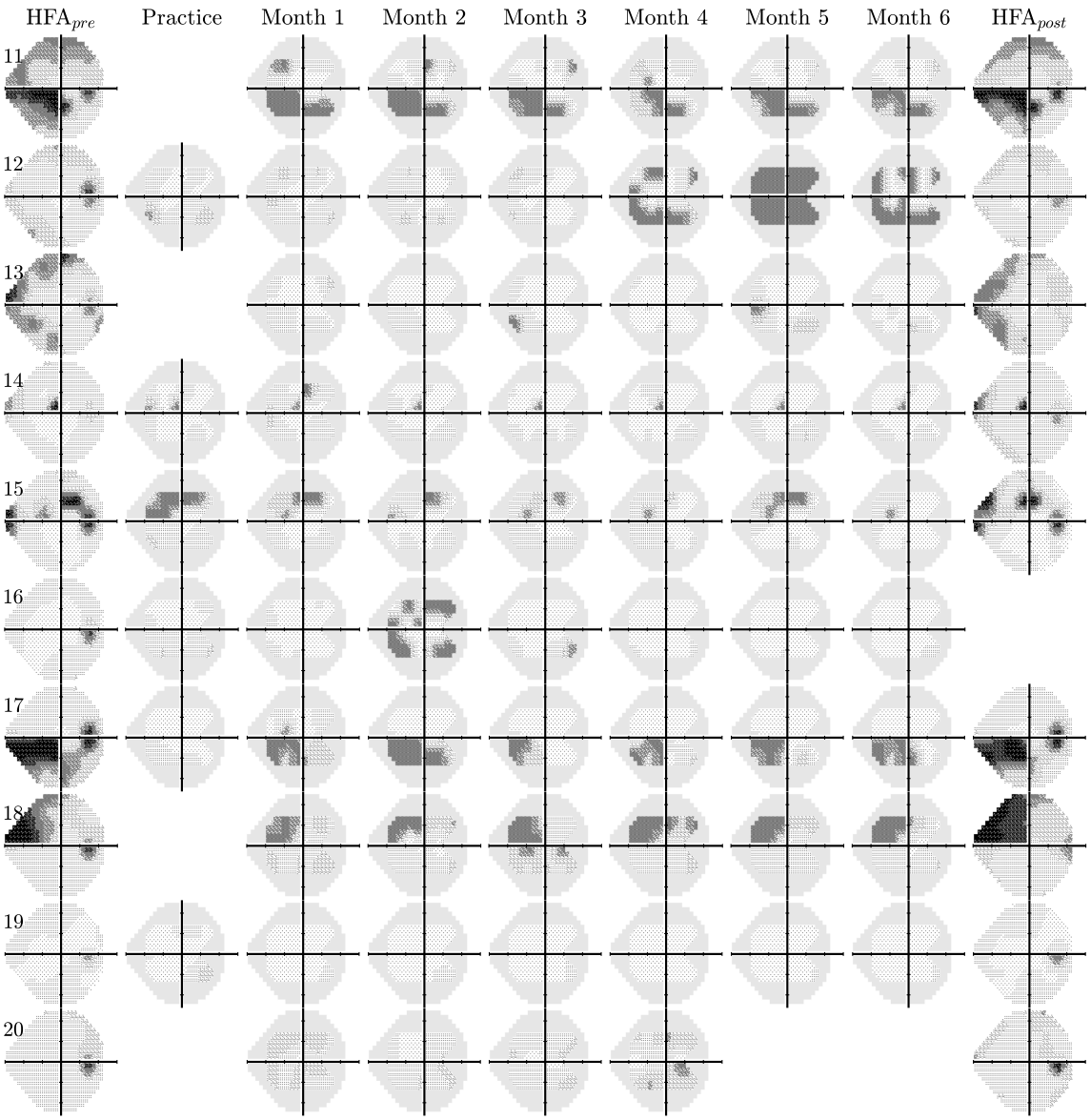

Fig S5. Raw right eye VF data for participants 11-20 (same format as Figure 4 of *Main Manuscript*).

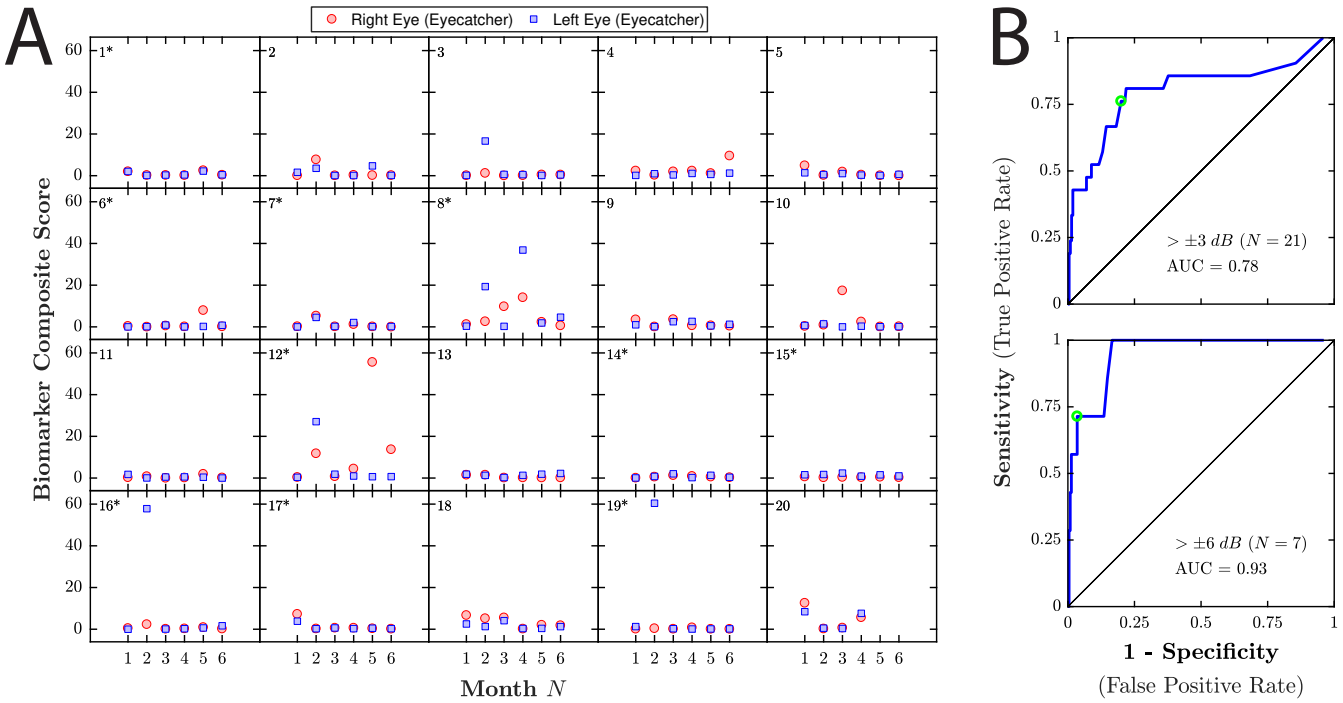

Fig S6. Data from the tablet's front-facing camera (and participant response latencies) were used identify anomalous tests. Following a method similar to *Jones et al, TVST, 2020*, a "Biomarker Composite Score" (BCS) was computed by linearly combining three variables: (1) mean response latency; (2) Variability in camera pixel intensity throughout the test (i.e., which would be high if somebody suddenly turned on a light or opened a door); (3) an OpenFace facial expression metric. Formally, $BCS = \left[\frac{1}{L_\mu} \cdot \max\left(\Delta I_\mu\right) \cdot H^c\right]$, where L_μ is mean response latency, in seconds, across all trials, ΔI_μ is the difference in mean pixel intensity between consecutive video frames, H is estimated Happiness (a value from 0-1, from OpenFace), and c is the estimated confidence of H (a value from 0-1, from OpenFace). ***(A)*** Raw BCS for each individual test, for comparison with the MD scores shown in the same format in Figure 2 (*Main Manuscript*). ***(B)*** Corresponding Receiver Operating Characteristics (ROCs). Moderately anomalous test results (MD more than $\pm 3\,dB$ from the median) could be identified with reasonable sensitivity/specificity ($AUC = 0.78$). Highly anomalous test results (MD more than $\pm 6\,dB$ from the median) could be identified with high sensitivity/specificity ($AUC = 0.94$). In principle these tests could be excluded or down-weighted post-hoc, or individuals could be automatically invited to repeat such tests.

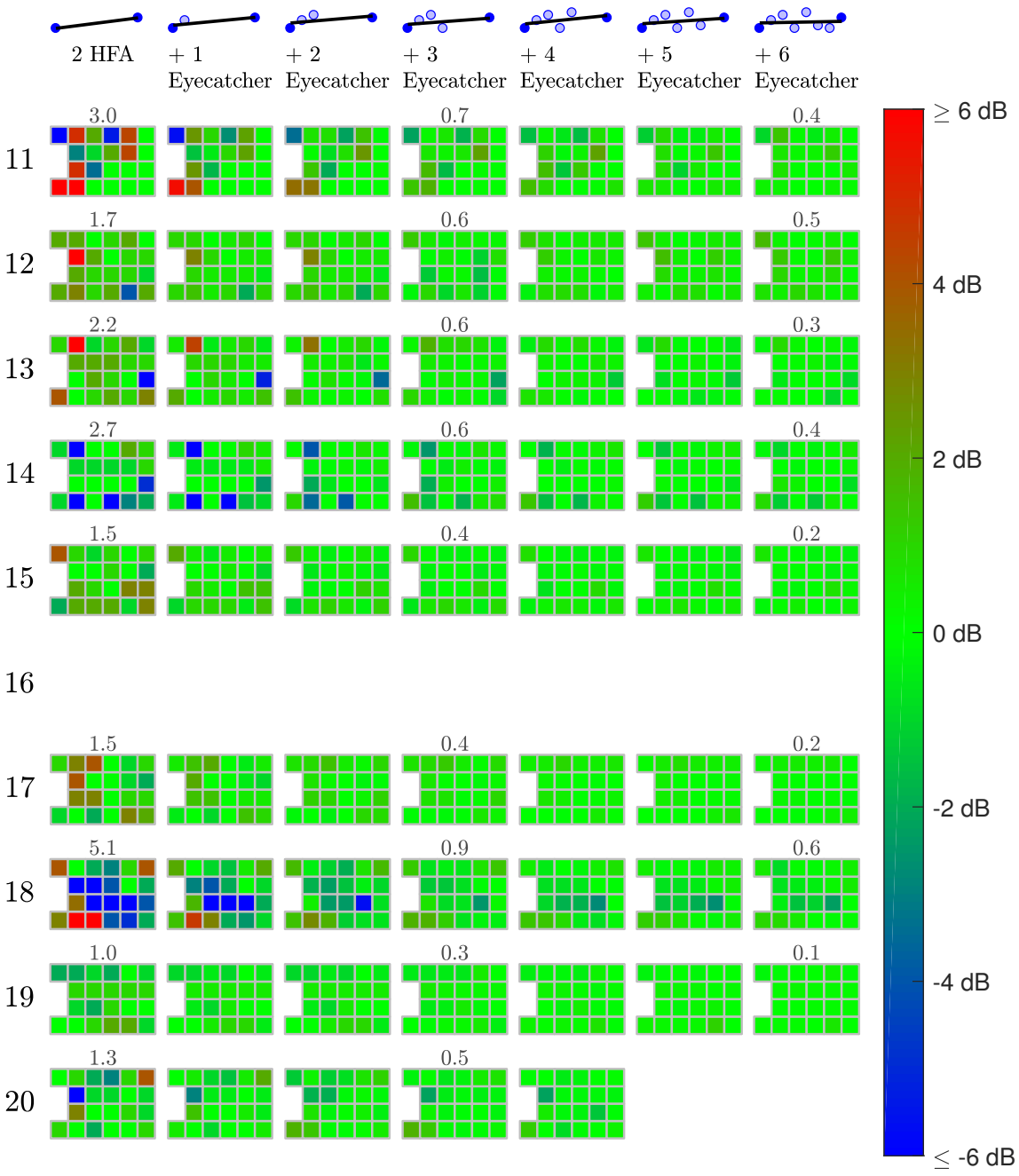

Fig S7. Benefit of home monitoring (reduction in rate-of-change measurement error) for the right eyes of participants 1-10 (same format as Figure 5A of *Main Manuscript*).

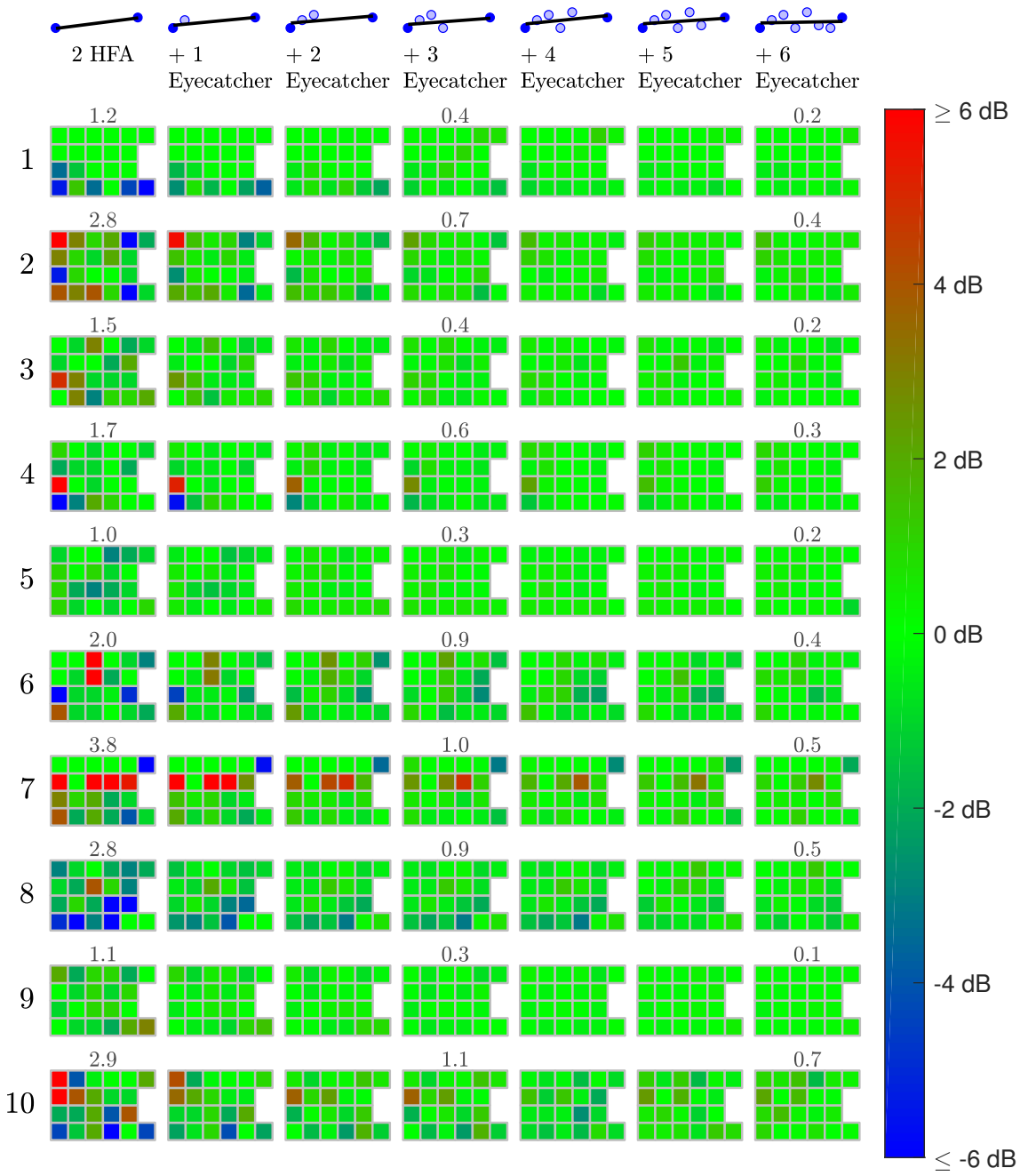

Fig S8. Benefit of home monitoring (reduction in rate-of-change measurement error) for the left eyes of participants 1-10 (same format as Figure 5A of *Main Manuscript*).

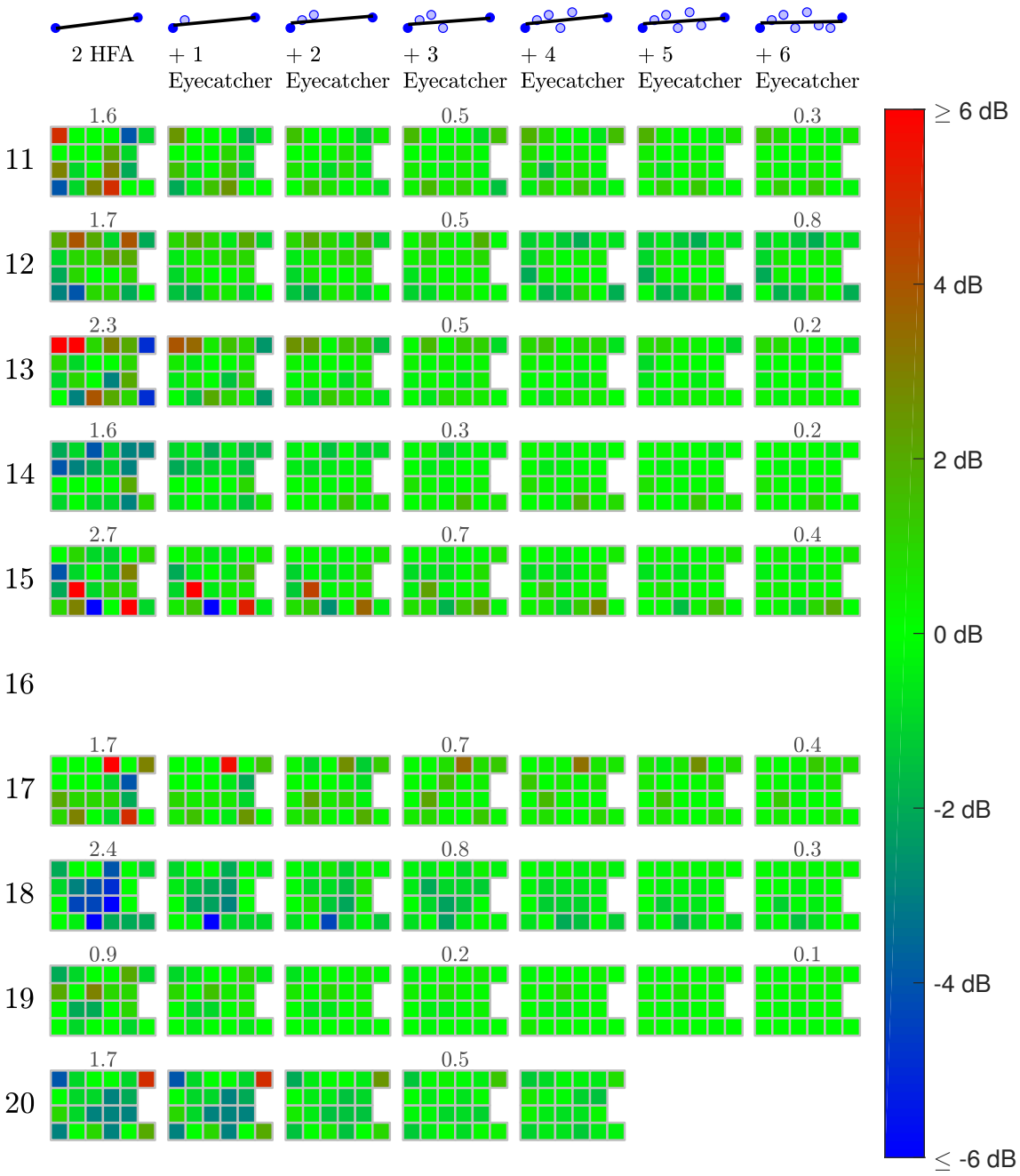

Fig S9. Benefit of home monitoring (reduction in rate-of-change measurement error) for the right eyes of participants 1-10 (same format as Figure 5A of *Main Manuscript*).

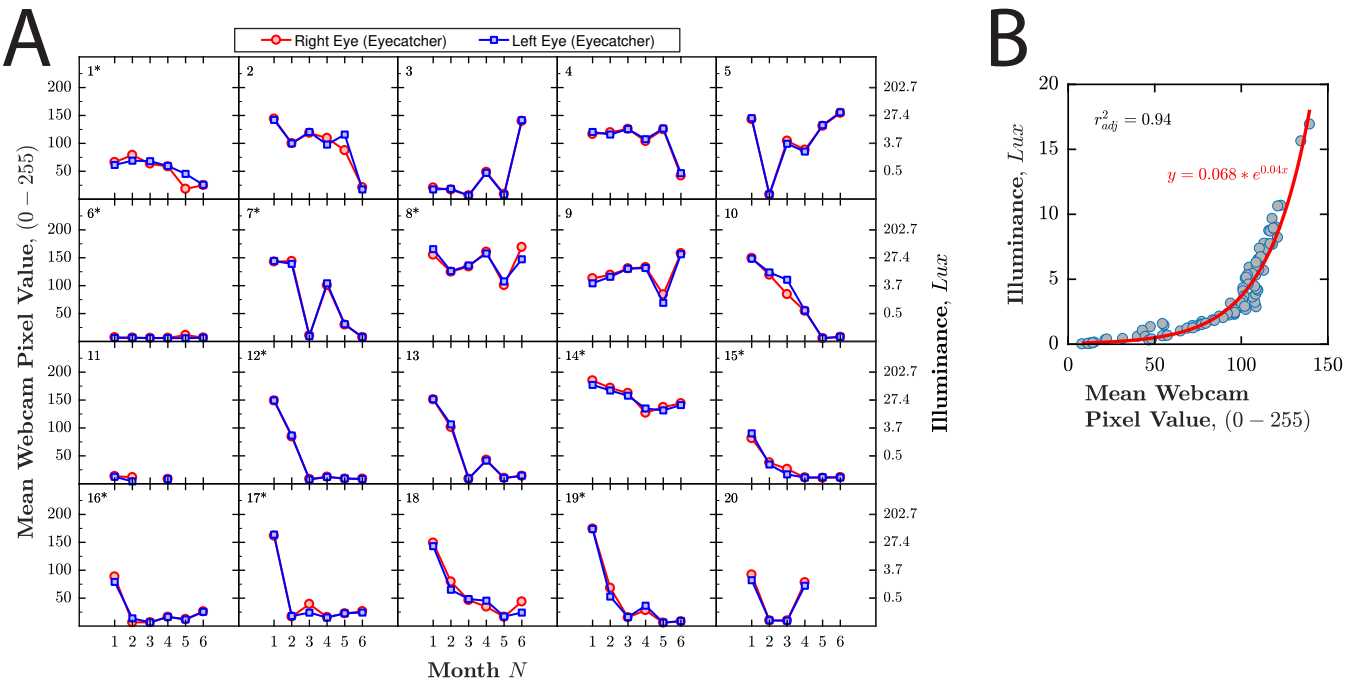

Fig S10. Ambient illumination levels during home testing. ***(A)*** Each marker represents the mean webcam pixel intensity value, averaged across all pixels and all frames. A secondary *y*-axis shows the corresponding estimated illuminance, in *Lux*, based on a the calibration data shown in (B). Note that for participant 20, data are missing for tests 5 and 6 as they stopped performing the test. For participant 11, data are missing for tests 3, 5, 6 because they chose to cover the camera on these tests. When questioned subsequently, they stated that this was because they found the recording light distracting (not for privacy). ***(B)*** Calibration data. Each marker ($N = 90$) represents raw measurements, made using an AOPUTTRIVER AP-881D Digital Lux Meter. The red line is the best fitting (least-square) exponential function, which provides an acceptable approximation of the data for present purposes. Note that these measurements pertain only to overall ambient lighting, and it is entirely conceivable that localized screen glare might have an even greater influence on visual field measurements.

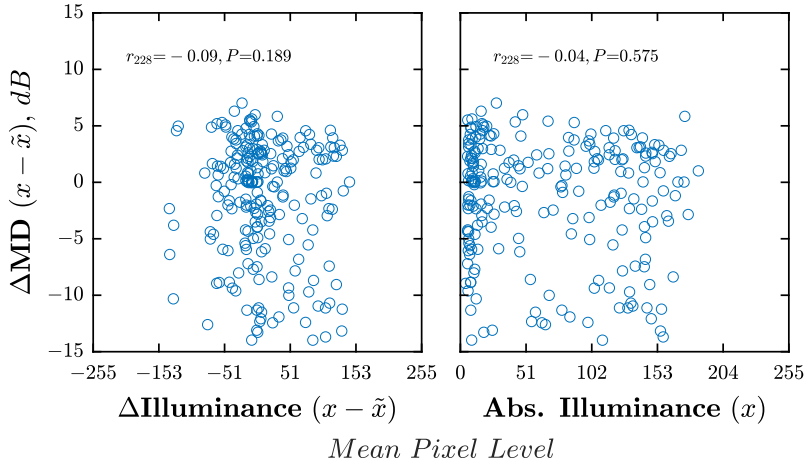

Fig S11. Relationship between ambient lighting and estimated VF loss. ***(A)*** There was no relationship between changes in illuminance and changes in test score [$r_{214} = 0.07, P = 0.320$]. ***(B)*** There was no relationship between absolute illuminance and absolute test score [$r_{214} = -0.06, P = 0.373$].

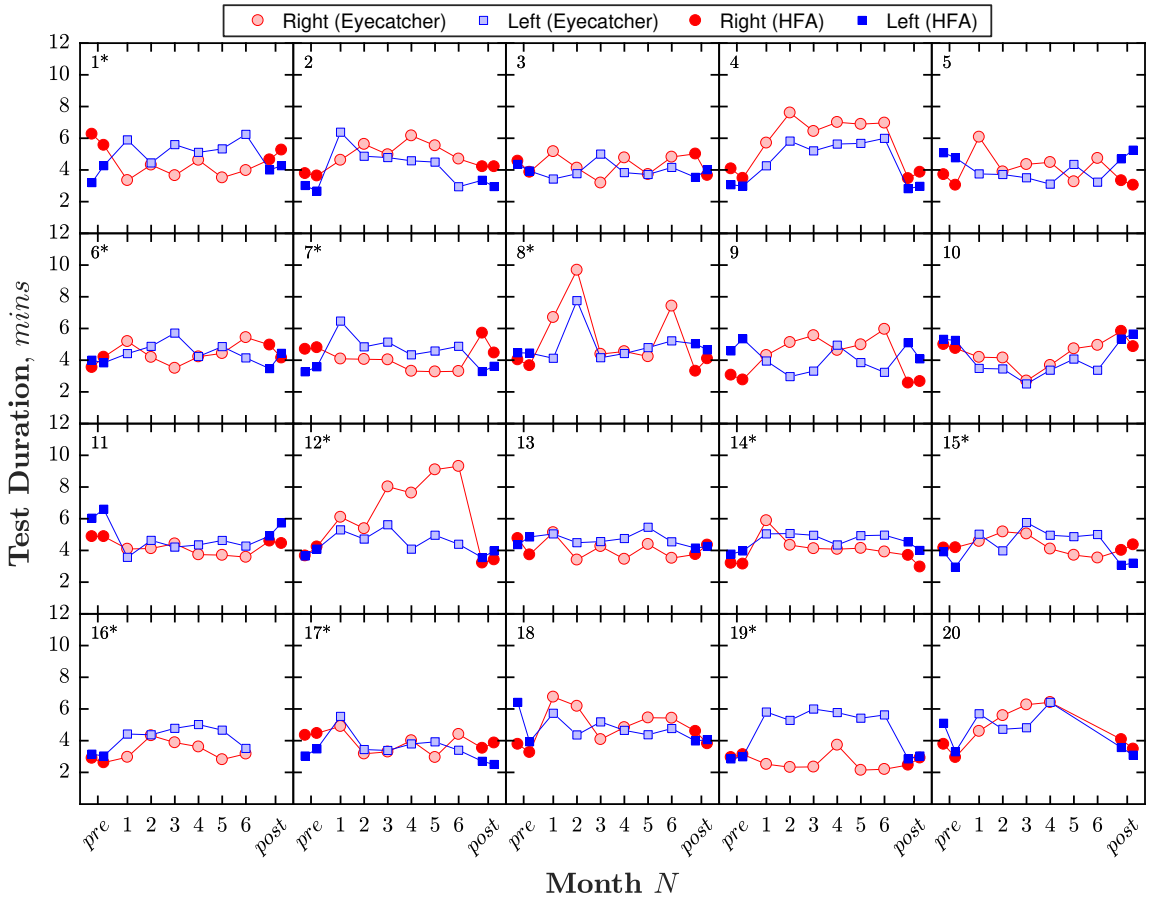

Fig S12. Test duration. Same format as Figure 2 of *Main Manuscript*.

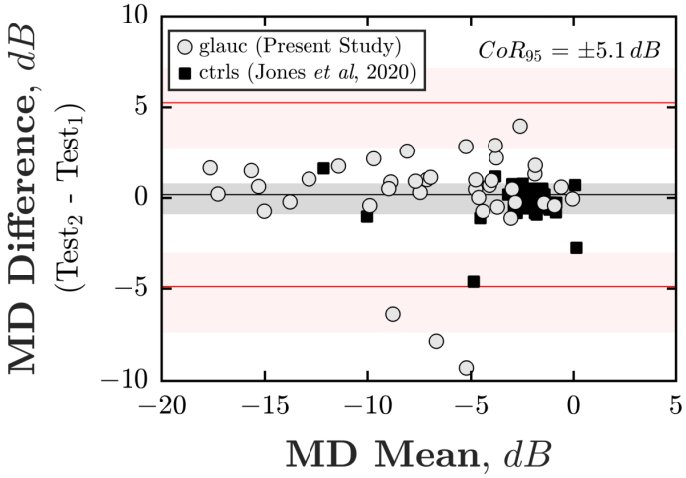

Fig S13. Reliability of each individual home-test (test-retest repeatability). Bland-Altman agreement from the first two Eyecatcher home tests (Month 2 - Month 1). Each marker corresponds to a single eye. Solid squares show analogous data from 46 control eyes, reported previously by *Jones et al, TVST, 2020*. These data are shown for illustration only, and were not included when computing the limits of agreement shown. If included, the 95% Coefficient of Repeatability was $\pm 2.8\,dB$.