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 Han[1], 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

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

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

1. INTRODUCTION

Assessment of the visual field via static threshold perimetry is a key element of modern ophthalmic assessments, where it is used routinely to diagnose and monitor common eye-diseases such as glaucoma and diabetes[1]. Traditionally, perimetry is performed using specialized devices such as the Humphrey Field Analyzer (Carl Zeiss Meditec Inc., Dublin, CA, USA) or Octopus Perimeter (Haag-Streit AG, Köniz, Switzerland): large, cumbersome machines in which the user places their head inside a white bowl, into which lights of variable intensity are project.

Recently, however, there has been a proliferation of 'screen-based' perimeters. These include portable, tablet devices[2–9], as well as eye-movement perimeters that use remote eye-tracking technology to position stimuli on the retina and record saccadic responses[9–16]. Although not without their drawbacks, such 'disruptive' devices have a number of potential advantages over tradition perimeters in terms of comfort and cost (see *Discussion: Implications and Applications*). One of the principle attractions of screen-based perimeters is that they use relatively ordinary commercial technology that is widely available and easily replaceable. The implication is that they could therefore be easily set-up and maintained in community settings or across the developing world. It is even conceivable that perimetry could become an 'app', which users download to run in the comfort of their own home, thereby conferring substantial benefits in terms of patient satisfaction[17] and financial savings[18].

In reality, however, the need for luminance calibration remains a significant impediment to the widespread use of screen-based perimeters. Thus, with a digital display, the commands that are sent to the screen are unitless values (e.g., numbers between 0--1024), but what we wish to manipulate is the luminance of the stimulus on the screen (e.g., in candelas per meter squared; cd/m^2). Perimetry therefore requires us to know how a given input (Command Level) translates to a given output (Luminance). This is the input-output function of the display device, and depends on a wide range of factors, including the model/make of the screen, its current settings, the model/make of the graphics card, the temperature of the room, how long the display has been active for, and what else is currently being displayed on the screen[19,20]. Because of this complexity, the input-output function can only be determined empirically, and must be measured for each individual device. Realistically, most of the potential operators of a screen-based perimeter --- be they patients or clinicians --- lack the necessary time, training, or equipment to perform such calibrations, meaning that screen-perimeters can only be constructed/maintained within the confines of specialized institutions: thereby nullifying one of their principle attractions.

The challenge of calibration is further complicated by the fact that digital displays tend to be spatially non-uniform[21]. This can be due to multiple factors (e.g., wear, uneven power distribution, imperfect panel fitting leading to 'light leak' at the edge of the screen, etc.), and means that the input-output function is liable to differ for every pixel. As a result, the input-output must be measured at multiple locations to ensure precise stimuli. For example, in the present experiment we calibrated the screen using a uniform grid of 10 by 8 locations, and interpolated between locations to provide coverage of every pixel. With a 10-bit display (i.e., 1024 luminance levels), this implies a total of 81,920 measurements, each of which should ideally be repeated multiple times for verification. Even using a programmable photometer, this process can take many hours, during which time a human operator needed to be present to manually reposition the photometer as required (though see Ref~[22,23]). Such an involved process of calibration is clearly incompatible with the notion of a perimeter that is cheap or widely available. In short, the need for calibration means that while anybody can acquire the technology necessary to perform screen-based perimetry, few people are in a position to use it appropriately.

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

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

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

2. METHODS

2.1. Overview

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

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

2.2. Participants

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

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

2.3. HFA Apparatus and Procedure

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

2.4. Screen-based Perimeter Apparatus and Procedure

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

This test is similar to the 'Eyecatcher' test that we reported previously[9], and used some of the same hardware and code. Eyecatcher, however, performs a quick suprathreshold evaluation of the para-central field, whereas the test reported here performed a full threshold evaluation across a modified 24-2 grid. The novel screen-based test is also similar in principle to other eye-movement perimeters[10–16]. Its key features are as follows.

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

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

As with the HFA, targets were 0.43° diameter (Goldmann III) circles of variable luminance, presented on a 24-2 grid, against a 10 cd/m^2 white background (Fig 1C). However, unlike with the HFA, participants responded by making an eye-movement towards the target location, rather than by pressing a button (see Fig 1E). Also, as shown in Figure 2C, the four test-points from the top and bottom of the standard 24-2 grid were omitted due to the dimensions of the screen. Finally, since the HFA's ('SITA-standard') thresholding algorithm is proprietary technology, the ZEST algorithm[28–31] was used to adapt stimuli and determine detection thresholds. As recommended by Turpin and colleagues[31], the ZEST prior was a bimodal probability density function, constructed by combining normative data for healthy and glaucomatous eyes. The likelihood function was a cumulative Gaussian, with a fixed slope of $\sigma = 1.25$, and a variable mean of $\mu = \langle 0, 1, 2, \ldots, 34 \rangle$. The growth pattern is given in Figure 2C. A dynamic termination criterion was used[32,33], in which the spread of the estimated posterior function was required to have a standard deviation of $\sigma \leq 1.5$ dB.

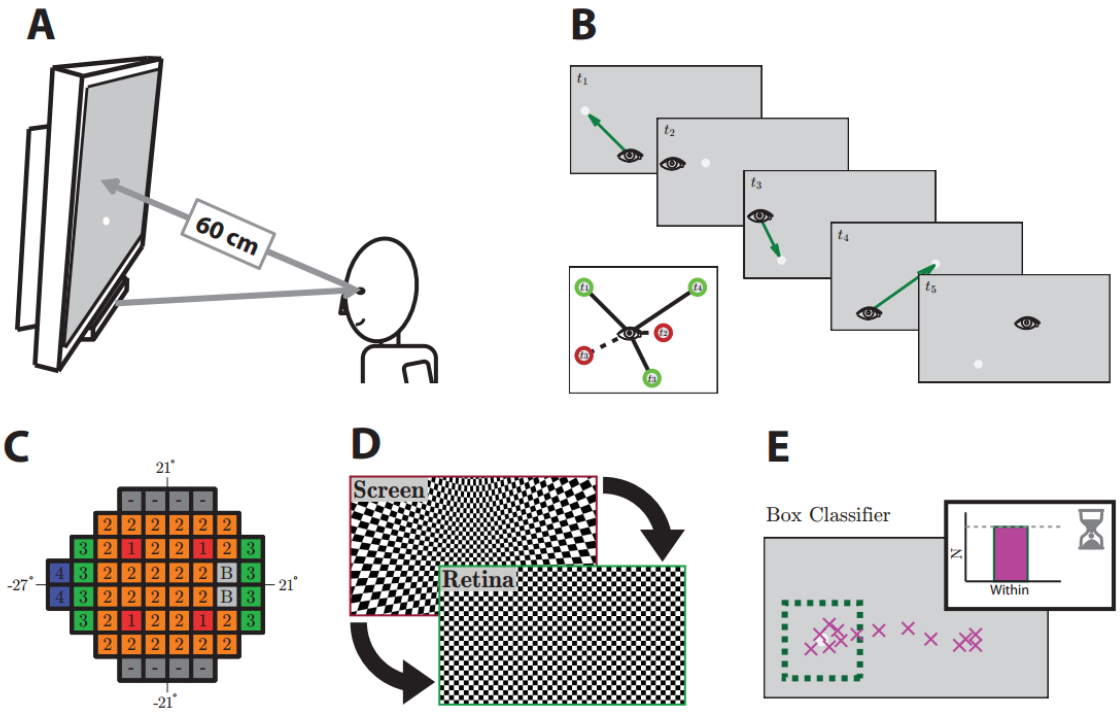

Figure 1. Screen-perimeter apparatus and procedures. **(A)** Hardware. Stimuli were presented on a 59.7 x 33.6 cm (2560 x 1440 pixel) LCD screen, viewed at a distance of 60 cm (i.e., 52.9° x 31.3° visual angle). An eye-tracker (Tobii EyeX) was mounted below the screen, and was used to measure gaze-location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

2.5. Screen Calibration

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

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

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

Estimated contrast sensitivity for each stimulus location was quantified as Differential Light Sensitivity (DLS): the smallest detectable difference in luminance, ΔL, between the target luminance, L_{targ}, and the background luminance, L_B. With both the novel perimetry measure, and the HFA reference measure, the value of L_B was fixed at 10 cd/m^2. The value of ΔL varied trial-by-trial according to an adaptive algorithm (ZEST or SITA Standard), in order to find the smallest value of ΔL that could be reliably detected on 50% of trials: ΔL_{jnd}.

Following standard perimetric convention[1,34,35], DLS values are reported in units of signal attenuation on an inverted log-scale:

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

where ΔL_{max} is the greatest displayable stimulus pedestal. For ease of comparison, this value was scaled identically for both the HFA and the novel screen-based test, using the maximum displayable pedestal of the novel test (ΔL_{max} = 225 cd/m^2). For all tests, DLS values therefore varied from 0 dB and 34 dB, with higher values indicating greater sensitivity.

3. RESULTS

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

As expected, the input-output functions from the manual and automatic luminance calibrations were virtually indistinguishable when measurements were made at the same location, i.e., at the single region of the screen that the integrated photometer samples from [*Pearson Correlation*; $r_{1022} \approx 1.0$, $P \ll 0.001$]. However, the automatic calibration assumes uniformity across the screen, and the display's spatial uniformity --- while far superior to a standard LCD monitor --- was only approximate. The standard deviation in luminance across the screen was ~3%: a magnitude similar to the Just Noticeable Difference for human contrast discrimination[36]. And some areas of the screen were ~17% more intense than others (see §3.6 for graphical illustration). Whether these imperfections are large enough to affect perimetric measurements remained an empirical question, however, which we turn to next.

3.2 Accuracy: Mean differential light sensitivity (MS)

Mean Sensitivity (MS) estimates for the three test conditions are shown in Figure 2. Group-mean MS values were 18.8 dB (Manual), 19.1 dB (Auto), and 19.1 dB (HFA). None of these values were significantly different from each other [3 paired *t*-tests; all $t_{63} \leq 1.16$; $P \geq 0.251$], or from the normative value of 19.2 dB reported previously by Brenton and Phelps[37] [3 one-sample *t*-tests; all $t_{63} \leq 1.40$; $P \geq 0.167$]. (NB: Breton and Phelps report a peak value of 30.7 dB. However, following perimetric convention this value was rescaled to 19.2 dB based on the maximum luminance output of our screen; see Eq 1). At an individual level, the results of the three tests were also positively correlated [3 Pearson correlations; $r_{62} \geq 0.50$, $p \leq 0.001$], with participants who scored higher on one test condition tending to score higher on other conditions too (Fig 2C). Taken together, these findings indicate that all three test conditions gave quantitatively similar results.

As is common in perimetry[38], a small but consistent practice effect was observed in all three test-conditions (Fig 2A), with mean sensitivity increasing between runs one and two by an average of 0.60 dB (Manual), 0.66 dB (Auto) and 0.73 dB (HFA). This difference was significant for the Auto [$t_{31} = 2.88$, $P = 0.007$] and HFA conditions [$t_{31} = 4.43$, $P < 0.001$], though did not reach significance in the Manual condition [$t_{31} = 1.55$, $P = 0.131$, *n.s.*].

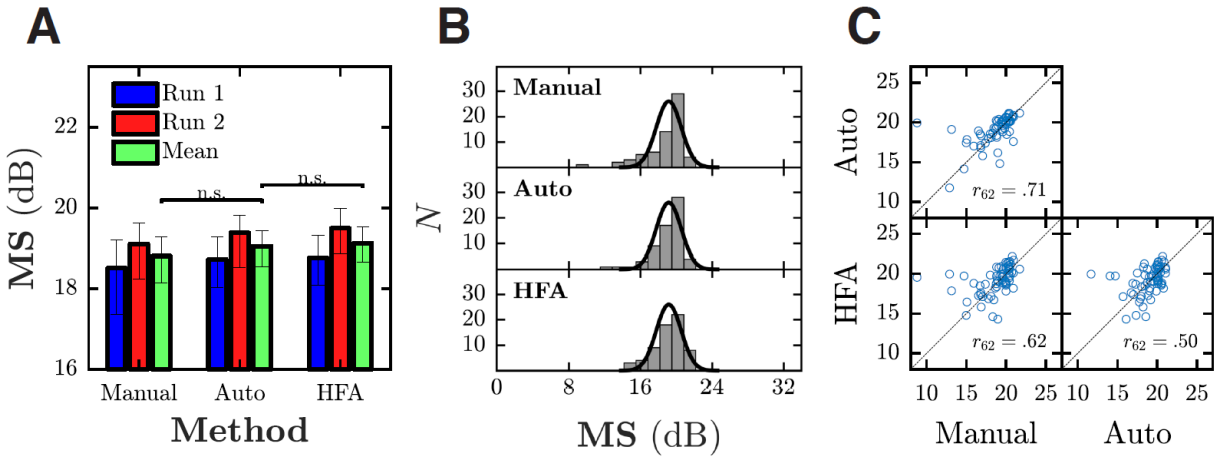

Figure 2. Estimates of mean sensitivity, MS, for each of the three test conditions. ***(A)*** Group-mean data [±95% CI], for the first and second run of each test, and the mean average of the two. ***(B)*** Histograms showing the distribution of results for all 64 (32x2) tests in each condition. Gaussian distributions show previously published normative data for the HFA[37] (21 healthy adults, aged 20-29 years), scaled to the same units as the present data using Eq 1. ***(C)*** Scatter-plot showing within-subject correlations. Markers show MS scores for

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.

3.3 Accuracy: Pointwise sensitivity (PWS)

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

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

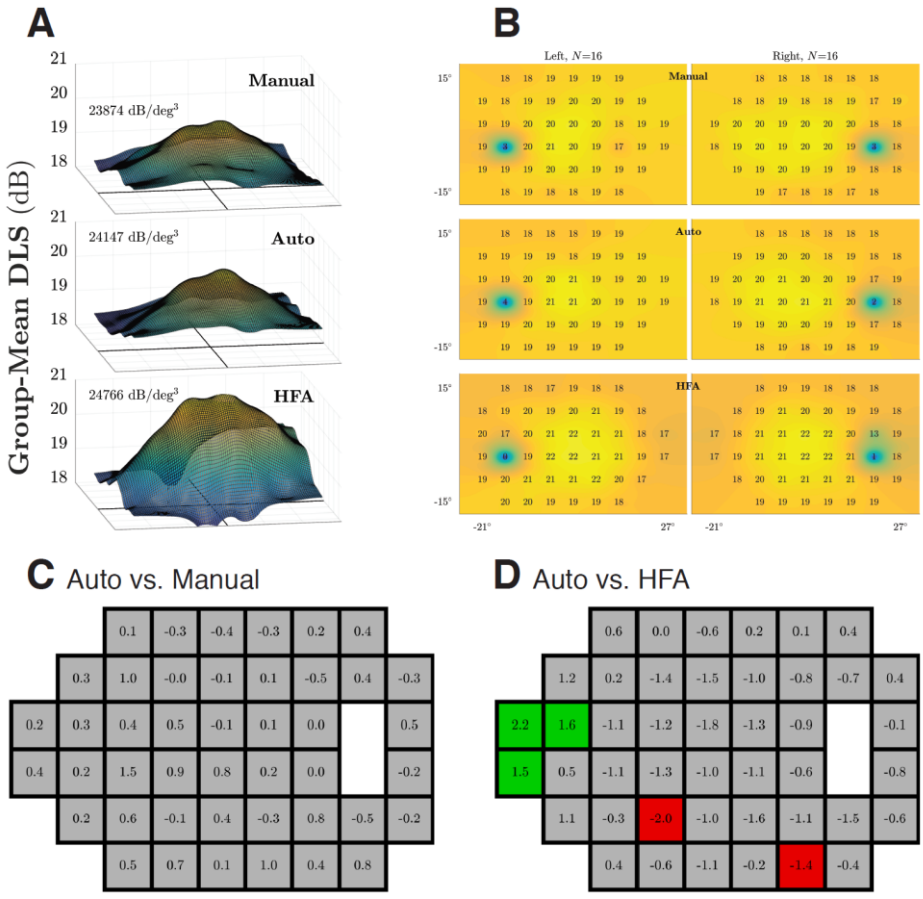

Figure 3. Distribution of pointwise visual sensitivity (DLS) estimates across the visual field. ***(A)*** Three dimensional 'Hill of Vision' plots for the new each of the three test conditions. Surfaces fitted using spring-regularized nearest-neighbor interpolation, and then smoothed using a moving-average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

3.4 Reliability: Test-retest repeatability

Bland-Altman analyses[39] were used to assess the reliability of the mean sensitivity (MS) estimates (Figure 4A). Across the three conditions, the 95% Coefficient of Repeatability [CoR_{95}] was 2.3 dB (Manual), 2.6 dB (Auto), and 1.8 dB (HFA). Using a bootstrapping procedure analogous to a *t*-test, these differences were found to be non-significant, both when comparing Auto vs. Manual ($P = 0.600$), and when comparing Auto vs. HFA ($P = 0.134$). These findings indicate that all three test conditions were similarly reliable (precise), and that using the display-screen's internal calibration did not reduce the overall reliability of the screen-based perimetry test.

To evaluate reliability at the level of individual PWS estimates, and to assess whether reliability varied across the visual field, these Bland-Altman analyses were repeated for each of the 44 individual test locations (Fig 5). When comparing Auto vs. Manual calibration, the CoR_{95} values were observed to differ significantly ($P < 0.01$) at one location only. This single difference was likely due to chance (i.e., given the $\alpha = 0.01$ significance level and the number of multiple comparisons). When comparing Auto vs. HFA, the reliability of the novel screen-based test was significantly lower for 3 central locations, and significantly higher for one peripheral location.

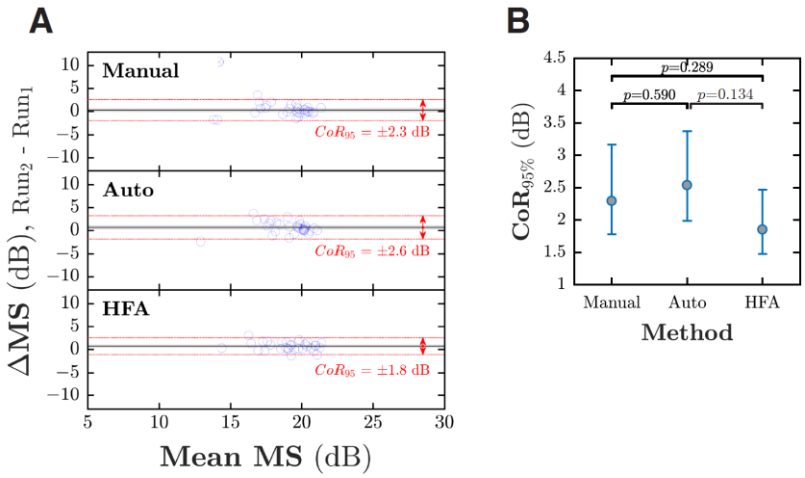

Figure 4. Test-retest repeatability for MS values. ***(A)*** Bland-Altman plots of mean sensitivity. Each marker represents a single participant. Dashed red lines indicate the 95% limits of agreement (μ ± CoR_{95}). In the Manual condition, one point [black cross] was excluded as an outlier). ***(B)***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

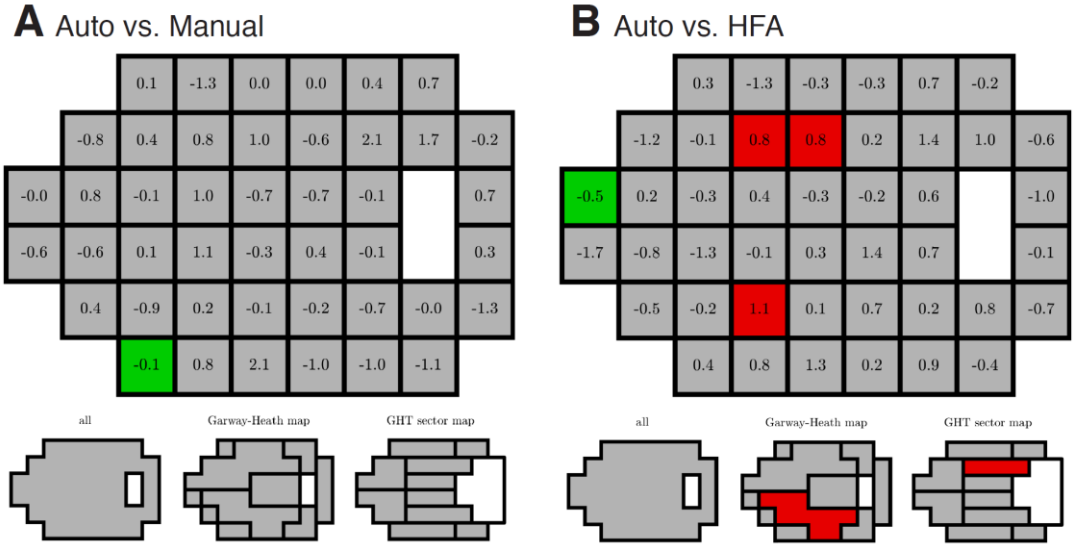

Figure 5. Differences in test-retest repeatability for individual PWS values. ***(A)*** Comparison of Auto vs. Manual. The top panel shows differences in CoR_{95} for each location on the test grid ($CoR_{Auto} - CoR_{Manual}$). The bottom panel shows the same data grouped into various sub-regions, based on three published visual-field maps[40]. ***(B)*** Same as (A), comparing Auto vs. HFA measures ($CoR_{Auto} - CoR_{HFA}$).

3.5 Test duration

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

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

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

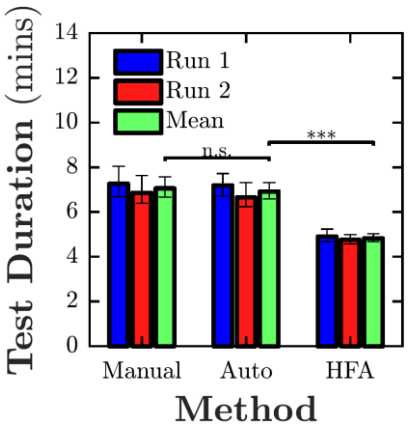

Figure 6. Group-Mean [±95% CI], test durations, for each of the three test conditions. Same format as the MS values presented in Figure 2A.

3.6 Photometric screen characterization: 20 months later

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

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

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

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

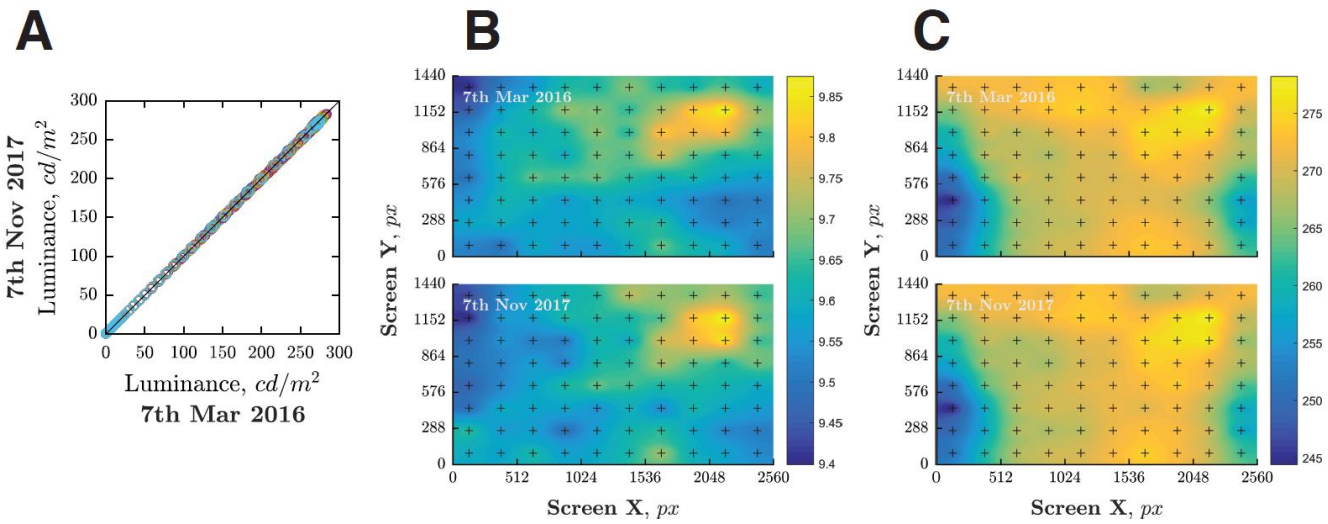

Figure 7. Photometric luminance measurement for the display screen, before and after 20 months of regular use. **(*A*)** Comparison of individual measurement at each of 80 screen locations. Measurements were made with a ColorCal MK II colorimeter. The black diagonal line denotes unity (perfect correlation).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

4. DISCUSSION

The goal of the present study was to assess whether screen-based perimetry could be performed without the need for any manual calibration, using only the integrated photometer contained within 'professional' commercial monitors (i.e., and assuming uniformity across the screen). The photometric data showed that this internal calibration was imperfect, and resulted in measurable deviations in luminance across the screen. However, these deviations had no measurable effect on the behavioral results. Specifically: there were no detectable differences in mean sensitivity (MS) estimates, pointwise sensitivity (PWS) estimates, test-retest repeatability, or test duration (i.e., when comparing screen-based measurements made with and without full manual calibration). This suggests that self-calibrating screens are sufficiently accurate to support accurate threshold perimetry measurements.

More generally, the present data provide further evidence that screen-based perimeters – and eye-movement perimetry in particular – could provide viable alternatives to traditional standard automated perimeters [SAPs]. The accuracy and precision of the novel screen-based test were indistinguishable from those made using the HFA (the current clinical reference standard), although the HFA was ~30% faster. This may be of particular importance in situations where access to SAP assessment is limited, either by physical/cognitive impairment, or on the grounds of portability or cost (*see below*).

4.1 Study Limitations

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

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

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

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

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

One concern is that while automatic-calibration removes one key bottleneck, scientific expertise is still required to program the stimuli and create the tests. To militate against this, all of the code necessary to calibrate and run the tests described in the present paper have been made freely available online (see *Methods*). This is 'research grade' code, however, and in the longer term, it will be necessary to convert this into a more user-friendly 'click to run' app. It will also be necessary for any results to be reviewed and interpreted by a qualified clinician, for example via the sorts of cloud-based solutions already under evaluation elsewhere[44].

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

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

These outstanding limitations notwithstanding, we believe the current work marks a qualitative step forward. The equipment required is cheap, easy to use, and easy to replace. And unlike with traditional, dedicated perimeters, the equipment is inherently multipurpose. We have already shown, for example, that the same basic hardware can also be used to perform acuity assessments[47], and we routinely use the equipment described in the present study for other day-to-day tasks, such as playing videos to patients in-between tests (i.e., particularly when performing pediatric assessments[48]), or for performing general office work. For these reasons, we believe that a simple 'plug-and-play' perimeter

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

Acknowledgments

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

Disclosures:

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

REFERENCES

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

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

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