



City Research Online

City, University of London Institutional Repository

Citation: Han, H. K. & Jones, P. R. (2019). Plug and play perimetry: Evaluating the use of a self-calibrating digital display for screen-based threshold perimetry. *Displays*, 60, pp. 30-38. doi: 10.1016/j.displa.2019.08.006

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/23783/>

Link to published version: <https://doi.org/10.1016/j.displa.2019.08.006>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk

TITLE

Plug and play perimetry: Evaluating the use of a self-calibrating digital display for screen-based threshold perimetry

RUNNING TITLE

Plug and play perimetry

AUTHORS

Hyun Kyu Han¹, Pete R. Jones^{1,2}

¹ *Institute of Ophthalmology, University College London (UCL), UK.*

² *NIHR Moorfields Biomedical Research Centre, London, UK.*

NOTES

Figures and Tables are included in the body-text as low-res thumbnails. High quality versions will be uploaded separately.

ABSTRACT

1 This study evaluated the feasibility of using a 'self-calibrating' display (EIZO CG277) to perform screen-
2 based threshold perimetry. Such displays incorporate their own integrated photometer, so could
3 potentially be used 'straight out of the box', without the need for time-consuming and costly
4 luminance calibration by skilled experts. Concerns remain, however, due to the fact that the internal
5 calibration of such devices is imperfect (i.e., is limited to a single screen location only) and due to
6 lingering doubts regarding the accuracy of screen-based perimetry in general. To evaluate such a
7 system, automated static threshold perimetry was performed in thirty-two normal-sighted adults. In
8 one condition, participants performed a novel screen-based perimetry test, for which the screen was
9 extensively calibrated using traditional photometric techniques/equipment. In a second condition, the
10 same test was performed, but the display was calibrated using only the screen's integrated
11 photometer (and assuming uniformity across the display). For reference, participants also completed a
12 traditional visual-field assessment using a Humphrey Field Analyzer (HFA). All three tests were
13 performed twice to assess test-retest repeatability (six tests total). The results showed no differences
14 when comparing screen-based perimetric measurements made with internal self-calibration vs full
15 manual calibration (either in terms of mean sensitivity, pointwise sensitivity, test-retest repeatability,
16 or test duration). Furthermore, the accuracy and precision of both were indistinguishable from the
17 current gold standard (HFA), although the HFA was approximately two minutes (~30%) faster. These
18 results indicate that self-calibrating commercial monitors can be used to perform screen-based
19 perimetry almost as well as current clinical devices, and without the need for any specialized
20 knowledge or equipment to setup or maintain. This could facilitate perimetric testing in currently
21 hard-to-reach settings, such as community centers, stroke wards, homes, rural locations, or
22 developing countries.

KEY WORDS: *Visual Fields; Perimetry; Contrast Sensitivity; Screen Calibration; Photometry; Eye-movements*

23 1. INTRODUCTION

24 Assessment of the visual field via static threshold perimetry is a key element of modern ophthalmic
25 assessments, where it is used routinely to diagnose and monitor common eye-diseases such as
26 glaucoma and diabetes¹. Traditionally, perimetry is performed using specialized devices such as the
27 Humphrey Field Analyzer (Carl Zeiss Meditec Inc., Dublin, CA, USA) or Octopus Perimeter (Haag-Streit
28 AG, Köniz, Switzerland): large, cumbersome machines in which the user places their head inside a
29 white bowl, into which lights of variable intensity are project.

30 Recently, however, there has been a proliferation of 'screen-based' perimeters. These include
31 portable, tablet devices²⁻⁹, as well as eye-movement perimeters that use remote eye-tracking
32 technology to position stimuli on the retina and record saccadic responses⁹⁻¹⁶. Although not without
33 their drawbacks, such 'disruptive' devices have a number of potential advantages over tradition
34 perimeters in terms of comfort and cost (see *Discussion: Implications and Applications*). One of the
35 principle attractions of screen-based perimeters is that they use relatively ordinary commercial
36 technology that is widely available and easily replaceable. The implication is that they could therefore
37 be easily set-up and maintained in community settings or across the developing world. It is even
38 conceivable that perimetry could become an 'app', which users download to run in the comfort of
39 their own home, thereby conferring substantial benefits in terms of patient satisfaction¹⁷ and financial
40 savings¹⁸.

41 In reality, however, the need for luminance calibration remains a significant impediment to the
42 widespread use of screen-based perimeters. Thus, with a digital display, the commands that are sent
43 to the screen are unitless values (e.g., numbers between 0--1024), but what we wish to manipulate is
44 the luminance of the stimulus on the screen (e.g., in candelas per meter squared; cd/m²). Perimetry
45 therefore requires us to know how a given input (Command Level) translates to a given output
46 (Luminance). This is the input-output function of the display device, and depends on a wide range of
47 factors, including the model/make of the screen, its current settings, the model/make of the graphics
48 card, the temperature of the room, how long the display has been active for, and what else is currently
49 being displayed on the screen^{19,20}. Because of this complexity, the input-output function can only be
50 determined empirically, and must be measured for each individual device. Realistically, most of the
51 potential operators of a screen-based perimeter --- be they patients or clinicians --- lack the necessary
52 time, training, or equipment to perform such calibrations, meaning that screen-perimeters can only be
53 constructed/maintained within the confines of specialized institutions: thereby nullifying one of their
54 principle attractions.

55 The challenge of calibration is further complicated by the fact that digital displays tend to be spatially
56 non-uniform²¹. This can be due to multiple factors (e.g., wear, uneven power distribution, imperfect
57 panel fitting leading to 'light leak' at the edge of the screen, etc.), and means that the input-output
58 function is liable to differ for every pixel. As a result, the input-output must be measured at multiple
59 locations to ensure precise stimuli. For example, in the present experiment we calibrated the screen
60 using a uniform grid of 10 by 8 locations, and interpolated between locations to provide coverage of
61 every pixel. With a 10-bit display (i.e., 1024 luminance levels), this implies a total of 81,920
62 measurements, each of which should ideally be repeated multiple times for verification. Even using a
63 programmable photometer, this process can take many hours, during which time a human operator
64 needed to be present to manually reposition the photometer as required (though see Ref~[^{22,23}]). Such
65 an involved process of calibration is clearly incompatible with the notion of a perimeter that is cheap
66 or widely available. In short, the need for calibration means that while anybody can acquire the
67 technology necessary to perform screen-based perimetry, few people are in a position to use it
68 appropriately.

69 Fortunately, recent advances in commercial hardware afford a possible solution. Demand from the
70 medical and creative sectors has led a number of companies to create ‘self-calibrating’ screens, such
71 as the EIZO CG277 (EIZO Corporation, Hakusan, Ishikawa, JP). These displays contain an integrated
72 photometer, meaning that they are able to autonomously measure their own input-output function
73 (i.e., at first launch, or overnight when not in use). If this integrated calibration were perfectly
74 accurate, then it would appear to follow, trivially, that accurate screen-based perimetry should be
75 possible, without the need for extraneous calibration equipment or technical skills.

76 In practice, however, two key concerns remain. First, the self-calibration of the EIZO CG277 is
77 imperfect. The integrated photometer is only able to sample a single screen location. And while the
78 panel itself is designed to be highly uniform (each panel undergoes a factory calibration and is issued
79 with a certificate of uniformity), substantial variations in light level exist across the screen (see
80 *Results*). Whether these imperfections are great enough to affect perimetric measurements depends
81 on a wide range of factors, including the amount of intrinsic noise in the test itself^{24,25}. Second, there
82 are lingering question marks about screen-perimetry in general, and whether even a perfectly
83 calibrated screen can ever provide accurate perimetric data, given, for example, the lack of control
84 over the test environment, or the distance of the observer’s head from the screen.

85 The present work examined both of these questions. To examine whether an imperfect self-calibration
86 adds measurable noise or bias to perimetric data, $N=32$ normally-sighted participants performed a
87 screen-based threshold-perimetry test multiple times. In one condition (“Auto”), the display panel was
88 calibrated using only the screen’s integrated photometer. In a second condition (“Manual”) the display
89 panel underwent a traditional manual calibration procedure, including extensive measurements and
90 validation using several third-party photometers (see *Methods*). The experimental hypothesis was that
91 the two sets of results would not differ, either in terms of accuracy, test-retest reliability, or speed.
92 Furthermore, to examine the validity of screen-perimetry in general, all participants also went a full
93 visual field assessment using an established reference standard (Humphrey Field Analyzer; HFA).
94 Ideally, the results of neither screen-based test should deviate from those from the HFA.

95 2. METHODS

96 2.1. Overview

97 Thirty-two adults with normal vision completed six automatic static threshold perimetry examinations
98 within a single session: (i) twice using a commercial Humphrey Field Analyzer (HFA); (ii) twice using a
99 novel screen-based perimeter with a standard photometric calibration applied (“Manual”); and (iii)
100 twice using the same screen-based perimeter, but calibrated using only the screen’s own integrated
101 photometer (“Auto”).

102 All six examinations were interleaved within a single session (ABCABC), with the starting method
103 randomly counterbalanced between subjects. All examinations were carried out monocularly in one
104 eye, with test-eye counterbalanced between subjects (16 left-eye only, 16 right-eye only). Testing was
105 carried out in a quiet room, under mesopic lighting (HFA: 0.09 lx; Eye-tracking: 0.07 lx), as measured
106 using an Amprobe LM-120 Light Meter (Danaher Corporation, Washington D.C., USA).

107 2.2. Participants

108 Participants were 32 healthy adults (23 female), aged 19.4–31.0 years ($M = 24.4$; $SD = 3.6$), with no
109 previous experience of visual field testing. Normal vision was assessed by ETDRS recognition acuity (all
110 ≤ 0.3 logMAR; $M = 0.07$) and self-report medical histories. Contact lens wearers were included in the
111 study but glasses were not allowed.

112 All participants were recruited through the UCL Psychology Department subject pool, and received
113 £8/h compensation for their time. The research was carried out in accordance with the Declaration of
114 Helsinki, and was approved by the UCL Ethics Committee. Informed, written consent was obtained
115 prior to testing.

116 2.3. HFA Apparatus and Procedure

117 HFA testing was performed using a Humphrey Field Analyzer II: Model 740i (Carl Zeiss Meditec Inc.,
118 Dublin, CA, USA). Participants completed a standard 24-2 Threshold test, using Goldmann III/0.43°
119 stimuli, a SITA Standard thresholding algorithm, and a 10cd/m² white background.

120 2.4. Screen-based Perimeter Apparatus and Procedure

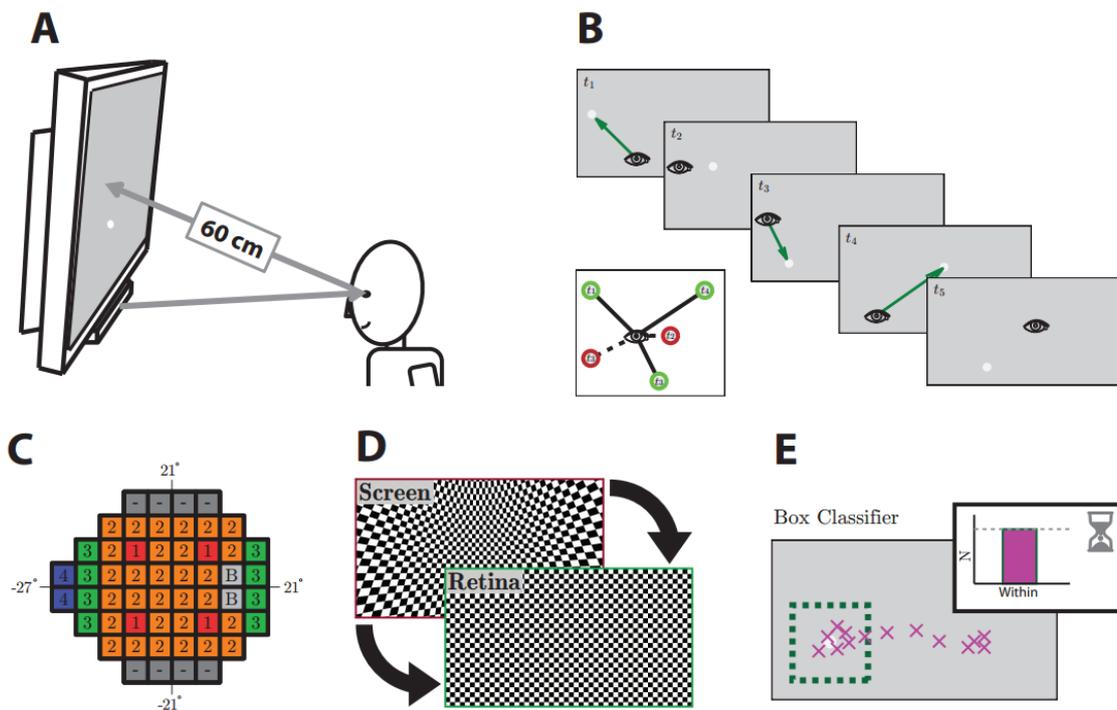
121 The novel screen-based perimeter was an eye-movement perimeter, in which a remote eye-tracker
122 was used to position dots of light on a screen, relative to the current point of fixation, and in which
123 participants responded by making eye-movements towards seen targets (see Fig 1). Source code for
124 an early version of the test is freely available online (<https://github.com/petejonze/visfield>).

125 This test is similar to the ‘Eyecatcher’ test that we reported previously⁹, and used some of the same
126 hardware and code. Eyecatcher, however, performs a quick suprathreshold evaluation of the para-
127 central field, whereas the test reported here performed a full threshold evaluation across a modified
128 24-2 grid. The novel screen-based test is also similar in principle to other eye-movement perimeters^{10–}
129 ¹⁶. Its key features are as follows.

130 The screen-based perimetry hardware are shown in Figure 1A, and consisted principally of: an
131 ordinary desktop computer, running Windows 7 (Microsoft Corporation, Redmond, WA, USA); a 10-bit
132 LCD (IPS) monitor (EIZO CG277; EIZO Corporation, Hakusan, Ishikawa, JP); a 10-bit graphics card
133 (Nvidia Quadro K620; Nvidia Corporation, Santa Clara, CA, USA); and a near-infrared remote eye-
134 tracker (Tobii EyeX; Tobii Technology, Stockholm, Sweden). Stimuli were generated in MATLAB R2014b
135 (32-bit; The MathWorks, Natick, USA) using Psychtoolbox v3.0.11^{26,27}. Eye-tracking data were retrieved

136 from the Tobii EyeX engine (v1.2.0) using custom C code, and were processed using custom MATLAB
 137 code.

138 As with the HFA, targets were 0.43° diameter (Goldmann III) circles of variable luminance, presented
 139 on a 24-2 grid, against a 10 cd/m² white background (Fig 1C). However, unlike with the HFA,
 140 participants responded by making an eye-movement towards the target location, rather than by
 141 pressing a button (see Fig 1E). Also, as shown in Figure 2C, the four test-points from the top and
 142 bottom of the standard 24-2 grid were omitted due to the dimensions of the screen. Finally, since the
 143 HFA's ('SITA-standard') thresholding algorithm is proprietary technology, the ZEST algorithm²⁸⁻³¹ was
 144 used to adapt stimuli and determine detection thresholds. As recommended by Turpin and
 145 colleagues³¹, the ZEST prior was a bimodal probability density function, constructed by combining
 146 normative data for healthy and glaucomatous eyes. The likelihood function was a cumulative
 147 Gaussian, with a fixed slope of $\sigma = 1.25$, and a variable mean of $\mu = \langle 0, 1, 2, \dots, 34 \rangle$. The growth pattern
 148 is given in Figure 2C. A dynamic termination criterion was used^{32,33}, in which the spread of the
 149 estimated posterior function was required to have a standard deviation of $\sigma \leq 1.5$ dB.



150
 151 **Figure 1.** Screen-perimetry apparatus and procedures. **(A)** Hardware. Stimuli were presented on a 59.7 x 33.6 cm (2560 x 1440 pixel)
 152 LCD screen, viewed at a distance of 60 cm (i.e., 52.9° x 31.3° visual angle). An eye-tracker (Tobii EyeX) was mounted below the screen,
 153 and was used to measure gaze-location and head-position. This allowed stimuli to be localized in size and location on the retina, and to
 154 evaluate eye-movement responses. Head position and gaze-location were unconstrained, and accounted for in software. **(B)** Example
 155 trial sequence. Goldmann III targets of variable intensity were placed relative to the current point of fixation. **(C)** Test-grid and growth-
 156 pattern. Targets were located on a 24-2 perimetric grid. The ZEST algorithm tested groups of locations in four discrete 'waves', following
 157 the order shown. Each point was tested independently; however normative data and estimates from earlier test-points were used to
 158 inform starting values. The blind-spot points ("B") were tested throughout, independent of the growth pattern. **(D)** Stimulus warping. A
 159 corrective distortion was applied (in software) to ensure a constant stimulus size/shape on the retina, despite the use of tangent-screen
 160 presentation. For example, stimuli in the far periphery of the screen were physically larger (in pixels) than those presented centrally, and
 161 were spatially distorted to maintain a circular shape on the retina. Stimuli were also scaled as necessary based on viewing distance (i.e.,
 162 since head-position was not constrained). **(E)** Eye-movement classification. Responses were deemed a 'Hit' if N gaze-estimates (purple
 163 crosses - sampled at 50 Hz from the eye-tracker) fell within a $D^\circ \times D^\circ$ box centered on the target location (green dashed line), within R
 164 seconds of stimulus onset. The parameters N , D , and R varied as a function of stimulus eccentricity (e.g., for a target at $\langle +9^\circ, +9^\circ \rangle$: $N = 6$,
 165 $D = 2.77$, $R = 1.62$).

166 **2.5. Screen Calibration**

167 For the “Manual” calibration condition: empirical measurements of luminance were performed using
 168 a ColorCal MK II colorimeter (Cambridge Research Systems, Cambridge, UK). To correct for spatial non-
 169 stationarities, input-output functions were measured independently for 80 (8 x 10) uniformly-spaced
 170 screen locations. Two-dimensional tensor-product linear-interpolation was then used to compute the
 171 appropriate calibration for every screen location (pixel). Calibrations were validated using a Minolta
 172 CS-100 colorimeter (Minolta Camera Co., Osaka, Japan). Key outcomes of this calibration process are
 173 reported in the results.

174 For the “Auto” calibration condition: screen calibration was carried out autonomously at a single
 175 location, using the EIZO CG277’s integrated photometer. The photometer was interfaced using custom
 176 C/Matlab code. This code was written for present work, and is freely available online under a non-
 177 commercial license (GNU GPL v3.0): <https://github.com/peteionze/myEIZOSensor>.

178 **2.6. Measurement and reporting of sensitivity**

179 Estimated contrast sensitivity for each stimulus location was quantified as Differential Light Sensitivity
 180 (DLS): the smallest detectable difference in luminance, ΔL , between the target luminance, L_{targ} , and
 181 the background luminance, L_B . With both the novel perimetry measure, and the HFA reference
 182 measure, the value of L_B was fixed at 10 cd/m². The value of ΔL varied trial-by-trial according to an
 183 adaptive algorithm (ZEST or SITA Standard), in order to find the smallest value of ΔL that could be
 184 reliably detected on 50% of trials: ΔL_{jnd} .

185 Following standard perimetric convention^{1,34,35}, DLS values are reported in units of signal attenuation
 186 on an inverted log-scale:

$$DLS_{dB} = 10 \log_{10} \left(\frac{\Delta L_{max}}{\Delta L_{jnd}} \right) \quad (1)$$

187 where ΔL_{max} is the greatest displayable stimulus pedestal. For ease of comparison, this value was
 188 scaled identically for both the HFA and the novel screen-based test, using the maximum displayable
 189 pedestal of the novel test ($\Delta L_{max} = 225$ cd/m²). For all tests, DLS values therefore varied from 0 dB and
 190 34 dB, with higher values indicating greater sensitivity.

191 **3. RESULTS**

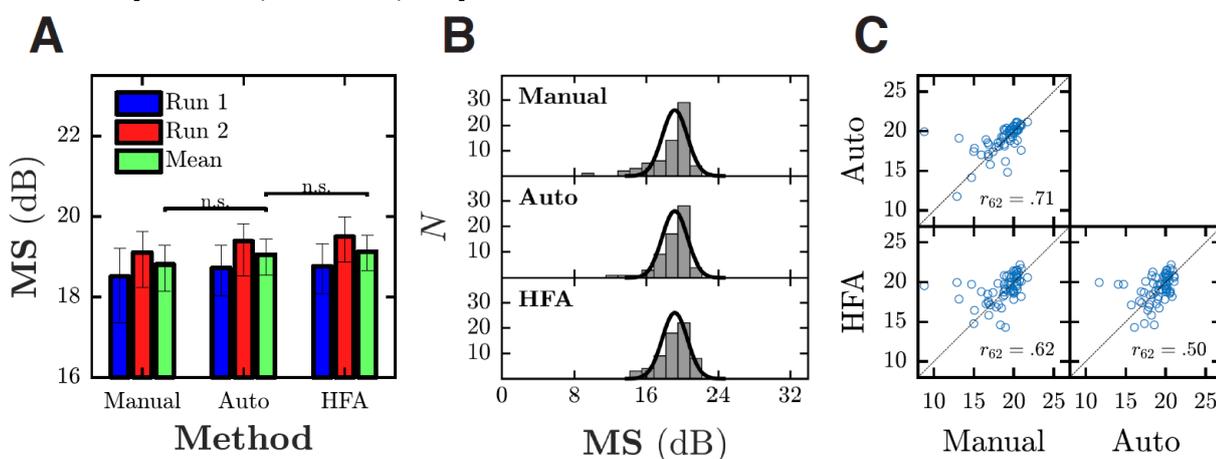
192 **3.1 Initial photometric characterization: Manual vs Automatic (internal) calibration**

193 As expected, the input-output functions from the manual and automatic luminance calibrations were
 194 virtually indistinguishable when measurements were made at the same location, i.e., at the single
 195 region of the screen that the integrated photometer samples from [Pearson Correlation; $r_{1022} \approx 1.0$, P
 196 $\ll 0.001$]. However, the automatic calibration assumes uniformity across the screen, and the display's
 197 spatial uniformity --- while far superior to a standard LCD monitor --- was only approximate. The
 198 standard deviation in luminance across the screen was $\sim 3\%$: a magnitude similar to the Just
 199 Noticeable Difference for human contrast discrimination³⁶. And some areas of the screen were $\sim 17\%$
 200 more intense than others (see §3.6 for graphical illustration). Whether these imperfections are large
 201 enough to affect perimetric measurements remained an empirical question, however, which we turn
 202 to next.

203 **3.2 Accuracy: Mean differential light sensitivity (MS)**

204 Mean Sensitivity (MS) estimates for the three test conditions are shown in Figure 2. Group-mean MS
 205 values were 18.8 dB (Manual), 19.1 dB (Auto), and 19.1 dB (HFA). None of these values were
 206 significantly different from each other [3 paired t -tests; all $t_{63} \leq 1.16$; $P \geq 0.251$], or from the normative
 207 value of 19.2 dB reported previously by Brenton and Phelps³⁷ [3 one-sample t -tests; all $t_{63} \leq 1.40$; $P \geq$
 208 0.167]. (NB: Breton and Phelps report a peak value of 30.7 dB. However, following perimetric
 209 convention this value was rescaled to 19.2 dB based on the maximum luminance output of our screen;
 210 see Eq 1). At an individual level, the results of the three tests were also positively correlated [3
 211 Pearson correlations; $r_{62} \geq 0.50$, $p \leq 0.001$], with participants who scored higher on one test condition
 212 tending to score higher on other conditions too (Fig 2C). Taken together, these findings indicate that
 213 all three test conditions gave quantitatively similar results.

214 As is common in perimetry³⁸, a small but consistent practice effect was observed in all three test-
 215 conditions (Fig 2A), with mean sensitivity increasing between runs one and two by an average of 0.60
 216 dB (Manual), 0.66 dB (Auto) and 0.73 dB (HFA). This difference was significant for the Auto [$t_{31} = 2.88$,
 217 $P = 0.007$] and HFA conditions [$t_{31} = 4.43$, $P < 0.001$], though did not reach significance in the Manual
 218 condition [$t_{31} = 1.55$, $P = 0.131$, *n.s.*].



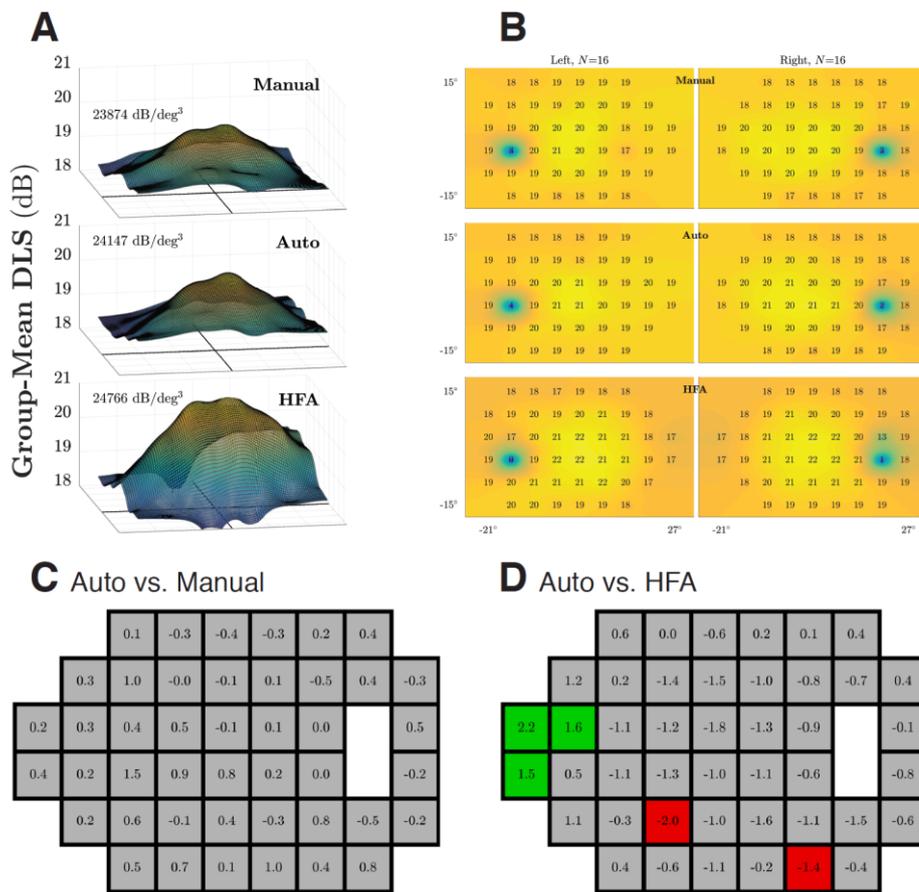
219 **Figure 2.** Estimates of mean sensitivity, MS, for each of the three test conditions. (A) Group-mean data [$\pm 95\%$ CI], for the first and
 220 second run of each test, and the mean average of the two. (B) Histograms showing the distribution of results for all 64 (32x2) tests in
 221 each condition. Gaussian distributions show previously published normative data for the HFA³⁷ (21 healthy adults, aged 20-29 years),
 222 scaled to the same units as the present data using Eq 1. (C) Scatter-plot showing within-subject correlations. Markers show MS scores for
 223

224 individual tests. The dashed black line is the identity line: if all data fell along this line then that would indicate perfect agreement
225 between the two tests.

226 **3.3 Accuracy: Pointwise sensitivity (PWS)**

227 With both Manual and Auto calibration, the screen-based perimeter was able to detect normal
 228 variations in visual sensitivity across the visual field, with gradations in sensitivity evident between
 229 paracentral and peripheral test locations (the ‘Hill of Vision’; Fig 3A and 3B). Furthermore, the screen-
 230 based perimeter exhibited enough spatiotemporal specificity to isolate the physiological blind-spot
 231 (Fig 3B). Thus, sensitivity estimates at $\langle \pm 15^\circ, -3^\circ \rangle$ were significantly lower than at any of the
 232 surrounding locations, both in the Manual [8 paired t -tests; all $t_{31} \geq 15.39$, all $P \ll 0.001$], and Auto test
 233 conditions [all $t_{31} \geq 16.86$, all $P \ll 0.001$].

234 To formally assess whether there was any systematic difference in pointwise sensitivity (PWS)
 235 estimates between conditions, we used independent Wilcoxon rank-sum tests to test for difference in
 236 each of the 44 test locations, both when comparing Auto vs Manual calibration (Fig 3C), and Auto vs
 237 HFA (Fig 3D). The results are shown in Figure 3C-D, with significant difference ($P < 0.01$) highlighted in
 238 green (Auto higher) and red (Auto lower). When comparing Auto vs Manual calibration, no significant
 239 PWS differences were observed, further confirming that the Auto calibration has no measurable effect
 240 on accuracy. When comparing between Auto and HFA, there was a general tendency for the HFA to
 241 report higher sensitivities at more central locations, and lower sensitivities more peripheral/nasal
 242 locations, as illustrated by the steeper gradients in Figure 3A. However, as shown in Figure 3D, these
 243 pointwise differences were only significant at 5 of the 44 individual locations (11.4%), and may be due,
 244 in part, to the number of (multiple) comparisons.

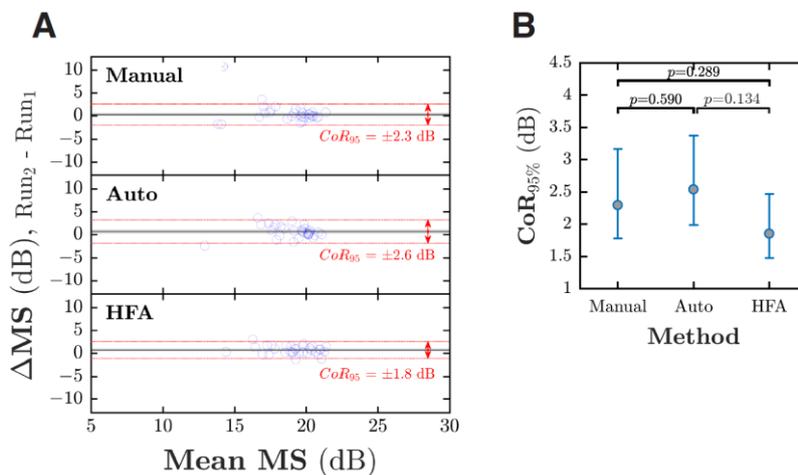


245 **Figure 3.** Distribution of pointwise visual sensitivity (DLS) estimates across the visual field. **(A)** Three dimensional ‘Hill of Vision’ plots for
 246 each of the three test conditions. Surfaces fitted using spring-regularized nearest-neighbor interpolation, and then smoothed
 247 using a moving-average rectangular filter. The two blind-spot locations were excluded from fits. A top-down view of these hills is given in
 248 Panel B. **(B)** Group-mean DLS values for each eye (columns) and test-condition (rows). **(C)** Differences in DLS values between the Auto
 249 and Manual conditions ($DLS_{Auto} - DLS_{Manual}$). Shading indicates bootstrapped significance-tests (Red: Auto lower; Green: Auto higher;
 250 Grey: no significant difference; $\alpha = 0.01$). **(D)** Same as (C), comparing Auto and HFA conditions ($DLS_{Auto} - DLS_{HFA}$).

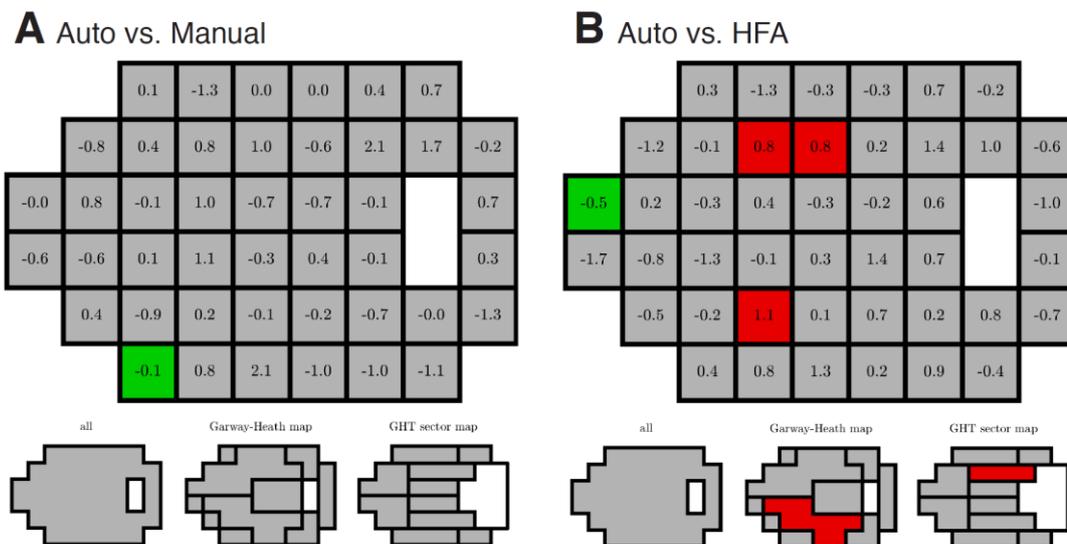
252 **3.4 Reliability: Test-retest repeatability**

253 Bland-Altman analyses³⁹ were used to assess the reliability of the mean sensitivity (MS) estimates
 254 (Figure 4A). Across the three conditions, the 95% Coefficient of Repeatability [CoR₉₅] was 2.3 dB
 255 (Manual), 2.6 dB (Auto), and 1.8 dB (HFA). Using a bootstrapping procedure analogous to a *t*-test,
 256 these differences were found to be non-significant, both when comparing Auto vs. Manual (*P* = 0.600),
 257 and when comparing Auto vs. HFA (*P* = 0.134). These findings indicate that all three test conditions
 258 were similarly reliable (precise), and that using the display-screen’s internal calibration did not reduce
 259 the overall reliability of the screen-based perimetry test.

260 To evaluate reliability at the level of individual PWS estimates, and to assess whether reliability varied
 261 across the visual field, these Bland-Altman analyses were repeated for each of the 44 individual test
 262 locations (Fig 5). When comparing Auto vs. Manual calibration, the CoR₉₅ values were observed to
 263 differ significantly (*P* < 0.01) at one location only. This single difference was likely due to chance (i.e.,
 264 given the α = 0.01 significance level and the number of multiple comparisons). When comparing Auto
 265 vs. HFA, the reliability of the novel screen-based test was significantly lower for 3 central locations,
 266 and significantly higher for one peripheral location.



267 **Figure 4.** Test-retest repeatability for MS values. **(A)** Bland-Altman plots of mean sensitivity. Each marker represents a single participant.
 268 Dashed red lines indicate the 95% limits of agreement ($\mu \pm \text{CoR}_{95}$). In the Manual condition, one point [black cross] was excluded as an
 269 outlier. **(B)** Comparison of CoR₉₅ values for mean sensitivity (MS). Error bars indicate bootstrapped 95% confidence intervals. There
 270 were no significant differences in repeatability between any of the three measures.
 271



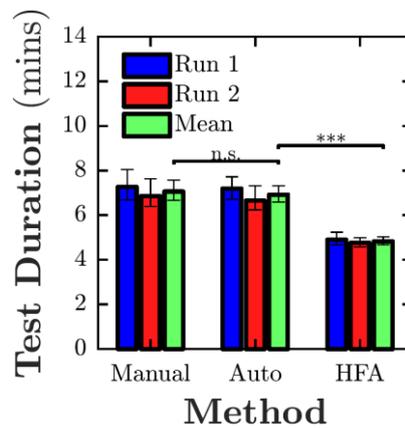
272 **Figure 5.** Differences in test-retest repeatability for individual PWS values. **(A)** Comparison of Auto vs. Manual. The top panel shows
 273 differences in CoR₉₅ for each location on the test grid (CoR_{Auto} – CoR_{Manual}). The bottom panel shows the same data grouped into various
 274 sub-regions, based on three published visual-field maps⁴⁰. **(B)** Same as (A), comparing Auto vs. HFA measures (CoR_{Auto} – CoR_{HFA}).
 275

276 **3.5 Test duration**

277 Grand mean test durations were 7.1 min (Manual), 6.92 min (Auto), and 4.83 min (HFA). The
 278 difference in test duration was significant when comparing Auto vs. HFA [paired t -tests; $t_{63} = 11.43$, $P \ll$
 279 0.001], but not when comparing Manual vs. Auto [$t_{63} = -0.88$, $P = 0.385$, $n.s.$]. This indicates that the
 280 screen perimeter was slower than the HFA, but that using the display-screen’s internal calibration did
 281 not affect the speed of the screen-based perimetry test.

282 The difference versus the HFA is most likely due to the fact screen-based perimeter contained a large
 283 number of additional trials that the HFA did not, including trials to assess false-positive and false-
 284 negative rates, ‘calibration’ trials (to calibrate the eye-tracker), and ‘refixation’ trials (to allow locations
 285 to be tested if they would otherwise fall off the edge of the screen). Conversely, it is important to note
 286 the HFA measured an additional 8 test locations that the screen-based perimeter did not (see Fig 1C)

287 As with the sensitivity scores presented previously (Fig 2A), there was some indication of a practice
 288 effect on test durations, with durations decreasing across repetitions in all three conditions. The
 289 mean-average reduction was 22 seconds. However, these differences were not significant for any of
 290 the three conditions [3 paired t -tests; $t_{31} \leq 1.73$, $P \geq 0.094$, $n.s.$].



291 **Figure 6.** Group-Mean [±95% CI], test durations, for each of the three test conditions. Same format as the MS values presented in
 292 Figure 2A.
 293

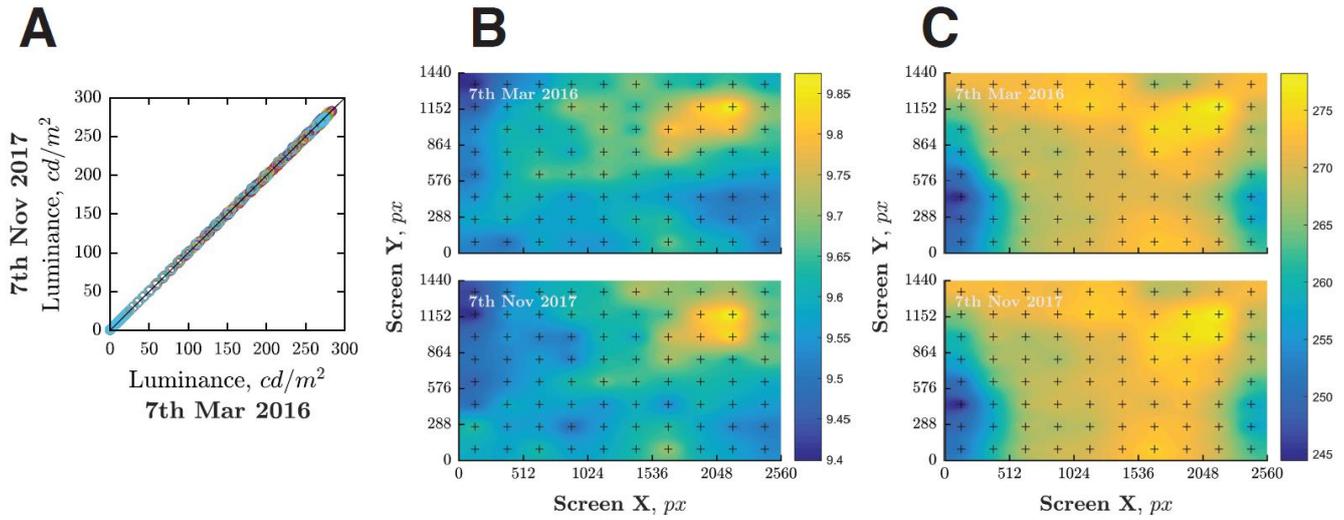
294 **3.6 Photometric screen characterization: 20 months later**

295 The foregoing results suggested that there is no measurable change in accuracy, reliability, or test
 296 duration when using the display-screen’s internal calibration (“Auto”) versus full manual calibration
 297 (“Manual”), and that both methods provide broadly similar perimetric data to the reference standard
 298 (HFA).

299 However, when this study was conducted the display panel was less than one year old. It is possible
 300 that the uniformity of the screen may deteriorate over time (e.g., due to natural wear and tear). This
 301 would compromise the internal calibration, as it measures only from a single location on the screen,
 302 and may ultimately introduce noticeable measurement error.

303 To assess whether this is the case, we made further photometric measurements of the screen after 20
 304 months of regular use. As shown in Figure 7, there was near perfect agreement between the two sets
 305 of measurements [Pearson Correlation; $r_{78} \approx 1.0$, $P \ll 0.001$]. The mean percentage change was 0.4%,
 306 and the spatial pattern of results was highly conserved both at low (Fig 7B) and high (Fig 7C) intensity
 307 levels.

308 In short, the luminance properties of the display screen (EIZO CG277) remained highly stable after 20
 309 months of regular use. There is therefore no reason to suppose that the accuracy or reliability of any
 310 perimetric measurements would decrease over a reasonable period of use.



311
 312 **Figure 7.** Photometric luminance measurement for the display screen, before and after 20 months of regular use. **(A)** Comparison of
 313 individual measurement at each of 80 screen locations. Measurements were made with a ColorCal MK II colorimeter. The black diagonal
 314 line denotes unity (perfect correlation). **(B)** Measurements for points close to 10 cd/m² (background intensity). Black crosses denote the
 315 test locations, fitted with a surface in the same manner as Figure 3A. The standard deviation was 0.11 cd/m²; the difference between the
 316 smallest and greatest value was 0.66 cd/m² (6.88%). **(C)** Measurements for points close to 275 cd/m² (maximum intensity). The standard
 317 deviation was 8.08 cd/m²; the difference between the smallest and greatest value was 45.63 cd/m² (17.47%).

318 **4. DISCUSSION**

319 The goal of the present study was to assess whether screen-based perimetry could be performed
320 without the need for any manual calibration, using only the integrated photometer contained within
321 'professional' commercial monitors (i.e., and assuming uniformity across the screen). The photometric
322 data showed that this internal calibration was imperfect, and resulted in measurable deviations in
323 luminance across the screen. However, these deviations had no measurable effect on the behavioral
324 results. Specifically: there were no detectable differences in mean sensitivity (MS) estimates,
325 pointwise sensitivity (PWS) estimates, test-retest repeatability, or test duration (i.e., when comparing
326 screen-based measurements made with and without full manual calibration). This suggests that self-
327 calibrating screens are sufficiently accurate to support accurate threshold perimetry measurements.
328 More generally, the present data provide further evidence that screen-based perimeters – and eye-
329 movement perimetry in particular – could provide viable alternatives to traditional standard
330 automated perimeters [SAPs]. The accuracy and precision of the novel screen-based test were
331 indistinguishable from those made using the HFA (the current clinical reference standard), although
332 the HFA was ~30% faster. This may be of particular importance in situations where access to SAP
333 assessment is limited, either by physical/cognitive impairment, or on the grounds of portability or cost
334 (*see below*).

335 **4.1 Study Limitations**

336 The present data were derived from healthy observers, not patients. They should therefore not be
337 taken as hard evidence for the general efficacy of screen-based perimetry in clinical practice (for this,
338 see Refs [3-5,9-12,18,41]). However, the use of healthy observers is unlikely to have affected the
339 conclusions of the present study. A system which is capable of observing reliable differences in the
340 healthy eye between central and peripheral locations (see Fig 3A) should be more than capable of
341 detecting/monitoring 'clinically significant' deviations (i.e., which tend to be far larger). In fact, any
342 residual calibration-error is likely to be of even less of a concern for patients than in the present
343 cohort of observers, since measurement variability is known to increase as a function of decreased
344 sensitivity^{42,43}, and so would be expected to further swamp any effects of stimulus imperfections.

345 A second potential concern is the fact that the screen-based perimeter used in the present study
346 differed in several ways to the reference device (HFA). Method of response differed (eye-movements
347 vs. button press), their underlying psychophysical algorithms differed (ZEST vs. the proprietary SITA
348 algorithm), and the HFA uses mandatory fixation targets and head-restraints, whereas our novel
349 screen perimeter used eye- and head-tracking to perform gaze-contingent stimulus placement and
350 dynamic size-scaling. They did, however, measure the same variable (differential light sensitivity to
351 Goldmann III targets across a 24-2 grid), and their results were numerically scaled to be directly
352 comparable (see §2.6). The fact that the output data were in such close agreement *despite* these
353 technical differences we take as particularly strong evidence that screen-based perimetry is capable of
354 replicating 'gold standard' measurements.

355 It should also be noted that none of the present findings would be expected to differ had a response
356 button been used instead for the screen-based perimetry. We chose to concentrate on eye-movement
357 perimetry due to its greater ease-of-use and patient-satisfaction (see Ref^[9]), and also because in we
358 are interested in applying the technology in future to individuals who are unable to comply with the
359 demands of traditional, button-press perimetry: either because they are either unable to maintain
360 fixation, or because they cannot press a response button reliably. However, for the purposes of the
361 present study, no qualitative difference were observed during piloting when a button was used
362 instead.

363 **4.2 Implications, Applications, and Limitations of Screen-Based Perimetry**

364 The present results are exciting because they mean that a rigorous threshold perimetry test could one
365 day be distributed as an app. An individual with limited technical knowledge could buy the equipment
366 described in the present study today (all of which is widely commercially available), and by installing
367 the appropriate software, would possess a functioning threshold perimeter, without the need for any
368 expert knowledge, complex assembly, or costly maintenance. It is unlikely that such devices would
369 replace the specialized perimetric devices that exist currently. Instead, we see the two devices as
370 complementary. Screen-based perimeters could, for example, allow visual field assessments to be
371 carried out in non-conventional settings, such as by the bedside in the case of stroke, or out in the
372 community as part of screening or case-finding programs. Alternatively, screen-perimetry could be
373 used to provide supplemental home monitoring for chronic progressive diseases such as glaucoma,
374 with patients reporting for formal clinical evaluation if sudden deterioration were detected. Before
375 these possibilities can be realized, however, several key hurdles remain.

376 One concern is that while automatic-calibration removes one key bottleneck, scientific expertise is still
377 required to program the stimuli and create the tests. To militate against this, all of the code necessary
378 to calibrate and run the tests described in the present paper have been made freely available online
379 (see *Methods*). This is 'research grade' code, however, and in the longer term, it will be necessary to
380 convert this into a more user-friendly 'click to run' app. It will also be necessary for any results to be
381 reviewed and interpreted by a qualified clinician, for example via the sorts of cloud-based solutions
382 already under evaluation elsewhere⁴⁴.

383 A second concern is access to the requisite hardware which, though easily available in stores
384 worldwide, is not yet present in the average home or workplace. Thus, while the screen-based
385 perimeter in the present study ran on ordinary desktop computer, it used a professional-grade
386 monitor (EIZO CG277) and a professional-grade graphics card (Nvidia Quadro K620). In contrast, most
387 consumer-oriented screens are highly non-uniform, and lack the necessary bit-rate or photometric
388 sensors. In addition, the present test also employed an inexpensive an eye-tracker (Tobii EyeX) which
389 is a further prerequisite (although such technology is already being built into certain laptops and
390 monitors, and in future may be replaced with data from an ordinary webcam^{45,46}). Together, these
391 additional hardware requirements mean that we remain short of our ideal goal of a pure 'software
392 perimeter' that requires no non-standard hardware to run. We are optimistic, however, that this gap
393 will continue to diminish as the necessary technologies become increasing mainstream.

394 A third, related concern is cost. At the time of writing, the additional hardware (monitor, graphics card,
395 and eye-tracker) cost approximately £2000 in total (without tax). This is not a trivial amount of money
396 for a consumer. However, it is an order of magnitude cheaper than standard perimetric devices, and in
397 healthcare terms is similar to the price of a printed letter chart. As such, we do not envisage cost to be
398 a key limiting factor. Indeed, part of the appeal of screen-based perimeters might be their relatively
399 low-cost, particularly in developing countries where healthcare providers are not already so heavily
400 invested in 'gold standard' perimetric equipment. As discussed previously, however, we do not
401 necessarily view screen-based perimetry as a like-for-like alternative to existing specialized devices,
402 but rather as a way of expanding access to visual field assessments.

403 These outstanding limitations notwithstanding, we believe the current work marks a qualitative step
404 forward. The equipment required is cheap, easy to use, and easy to replace. And unlike with
405 traditional, dedicated perimeters, the equipment is inherently multipurpose. We have already shown,
406 for example, that the same basic hardware can also be used to perform acuity assessments⁴⁷, and we
407 routinely use the equipment described in the present study for other day-to-day tasks, such as playing
408 videos to patients in-between tests (i.e., particularly when performing pediatric assessments⁴⁸), or for
409 performing general office work. For these reasons, we believe that a simple 'plug-and-play' perimeter

410 could be a highly attractive proposition, particularly in circumstances where 'gold standard' devices
411 such as the HFA are not viable alternatives.

412 **Acknowledgments**

413 This work was supported by the NIHR Biomedical Research Centre located at (both) Moorfields Eye
414 Hospital and UCL Institute of Ophthalmology.

415 **Disclosures:**

- 416 • **H. K. Han**, None;
417 • **P. R. Jones**, None;

418 REFERENCES

- 419 1. Henson, D. B. *Visual fields*. 2000;(Butterworth-Heinemann Medical, 2000).
- 420 2. Anderson, A. J., Bedggood, P. A., Kong, Y. X. G., Martin, K. R. & Vingrys, A. J. Can Home
421 Monitoring Allow Earlier Detection of Rapid Visual Field Progression in Glaucoma?
422 *Ophthalmology* 2017;doi:https://doi.org/10.1016/j.ophtha.2017.06.028
- 423 3. Schulz, A. M., Graham, E. C., You, Y., Klistorner, A. & Graham, S. L. Performance of iPad based
424 threshold perimetry in glaucoma and controls. *Clin. Experiment. Ophthalmol.* 2017;
- 425 4. Kong, Y. X. G., He, M., Crowston, J. G. & Vingrys, A. J. A comparison of perimetric results from a
426 tablet perimeter and Humphrey field analyzer in glaucoma patients. *Transl. Vis. Sci. Technol.*
427 2016;5:2
- 428 5. Vingrys, A. J., Healey, J. K., Liew, S., Saharinen, V., Tran, M., Wu, W. & Kong, G. Y. X. Validation of
429 a Tablet as a Tangent Perimeter. *Transl. Vis. Sci. Technol.* 2016;5:3
- 430 6. Prea, S. M., Kong, Y. X. G., Mehta, A., He, M., Crowston, J. G., Gupta, V., Martin, K. R. & Vingrys,
431 A. J. Six-month Longitudinal Comparison of a Portable Tablet Perimeter With the Humphrey
432 Field Analyzer. *Am. J. Ophthalmol.* 2018;190:9–16
- 433 7. Johnson, C. A., Thapa, S., Kong, Y. X. G. & Robin, A. L. Performance of an iPad application to
434 detect moderate and advanced visual field loss in Nepal. *Am. J. Ophthalmol.* 2017;182:147–154
- 435 8. Nesaratnam, N., Thomas, P. B. M., Kirolos, R., Vingrys, A. J., Kong, G. Y. X. & Martin, K. R. Tablets
436 at the bedside-iPad-based visual field test used in the diagnosis of Intracellar
437 Haemangiopericytoma: a case report. *BMC Ophthalmol.* 2017;17:53
- 438 9. Jones, P. R., Smith, N. D., Bi, W. & Crabb, D. P. Portable Perimetry Using Eye-Tracking on a Tablet
439 Computer--A Feasibility Assessment. *Transl. Vis. Sci. Technol.* 2019;8:17
- 440 10. Murray, I. C., Fleck, B. W., Brash, H. M., MacRae, M. E., Tan, L. L. & Minns, R. A. Feasibility of
441 saccadic vector optokinetic perimetry: a method of automated static perimetry for children
442 using eye tracking. *Ophthalmology* 2009;116:2017–2026
- 443 11. Murray, I. C., Cameron, L. A., McTrusty, A. D., Perperidis, A., Brash, H. M., Fleck, B. W. & Minns,
444 R. A. Feasibility, Accuracy, and Repeatability of Suprathreshold Saccadic Vector Optokinetic
445 Perimetry. *Transl. Vis. Sci. Technol.* 2016;5:15
- 446 12. McTrusty, A. D., Cameron, L. A., Perperidis, A., Brash, H. M., Tatham, A. J., Agarwal, P. K., Murray,
447 I. C., Fleck, B. W. & Minns, R. A. Comparison of Threshold Saccadic Vector Optokinetic Perimetry
448 (SVOP) and Standard Automated Perimetry (SAP) in Glaucoma. Part II: Patterns of Visual Field
449 Loss and Acceptability. *Transl. Vis. Sci. Technol.* 2017;6:4
- 450 13. Wroblewski, D., Francis, B. A., Sadun, A., Vakili, G. & Chopra, V. Testing of Visual Field with
451 Virtual Reality Goggles in Manual and Visual Grasp Modes. *Biomed Res. Int.* 2014;2014:206082
- 452 14. Jones, P. R., Kalwarowsky, S., Rubin, G. S. & Nardini, M. Automated static threshold perimetry
453 using a remote eye-tracker. *Investig. Ophthalmol. Vis. Sci.* 2015;56:3908--3908
- 454 15. Mazumdar, D., Pel, J. M., Panday, M., Asokan, R., Vijaya, L., Shantha, B., George, R. & others.
455 Comparison of saccadic reaction time between normal and glaucoma using an eye movement
456 perimeter. *Indian J. Ophthalmol.* 2014;62:55–59
- 457 16. Pel, J. J. M., van Beijsterveld, M. C. M., Thepass, G. & van der Steen, J. Validity and Repeatability
458 of Saccadic Response Times Across the Visual Field in Eye Movement Perimetry. *Transl. Vis. Sci.*
459 *Technol.* 2013;2:
- 460 17. Glen, F. C., Baker, H. & Crabb, D. P. A qualitative investigation into patients' views on visual field
461 testing for glaucoma monitoring. *BMJ Open* 2014;4:e003996
- 462 18. Anderson, A. J., Bedggood, P. A., Kong, Y. X. G., Martin, K. R. & Vingrys, A. J. Can Home
463 Monitoring Allow Earlier Detection of Rapid Visual Field Progression in Glaucoma?
464 *Ophthalmology* 2017;doi:https://doi.org/10.1016/j.ophtha.2017.06.028
- 465 19. Brainard, D. H., Pelli, D. G. & Robson, T. Display characterization. *Encycl. imaging Sci. Technol.*
466 2002;

- 467 20. Metha, A. B., Vingrys, A. J. & Badcock, D. R. Calibration of a color monitor for visual
468 psychophysics. *Behav. Res. Methods* 1993;25:371–383
- 469 21. Ghodrati, M., Morris, A. P. & Price, N. S. The (un)suitability of modern liquid crystal displays
470 (LCDs) for vision research. *Front. Psychol.* 2014;6:303
- 471 22. Perperidis, A., Murray, I., Brash, H., McTrusty, A., Cameron, L., Fleck, B. & Minns, R. Correcting
472 LCD luminance non-uniformity for threshold Saccadic Vector Optokinetic Perimetry (SVOP).
473 in *Eng. Med. Biol. Soc. (EMBC), 2013 35th Annu. Int. Conf. IEEE* 2013;1636–1639(2013).
- 474 23. Kimpe, T., Xthona, A., Matthijs, P. & De Paepe, L. Solution for nonuniformities and spatial noise
475 in medical LCD displays by using pixel-based correction. *J. Digit. Imaging* 2005;18:209–218
- 476 24. Eijkman, E. & Vendrik, A. J. H. Can a sensory system be specified by its internal noise? *J. Acoust.*
477 *Soc. Am.* 1965;37:1102–1109
- 478 25. Lu, Z. L. & Doshier, B. A. Characterizing observers using external noise and observer models:
479 assessing internal representations with external noise. *Psychol. Rev.* 2008;115:44–82
- 480 26. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* 1997;10:433–436
- 481 27. Pelli, D. G. The VideoToolbox software for visual psychophysics: Transforming numbers into
482 movies. *Spat. Vis.* 1997;10:437–442
- 483 28. Vingrys, A. J. & Pianta, M. J. A new look at threshold estimation algorithms for automated static
484 perimetry. *Optom. Vis. Sci.* 1999;76:588–595
- 485 29. King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C. & Supowit, A. Efficient and unbiased
486 modifications of the QUEST threshold method: theory, simulations, experimental evaluation
487 and practical implementation. *Vision Res.* 1994;34:885–912
- 488 30. Turpin, A., McKendrick, A. M., Johnson, C. A. & Vingrys, A. J. Development of efficient threshold
489 strategies for frequency doubling technology perimetry using computer simulation. *Investig.*
490 *Ophthalmol. Vis. Sci.* 2002;43:322–331
- 491 31. Turpin, A., McKendrick, A. M., Johnson, C. A. & Vingrys, A. J. Properties of perimetric threshold
492 estimates from full threshold, ZEST, and SITA-like strategies, as determined by computer
493 simulation. *Invest. Ophthalmol. Vis. Sci.* 2003;44:4787–4795
- 494 32. Anderson, A. J. Utility of a dynamic termination criterion in the ZEST adaptive threshold
495 method. *Vision Res.* 2003;43:165–170
- 496 33. McKendrick, A. M. & Turpin, A. Advantages of terminating Zippy Estimation by Sequential
497 Testing (ZEST) with dynamic criteria for white-on-white perimetry. *Optom. Vis. Sci.*
498 2005;82:981–987
- 499 34. Weijland, A., Fankhauser, F., Bebie, H. & Flammer, J. *Automated Perimetry, 5th Edition.*
500 2004;(2004).
- 501 35. Schiefer, U., Pätzold, J., Dannheim, F., Artes, P. & Hart, W. Konventionelle Perimetrie. Teil 1
502 Einführung--Grundbegriffe. [Conventional techniques of visual field examination. Part I:
503 Introduction--basics]. *Ophthalmologe* 2005;102:627–646
- 504 36. Pelli, D. G. & Bex, P. Measuring contrast sensitivity. *Vision Res.* 2013;90:10–14
- 505 37. Brenton, R. S. & Phelps, C. D. The normal visual field on the Humphrey field analyzer.
506 *Ophthalmologica* 1986;193:56–74
- 507 38. Heijl, A., Lindgren, G. & Olsson, J. The effect of perimetric experience in normal subjects. *Arch.*
508 *Ophthalmol.* 1989;107:81–86
- 509 39. Bland, J. M. & Altman, D. G. Measuring agreement in method comparison studies. *Stat.*
510 *Methods Med. Res.* 1999;8:135–160
- 511 40. Boden, C., Chan, K., Sample, P. A., Hao, J., Lee, T.-W., Zangwill, L. M., Weinreb, R. N. &
512 Goldbaum, M. H. Assessing visual field clustering schemes using machine learning classifiers in
513 standard perimetry. *Invest. Ophthalmol. Vis. Sci.* 2007;48:5582–5590
- 514 41. Taylor, V., Glaze, S., Unwin, H., Bowman, R., Thompson, G. & Dahlmann-Noor, A. Saccadic vector
515 optokinetic perimetry in children with neurodisability or isolated visual pathway lesions:

- 516 observational cohort study. *Br. J. Ophthalmol.* 2016;100:1427–1432
- 517 42. Henson, D. B., Chaudry, S., Artes, P. H., Faragher, E. B. & Ansons, A. Response variability in the
518 visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes.
519 *Invest. Ophthalmol. Vis. Sci.* 2000;41:417–421
- 520 43. Heijl, A., Lindgren, A. & Lindgren, G. Test-retest variability in glaucomatous visual fields. *Am. J.*
521 *Ophthalmol.* 1989;108:130–135
- 522 44. Rono, H. K., Bastawrous, A., Macleod, D., Wanjala, E., DiTanna, G., Weiss, H. A. & Burton, M. J.
523 Smartphone-based screening for visual impairment in Kenyan school children: a cluster
524 randomised controlled trial. *Lancet Glob. Heal.* 2018;6:e924–e932
- 525 45. Sewell, W. & Komogortsev, O. Real-time eye gaze tracking with an unmodified commodity
526 webcam employing a neural network. in *CHI'10 Ext. Abstr. Hum. Factors Comput. Syst.*
527 2010;3739–3744(2010).
- 528 46. San Agustin, J., Skovsgaard, H., Hansen, J. P. & Hansen, D. W. Low-cost gaze interaction: ready to
529 deliver the promises. in *CHI'09 Ext. Abstr. Hum. Factors Comput. Syst.* 2009;4453–4458(2009).
- 530 47. Jones, P. R., Kalwarowsky, S., Atkinson, J., Braddick, O. J. & Nardini, M. Automated measurement
531 of resolution acuity in infants using remote eye-tracking. *Invest. Ophthalmol. Vis. Sci.*
532 2014;55:8102–8110
- 533 48. Wilson, C. E. in *Pediatr. Ophthalmol. Curr. thought a Pract. Guid.* (ed. Edward M. Wilson Richard
534 Saunders, T. R.) 2008;1–6(Springer Science & Business Media, 2008).
- 535