tvst

https://doi.org/10.1167/tvst.7.3.17

Article

A Method Using Goldmann Stimulus Sizes I to V–Measured Sensitivities to Predict Lead Time Gained to Visual Field Defect Detection in Early Glaucoma

Jack Phu[1,2], Sieu K. Khuu[2], Bang V. Bui[3], and Michael Kalloniatis[1,2]

[1] Centre for Eye Health, University of New South Wales Sydney, Sydney, NSW, Australia
[2] School of Optometry and Vision Science, University of New South Wales Sydney, Sydney, NSW, Australia
[3] Department of Optometry and Vision Sciences, University of Melbourne, Parkville, VIC, Australia

Correspondence: Michael Kalloniatis, Centre for Eye Health, University of New South Wales Sydney, Sydney 2052, NSW, Australia. e-mail: m.kalloniatis@unsw.edu.au

Received: 23 November 2017
Accepted: 24 April 2018
Published: 7 June 2018

Keywords: Humphrey Field Analyzer; contrast sensitivity; perimetry; Ricco's area; spatial summation

Citation: Phu J, Khuu SK, Bui BV, Kalloniatis M. A method using Goldmann stimulus sizes I to V–measured sensitivities to predict lead time gained to visual field defect detection in early glaucoma. Trans Vis Sci Tech. 2018;7(3):17, https://doi.org/10.1167/tvst.7.3.17

Purpose: To predict the lead time (difference in time taken for a visual field [VF] defect to be detected) obtained when using stimulus sizes within or near the size of the critical area of spatial summation (Ac), and to test these predictions using sensitivity measurements from a cohort of glaucoma patients.

Methods: Thirty-seven patients with early open-angle glaucoma and 60 healthy observers underwent VF testing on the Humphrey Field Analyzer in full threshold mode using Goldmann stimulus sizes I to V (GI–V) across the 30-2 test grid. We used the sensitivities measured using GI to V in healthy patients to predict the lead time gained by using stimulus sizes within the size of Ac at all locations within the 30-2 grid. Then, we used sensitivities measured in the glaucoma patients to test this predictive model.

Results: Median lead time to VF defect detection when using stimulus sizes within Ac compared with stimulus sizes larger than Ac was 4.1 years across the 30-2 test grid (interquartile range, 3.1 and 5.1 years). Sensitivities of the glaucoma patients showed good agreement with the predictive model of lead time gained (77.5%–84.3% were within $\pm$3 dB).

Conclusions: Our model predicted substantial lead time differences when using stimulus sizes within or near Ac. Such stimulus sizes could potentially detect VF defects, on average, 4 years earlier than current paradigms.

Translational Relevance: Stimulus sizes within or near Ac may be more suitable for early detection of glaucomatous VF defects. Larger stimulus sizes may be more suitable for later monitoring of established disease.

Introduction

The clinical standard for assessment of the visual field (VF) in glaucoma is standard automated perimetry (SAP), such as using the Humphrey Field Analyzer (HFA; reviewed in Jampel et al.[1] and Phu et al.[2]). Recently, imaging technologies, such as optical coherence tomography (OCT), have highlighted the discordance between structural and functional defects in glaucoma.[3,4] One suggestion for structure–function discordance, particularly in early glaucoma, is that current SAP protocols are relatively insensitive to early VF loss.[5] In support of this, a number of studies have shown that current SAP parameters—a fixed size achromatic stimulus (Goldmann size III, GIII) presented for a fixed duration (~100–200 ms)—are not optimal for revealing maximum threshold elevation in glaucoma, due to spatial and temporal summation characteristics, which are not explicitly considered in current SAP devices.[6–9]

Ricco's law of spatial summation relates luminance threshold (L) and stimulus area (A) mathematically: $k = L \times A^n$, where k is a constant, and n is the summation exponent (when $n = 1$, there is complete spatial summation; when $0 < n < 1$, partial summation holds).[10] When luminance and size are

plotted in log-log form, the slope of −1 denotes the region of complete spatial summation, the limit of which is known as Ricco's critical area (Ac). Within this region, there is a 1:1 relationship between luminance and stimulus area, while outside of Ac it changes in a nonlinear fashion up until the point of no summation. Ac is affected by a variety of factors, such as eccentricity[11–15] and disease.[6–9,16,17]

Recent studies have described the relationship between sensitivities measured using commercially available stimulus sizes, for example, the five Goldmann sizes (GI–V) available on instruments such as the HFA using a two-line fit: the first line representing the region of complete spatial summation, the second line indicating the tangential slope of partial summation, with the inflection point being an estimate of Ac.[6–9,11–14] Although there are limitations to this estimate of spatial summation characteristics,[15] it provides a basis for describing the relationship between sensitivities measured using such parameters.

Using the relationship between sensitivities measured using various stimulus sizes, it is possible to determine the stimulus size that balances the ability to reveal the largest sensitivity reduction in disease and an acceptable level of variability. Studies have shown that stimulus sizes near to or within Ac of healthy subjects show a greater sensitivity reduction in patients with ocular disease, such as glaucoma, despite greater variability compared with larger stimuli.[6–9] Further, a recent study has highlighted the importance of considering the signal-to-noise ratio (SNR) in terms of comparing clinical utility of a given stimulus in identifying abnormal visual function.[18] Specifically, Rountree et al.[18] demonstrated a greater SNR when using area-modulated stimuli, contrast modulated stimuli scaled to Ricco's area, and a combined modulation of both contrast and area. In applying these principles to the Goldmann stimulus sizes currently available in clinical perimeters, we test the hypothesis that stimulus sizes that are tailored to the location-specific Ac may also reduce the measurement variability in comparison to using the smallest available stimulus sizes on the instrument. Minimizing variability and maximizing dynamic range are particularly relevant at more peripheral test locations,[12,13,19] and in stages of worsening VF defects.[20–24]

In the present study, we used the relationship between the sensitivity measurements obtained using stimulus sizes GI to V to predict the lead time to VF defect detection in early glaucoma at specific locations within the 30-2 test grid. Here, lead time refers to the difference between a stimulus size tailored to the location-specific Ac across the VF (we refer to this as a Spatially Equated Stimulus [SES]) and usual care, which is the Goldmann size III (GIII) target. Given the recent interest in using Goldmann size V (GV),[25–27] due to its low variability and comparable performance to GIII, we also tested the lead time between SES and GV. Then, we tested this model using the sensitivities measured using GI to V in a group of glaucoma patients. To account for the variability of sensitivities measured at locations with advanced defects, we also examined the results when the SAP measurement floor was reached.

Methods

Observers and Patients

Sixty healthy observers (mean age: 42.5 years, standard deviation [SD]: 16.3 years; 29 males, 31 females) and 37 patients with open-angle glaucoma (mean age: 63.3 years, SD: 9.1 years; 25 males, 12 females) underwent VF testing on the HFA. Although the age distribution was slightly different between the two cohorts, we used previously published methods[6,8,11,12,28,29] and age-correction factors[30] (which is, in turn, similar to methods used in some perimetric algorithms[31]) to facilitate comparisons after sensitivity data were converted into a 50-year-old equivalent patient (see below).

Results from the healthy subjects and 19 of the glaucoma patients have been reported, in part, in previous studies in which the inclusion and exclusion criteria can be found.[6,8,12,29,32] The glaucoma patients had open-angle glaucoma; most patients ($n = 33$) had early glaucoma, and four patients had moderate glaucoma (Table 1).[33] Ethics approval was given by the relevant University of New South Wales Ethics committee. The observers gave written informed consent prior to data collection, and the research was conducted in accordance with the tenets of the Declaration of Helsinki.

Apparatus and Procedures

The HFA (HFA-ii; Carl Zeiss Meditec, Dublin, CA) was used to measure sensitivities at the 75 locations (including the fovea, and excluding the 2 points near to the physiologic blind spot) of the 30-2 test pattern in full threshold mode. Methods for conducting VFs and reporting results in decibels have been previously described,[6,8,12,29] but in short, the sensitivity values were taken directly from the HFA

Table 1. Characteristics of Study Participants

	Healthy ($n = 60$)	Glaucoma ($n = 37$)
Age, y, mean ± SD	42.5 ± 16.3	63.3 ± 9.1
Sex, *n*, male:female	29:31	25:12
Eye tested, *n*, right eye:left eye	37:23	15:22
Spherical equivalent refractive error, diopters (median, range)	−1.07 (+2.63 to −6.00)	−0.14 (+3.38 to −5.38)
Mean deviation, dB, mean ± SD	−0.74 ± 1.20	−3.08 ± 2.11
Pattern standard deviation, dB, mean ± SD	1.97 ± 0.53	4.37 ± 3.03

printout. Results were converted to right eye orientation for all subjects.

Age-Correction of Sensitivities and the Normative Distribution

Age-correction of sensitivities using location-specific age-correction factors[30] was performed to facilitate pooling of sensitivities across all healthy observers and glaucoma patients, as per previous methods.[6,8,12,29] As spatial summation characteristics do not change with age,[12,34–36] and because age-correction factors do not differ across size in a clinically significant fashion (mean difference in pairwise comparison across size-specific correction factors for GI to V: 0.08 ± 0.08 dB per decade),[29] the age-correction factors for GIII were applied to sensitivities across all stimulus sizes.[28] Overall, the effect of age upon sensitivity values was small, and was within the typical range of test–retest variability of VFs.[12,29,37]

We also performed a subanalysis using a subset of healthy subjects ($n = 21$, mean age: 60.5 years, SD: 7.4 years; 13 males, 8 females) to serve as an age-matched cohort for comparison (unpaired *t*-test, $P = 0.2371$), in case the analysis with the complete cohort was confounded by the presence of younger subjects. The trends evident when using this subgroup were the same as those in the age-corrected and pooled complete cohort of $n = 60$ (Supplementary Table S1 and Supplementary Figs. S1, S2). Thus, we continue to report the results when using the complete age-corrected cohort of 60 healthy subjects.

Spatially Equated Stimuli: A New Reference Stimulus Size for Lead Time Calculations

We wished to test a model that predicts the lead time between stimuli tailored to the Ac of healthy subjects and usual care (GIII) and a larger stimulus size (GV). Due to the limited number of commercially available stimulus sizes on clinical instruments, such as the HFA, recent studies have proposed a two-line fit to estimate the size of Ac,[6,8,11,12] a method which is also seen in laboratory-based studies[7,9,13,36] using a limited number of stimulus sizes (Fig. 1A). This technique may be contrasted with previous curvilinear descriptions of spatial summation functions, but in these instances, it may be difficult to estimate the limit of Ac.[15,17] Another technique that has been described is similar, but involves scaling energy output[7] or equating for stimulus size[6,12] with an initial slope of 0, representing the constant k in Bloch's and Ricco's law, respectively, and allowing the point of inflection and second slope to vary (to be able to identify subtle changes in Ac). This is a topic of ongoing debate (see Supplementary Figs. S3, S4). However, the amount of sensitivity reduction is known to change depending on whether the stimulus used to measure sensitivity is within or outside of complete spatial summation (Fig. 1B). Note that the functions shown in Figure 1 indicate a slightly upward but rather predominantly rightward shift in the glaucoma patient with relative to the healthy subject. Similar to the work of Redmond et al.,[9] this model suggests that sensitivity loss occurring in early glaucoma could be explained by changes in Ac alone, rather than by uniform alterations in sensitivity measured using different stimulus sizes.

The sensitivities measured in the healthy cohort formed the normative data and distribution limits for measuring VF defects in the glaucoma test group. The mean and 95% lower limit of the normative distribution for GI to V across the 30-2 test grid are shown in Figures 2A to 2E. In the present study, we tested the ability of a recently proposed map[6,8] guiding stimulus size selection for SES—the largest stimulus size within or near the size of Ac at each specific test location within the 30-2 test grid (Fig. 2F). Using SES has been shown to flag a greater number and depth of VF defects compared with GIII and GV.[6,8] Thus, we hypothesized that SES would predict lead time to defect detection compared with GIII and GV in glaucoma.

However, as GIII is the current clinical standard for VF testing, we compared the number of VF defects

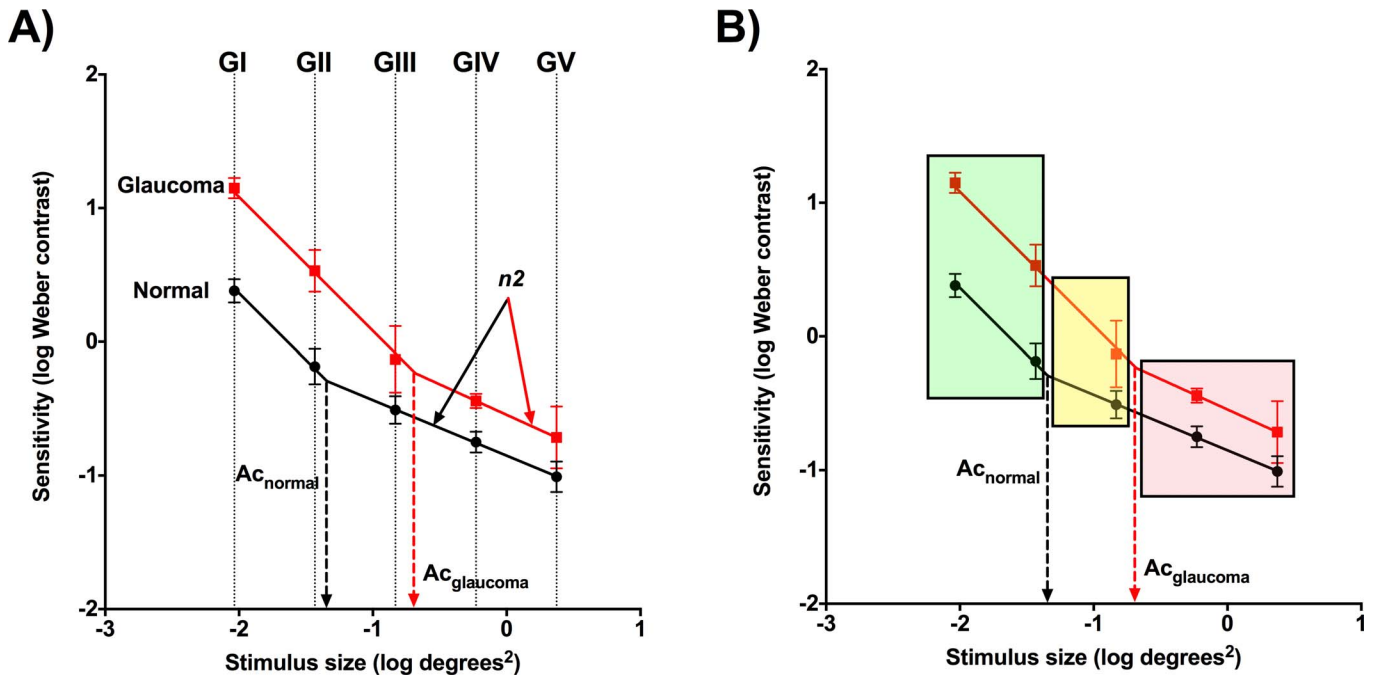

Figure 1. (A) An example of a two-line fit spatial summation function relating the mean sensitivities (log Weber contrast) measured using the five Goldmann stimulus sizes (each datum point: GI–V, log degrees^2). Two representative subjects are presented (healthy: *black*; glaucoma: *red*) at a single representative location, adapted from previous studies.[6–9] The point of inflection indicates the estimate of the critical area of spatial summation (Ac), and the second slope represents the tangential slope of partial summation (*n2*). (B) The same spatial summation function in (A) with *shaded regions* indicating different relative sensitivity reductions occurring in glaucoma when measured using different stimulus sizes, depending on whether they are within Ac. *Green shaded region*: stimuli within the Ac of both healthy and glaucoma. *Yellow*: less sensitivity reduction detected, indicates stimulus sizes larger than the normal Ac but within the Ac of glaucoma. *Red*: the least amount of measured sensitivity reduction, indicates a stimulus size larger than the Ac of both healthy and glaucoma patients. Note that in both (A) and (B), there is a predominantly rightward (and only slightly upward) shift in the function of the patient with early glaucoma, relative to the results of the healthy subject. The largely lateral shift of the function indicates the fact that sensitivity loss occurring in early glaucoma could be explained by changes in Ac alone (see Redmond et al.[9]).

detected using GI to V and SES to determine whether SES offered a balance between defect detection and stimulus dynamic range. We also determined the ratio of the depth of defect at each specific test point to determine a new "ground truth" for VF testing. Finally, in order to test the dynamic range of each stimulus size, we also determined the number of times each stimulus size (GI–V and SES) reached the instrument measurement floor.[23] Here, we hypothesized that SES, as shown in previous studies,[6,8] detects more defects compared with the current clinical standard while maintaining a suitable dynamic range for testing. These analyses are all provided in detail in the Supplementary Material (Supplementary Figs. S1, S2, and Supplementary Table S1).

Predicting Lead Time to VF Defect Detection in Glaucoma

In the present study, we use lead time to refer to the difference in time between when SES and GIII or GV detected an 'event': a sensitivity value that was below the location-specific 95% lower limit of the normative distribution. We used a sensitivity decline of −0.8 dB/y[38] to be the average rate of progression for SES, a stimulus size that was at or near Ac and hence more likely highlights glaucomatous damage prior to stimulus sizes larger than Ac. The change in sensitivity required to reach an 'event' was taken as the difference between the mean sensitivity and the 95% lower limit of the normative distribution at each specific location. Therefore, the time taken for SES (T_{SES}, in years) to detect an 'event' was calculated by dividing the sensitivity change by the rate of change (Equation 1):

$$T_{SES} = [(\text{mean sensitivity} - 95\%\ \text{lower limit})/-0.8] \quad (1)$$

The two-line nonlinear regression fit describing the relationship between sensitivities measured using GI to V suggests that the rate of change of stimuli larger

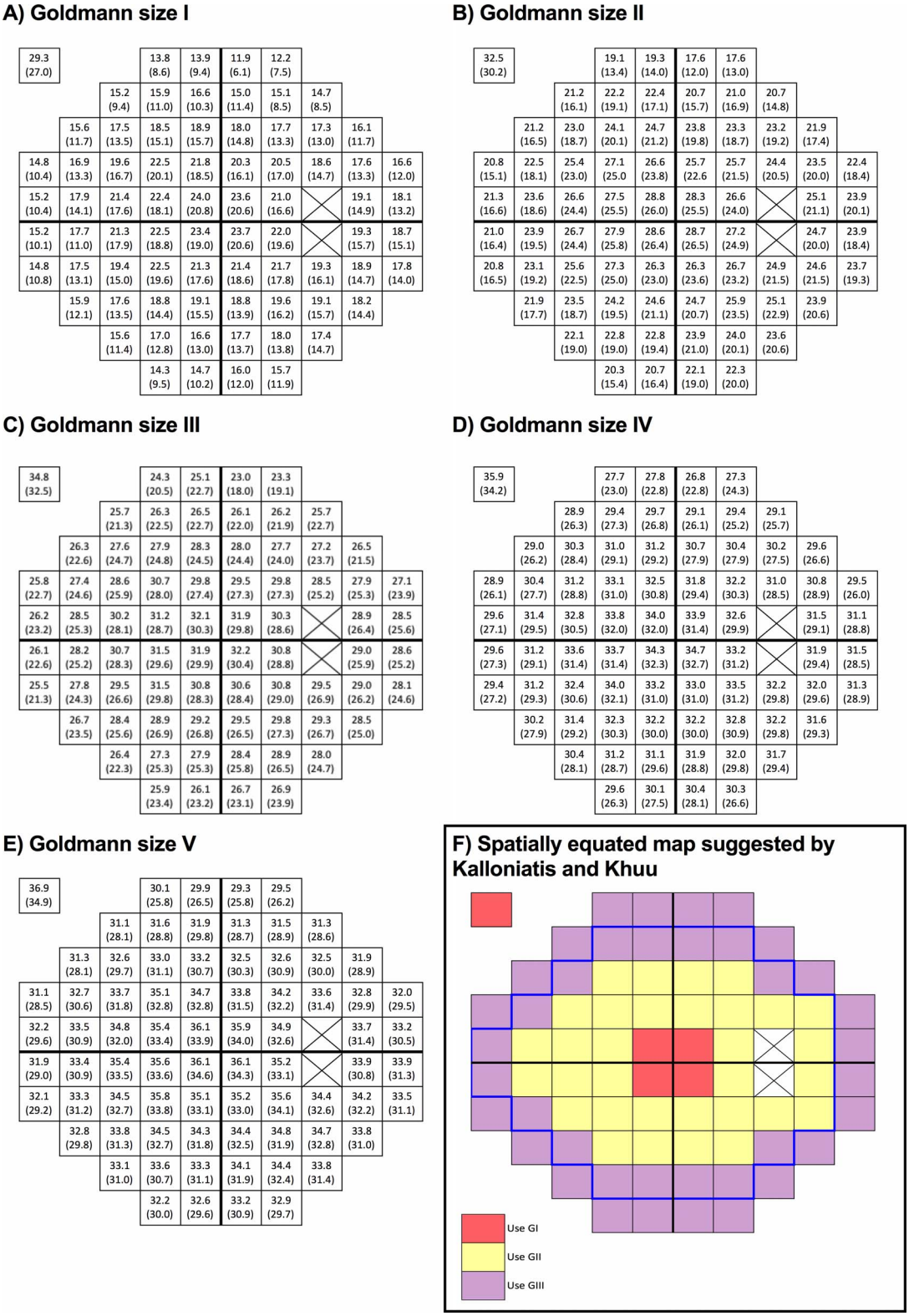

Figure 2. Mean and 95% distribution limits (i.e., the lower 5th percentile) at each test location within the 30-2 test grid (fovea offset to the upper left for clarity, and the two points next to the physiologic blind spot have been crossed out) for Goldmann sizes I to V (A–E). (F) The spatially equated stimulus map as suggested by Kalloniatis and Khuu,[6] in which GI to III are used at various locations in the 30-2 test grid in order to maximize the stimulus size while still being within or near complete spatial summation. The *blue border lines* indicate the extent of the 24-2 test grid.

than Ac is scaled by *n2*, the tangential slope of partial summation (Fig. 1A; see also Results and Discussion for further information). This is a slower rate of change as the stimulus size is larger than Ac and hence is within the region of partial summation, which reflects the change in the spatial summation function with the disease process. There is only a very slight shift upward, representing reductions in sensitivity, but rather the function undergoes a predominantly rightward shift reflecting increases in Ac (the sensitivity loss occurring in early glaucoma has been found to be explained by a change in Ricco's area alone[9]). Thus, when stimulus sizes at or smaller than Ac undergo a certain rate of change, stimulus sizes larger than Ac, due to the predominantly rightward shift of the function, undergo a slower rate of change, that is, one that is scaled by *n2*. Thus, the rate of change here was −0.8 multiplied by the location-specific tangential slope of partial summation *n2*. We determined *n2* using a segmental nonlinear regression (GraphPad Prism version 7; GraphPad Inc., La Jolla, CA) on the age-corrected normative data ($n = 60$). The first slope was fixed at −1 (complete spatial summation), and the second slope was taken as the tangential slope of partial summation *n2* (Supplementary Fig. S4F). The time taken for stimulus sizes larger than Ac at the relevant locations was calculated using Equation 2 (T_{GIII} or T_{GV}, in years, as appropriate):

$$T_{GIII} = [(\text{mean sensitivity} - 95\%\ \text{lower limit})/(-0.8 \times n2)] \quad (2)$$

The lead time was therefore the difference between T_{GIII} (or T_{GV}) and T_{SES} (in years). For example, a lead time of 4 years between SES and GIII means that SES detects an 'event' 4 years before GIII.

Testing the Predicting Model Using Sensitivities Measured in Patients With Glaucoma

We then used the sensitivities measured in patients with glaucoma to test the model predicting lead time. When the model predicts a lead time for 'event' detection, it also allows for predicting the sensitivity of GIII or GV when SES first detects an 'event'. The predicted sensitivity (S_P) is calculated by subtracting from the mean GIII (or GV) sensitivity the product of the rate of change, *n2* and the time elapsed since SES detected an 'event' (T_{SES}) (Equation 3):

$$S_p = \text{mean sensitivity} - (-0.8 \times n2 \times T_{SES}) \quad (3)$$

We then compared S_P with the actual sensitivity (S_A) measured using the HFA at the same location for the same patient using GIII and GV. The resultant difference, $S_P - S_A$, was determined at all locations with an 'event', to determine the level of agreement between predicted and actual sensitivities. A difference of 0 means that the predicted lead time, t_L, is accurate. Due to the natural variability of VFs, we used a cutoff of ±3 dB as the limits of agreement between S_P and S_A.[39,40]

Measurement Floor Effect

Sensitivities at or near the measurement floor are unlikely to provide further useful information about the depth of defect.[22,23] Here, we defined the measurement floor as the difference between 19 dB and the 95% lower limit of the normative distribution at each location in the test grid, as suggested by previous studies to be the lower limit of reliable measurement.[22,23] We converted the 19 dB of GIII into the equivalent decibel value (*F*) for GI, GII, GIV, and GV using the spatial summation functions of healthy observers (as per Fig. 1) and Equations 4 to 7, respectively. We determined the proportion of points that reached a 'floor' effect when using SES, and examined how these points affected the predicted sensitivities, S_P.

$$\text{For GI: } F = 19 - ([([\text{Ac} - (-0.831)] \times n2) + ([(-1.433) - \text{Ac}] \times -1) \times 10] + 6) \quad (4)$$

$$\text{For GII: } F = 19 - [([\text{Ac} - (-0.831)] \times n2) + ([(-1.433) - \text{Ac}] \times -1) \times 10] \quad (5)$$

$$\text{For GIV: } F = 19 - (n2 \times 6) \quad (6)$$

$$\text{For GV: } F = 19 - (n2 \times 12) \quad (7)$$

Statistical Analysis

Statistical analysis was conducted using GraphPad Prism Version 7. Outliers for the healthy observers were identified and excluded using the ROUT Method set at Q = 10%.[41] A D'Agostino and Pearson omnibus normality test ($\alpha = 0.05$) was performed on the healthy cohort for each location, and this did not yield significant results, showing that sensitivity values were normally distributed at all locations within the 30-2 test grid. The normally distributed sensitivity data contrasted with the work of Heijl et al.[30] This has been discussed in recent papers,[29,32] but, in short, ensured

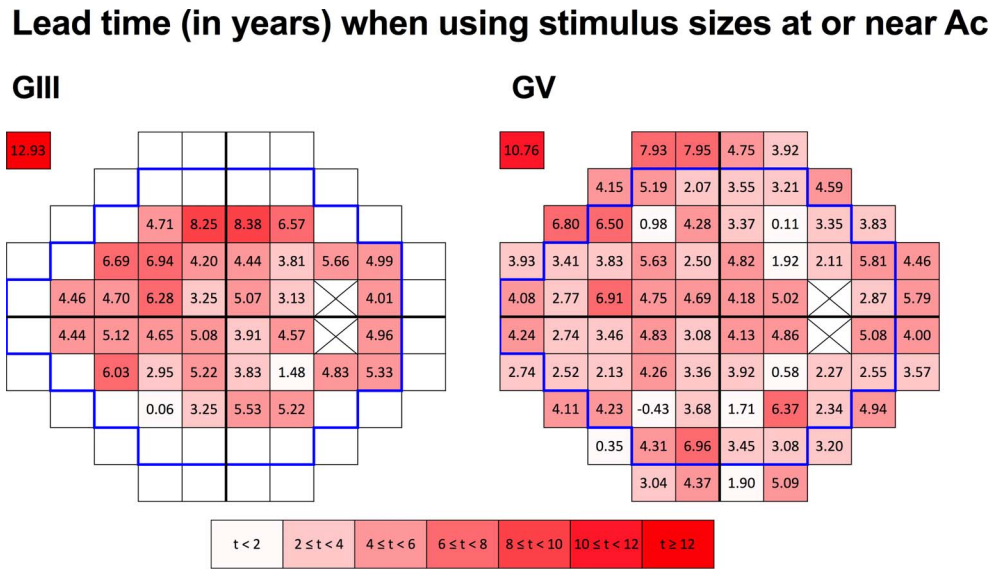

Figure 3. Lead time (in years) when using stimulus sizes at or near the size of Ac compared with GIII and GV across all test locations within the 30-2 test grid (the fovea has been offset to the upper left, and the two locations adjacent to the physiologic blind spot have been crossed out). A positive value indicates the number of years earlier that stimulus sizes at or near Ac detect an 'event' compared with GIII or GV (color coded by binned values of 2 years, with *darker values* indicating more years of lead time). Note that the *white empty cells* in the GIII map are test locations at which GIII is near Ac, and hence used within the spatially equated stimulus paradigm. The *blue border lines* indicate the extent of the 24-2 test grid.

that the obtained sensitivity values in the present study reliably reflected visual function by minimizing errors due to subject variability. Data were analyzed using descriptive statistics, paired *t*-tests, one-way analysis of the variance (ANOVA), and two-way repeated measures ANOVA. For magnitude of defects, we reported the difference from the arithmetic mean of healthy subjects, which was not significantly different to the geometric mean. For nonparametric data, we report median and interquartile ranges (IQR). Post hoc analyses (Tukey's multiple comparisons with Dunn's corrections at $\alpha = 0.05$) were performed when significant effects were found on ANOVAs. Pearson's *r* was used to examine correlations between predicted and actual sensitivities.

Results

Predicting Lead Time to Defect Detection When Using Spatially Equated Stimuli Compared With GIII and GV

The lead time (difference between T_{SES} and T_{GIII}/T_{GV}) at each location within the 30-2 test grid is shown in Figure 3. Across the 30-2 test grid, the median lead time using SES was 4.8 (IQR, 4.0 and 5.6) years and 3.9 (IQR, 2.8 and 4.8) years for GIII and GV, respectively. There was no systematic eccentricity-dependent effect. There were one and five locations at which the lead time was less than 1 year for GIII and GV respectively, and this was likely due to the high variance of GII measurements (as recommended by Kalloniatis and Khuu's[6] spatial map) at those specific locations. Notably, these locations were isolated, and did not form a cluster of contiguous points, which is typically used to delineate a glaucomatous defect.[33] Thus, overall, there were significant lead times when using SES relative to GIII and GV.

Testing the Predictive Model: Predictive Sensitivities Compared With Actual Measured Sensitivities

We then used the sensitivities measured using GI to V of the glaucoma patients ($n = 37$) to determine the accuracy of the predicted lead time. If the application of the scaled progression rate ($-0.8 \times n2$ dB/y) to predict sensitivities (S_P) shows reasonable agreement with the actual measured sensitivities (S_A) in glaucoma patients, then it suggests that the predictive model provides a valid estimation of lead time when using stimulus sizes at or within Ac. Thus, S_P and S_A were compared at locations that had an 'event' detected by SES.

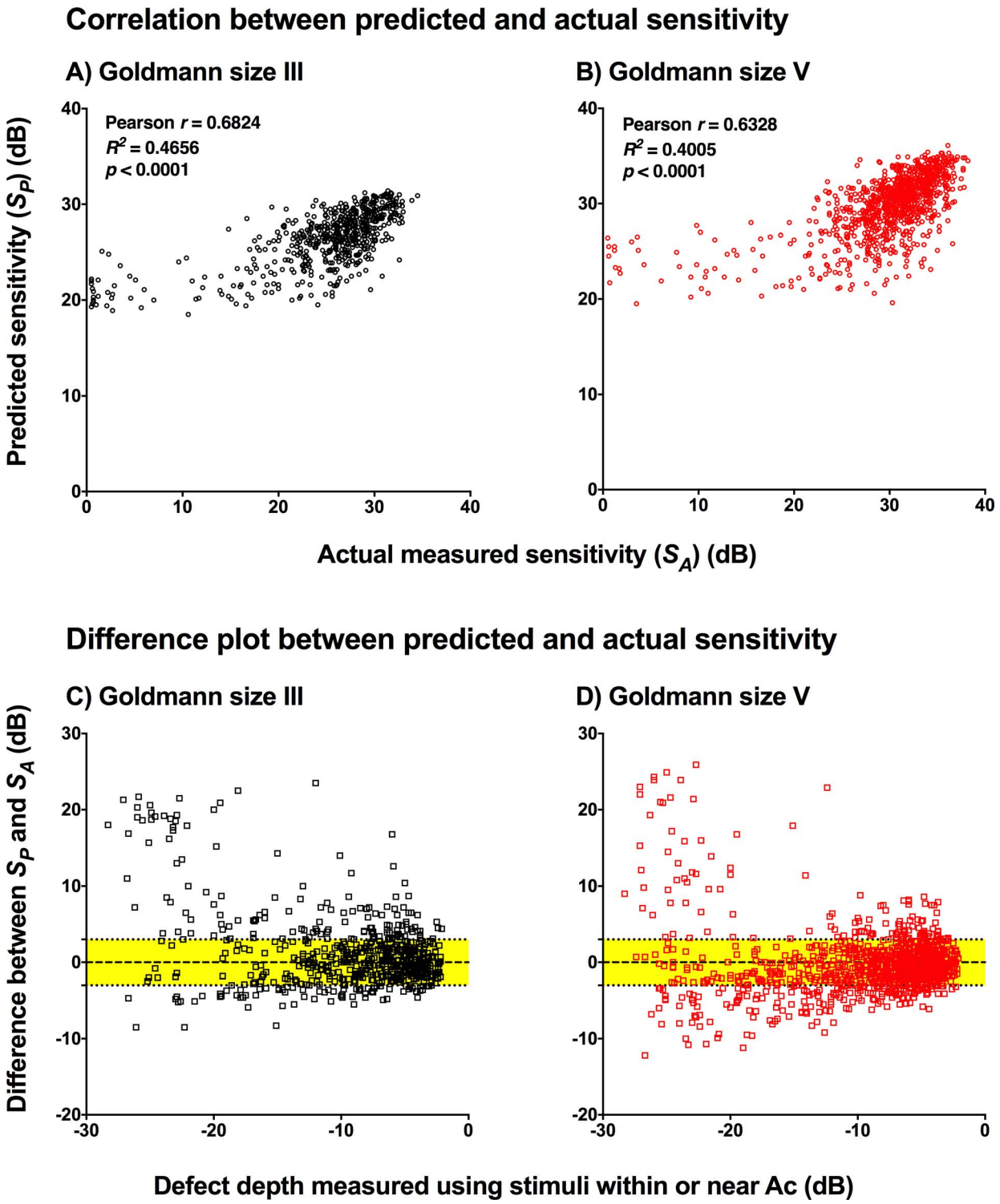

Figure 4. A comparison of sensitivity predicted by the model (S_P) and actual measured sensitivities (S_A) for Goldmann size III (GIII, *black*) and Goldmann size V (GV, *red*) at locations where stimulus size within or near the size of Ac (SES) detected an 'event'. *Top row*: the correlation between S_P and S_A for GIII (A) and GV (B). Pearson's *r*, R^2, and *P* values shown on each figure. *Bottom row*: difference plot between S_P and S_A (in dB) as a function of visual field defect depth (in dB) when measured SES for GIII (C) and GV (D). The *dashed black line* indicates no difference between S_P and S_A ($y = 0$), and the *yellow area* indicates the region of ± 3 dB, which is the approximate test–retest variability of the instrument.

Table 2. The Number of 'Events' Flagged by Each Stimulus Size. The Number of Points (*n*) Reaching the HFA Measurement Floor Effect (Equivalent to 19 dB Measured Using Goldmann Size III, GIII) for Each Stimulus Size (Goldmann Size I-V, GI-V)

	SES	GI	GII	GIII	GIV	GV
Number of 'events' detected (*n*)	1098	1209	1205	912	770	788
Number of times reaching the 19 dB (GIII equivalent) measurement floor (*n*)	415 (37.8%)	628 (51.9%)	570 (31.3%)	285 (20.1%)	155 (11.2%)	88 (37.8%)

The percentages indicate the proportion of 'events' within the total number of 'events' identified by each respective stimulus size.

S_P and S_A were correlated when measured using GIII ($r = 0.6824$, $R^2 = 0.4656$, $P < 0.0001$) and GV ($r = 0.6328$, $R^2 = 0.4005$, $P < 0.0001$; Fig. 4A). However, there were a number of points toward the left side of the figures where S_P was higher than S_A. These indicated that the model predicted a higher sensitivity at locations which had a lower actual sensitivity. When the difference between S_A and S_P was plotted as a function of the depth of defect revealed using a stimulus size within or near Ac (SES), the majority of values (GIII: 65.1%; GV: 71.4%) were within the ±3-dB range of test–retest variability of the instrument (Fig. 4B). This suggests that there is reasonable agreement between S_P and S_A. However, there was a greater difference between predicted and actual values when the depth was greater (i.e., more negative along the x axis) that could be due to factors including the reduced reliability of sensitivity measurements at greater defect depths (Fig. 4B).

The Effect of the Measurement Floor on the Agreement Between Predicted and Actual Sensitivities

As the largest differences between S_P and S_A occurred at locations with greater VF loss measured using SES, we then performed the same analysis after excluding points that reached the measurement floor of SES. As mentioned in the Methods, the measurement floor was a sensitivity of 19 dB for GIII, and was scaled for GI and GII, as SES represents different stimulus sizes used at different test locations within the 30-2 test grid (as per Fig. 3). These points were identified and excluded for subsequent analysis. Table 2 shows, first, the number of 'events' flagged by each stimulus size and SES. Table 2 also shows the number of points reaching the measurement floor of 19 dB (GIII equivalent and scaled for each stimulus size as per Equations 4–7), and its relative proportion within the total number of 'events' that it has flagged. Note that SES represented a combination of GI, GII, and GIII stimulus sizes as shown in Figure 2. Although SES identified a similar number of 'events' compared with GI and GII, it had relatively fewer 'events' that had reached the measurement floor. GIII, GIV, and GV all identified fewer 'events' and had correspondingly fewer occasions where 'events' reached the measurement floor.

When locations reaching the measurement floor were excluded from analysis, the correlation between S_P and S_A was slightly poorer than when all points were included for GIII ($r = 0.4828$, $R^2 = 0.2331$, $P < 0.0001$) and for GV ($r = 0.5597$, $R^2 = 0.3132$, $P < 0.0001$; Fig. 5A). However, there was a greater proportion of points where the difference between S_P and S_A were within ±3 dB (GIII: 75.7%; GV: 84.0%) compared with when all points were included for analysis (Fig. 5B). These results indicated that points reaching a measurement floor of 19 dB for GIII (i.e., established and deep VF defects) affected the agreement between S_P and S_A, and thus the accuracy of the predictive model.

Comparing a Uniform Rate of VF Decline With a Rate of Change Scaled by the Slope of Partial Summation

One of the assumptions outlined in the Methods section was the use of a scaled rate of change for sensitivities measured using stimulus sizes larger than Ac (i.e., $-0.8 \times n2$ dB/y). The actual numeric value of the rate of change itself is less important (see Discussion) than the differences in scaling between stimulus sizes smaller than or larger than Ac. The scaling assumes this relationship due to the predominantly rightward shift in the spatial summation function in glaucoma, as shown by previous studies,[6,8,9] which suggests that, to some degree, there is a difference in rate of change preserved until Ac becomes sufficiently large, at which one stimulus sizes larger than Ac will progress at a similar rate (Fig. 6A).

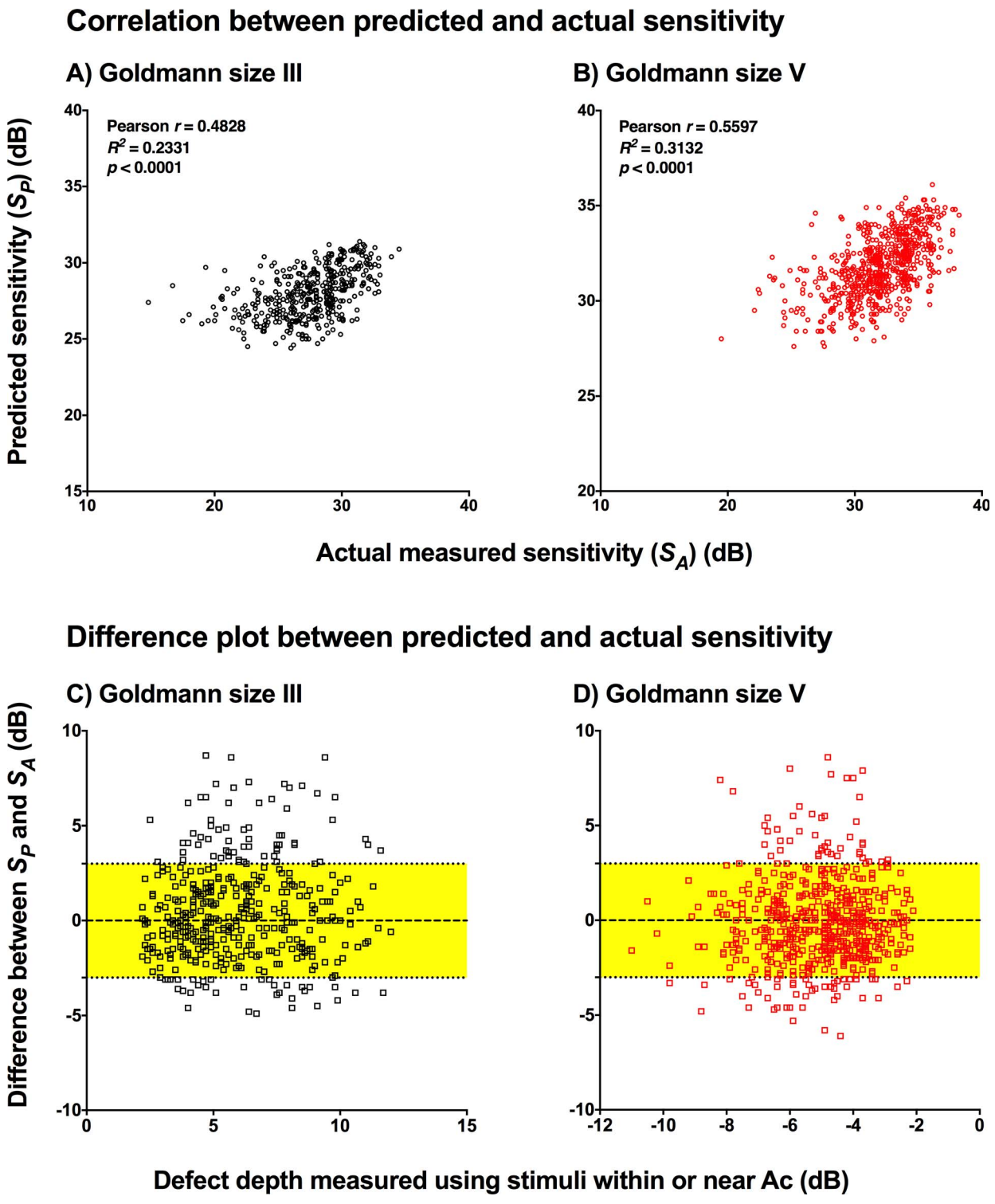

Figure 5. A comparison of sensitivity predicted by the model (S_P) and actual measured sensitivities (S_A) for Goldmann size III (GIII, *black*) and Goldmann size V (GV, *red*) at locations where stimulus size within or near the size of Ac (SES) detected an 'event' after excluding 'events', which had reached the measurement floor when using SES (not the difference in *x*-axis scale compared with Fig. 4). *Top row*: the correlation between S_P and S_A for GIII (A) and GV (B). Pearson's *r*, R^2, and *P* values shown on each figure. *Bottom row*: difference plot between S_P and S_A (in dB) as a function of visual field defect depth (in dB) when measured SES for GIII (C) and GV (D). The *dashed black line* indicates no difference between S_P and S_A ($y = 0$), and the *yellow area* indicates the region of ± 3 dB, which is the approximate test–retest variability of the instrument.

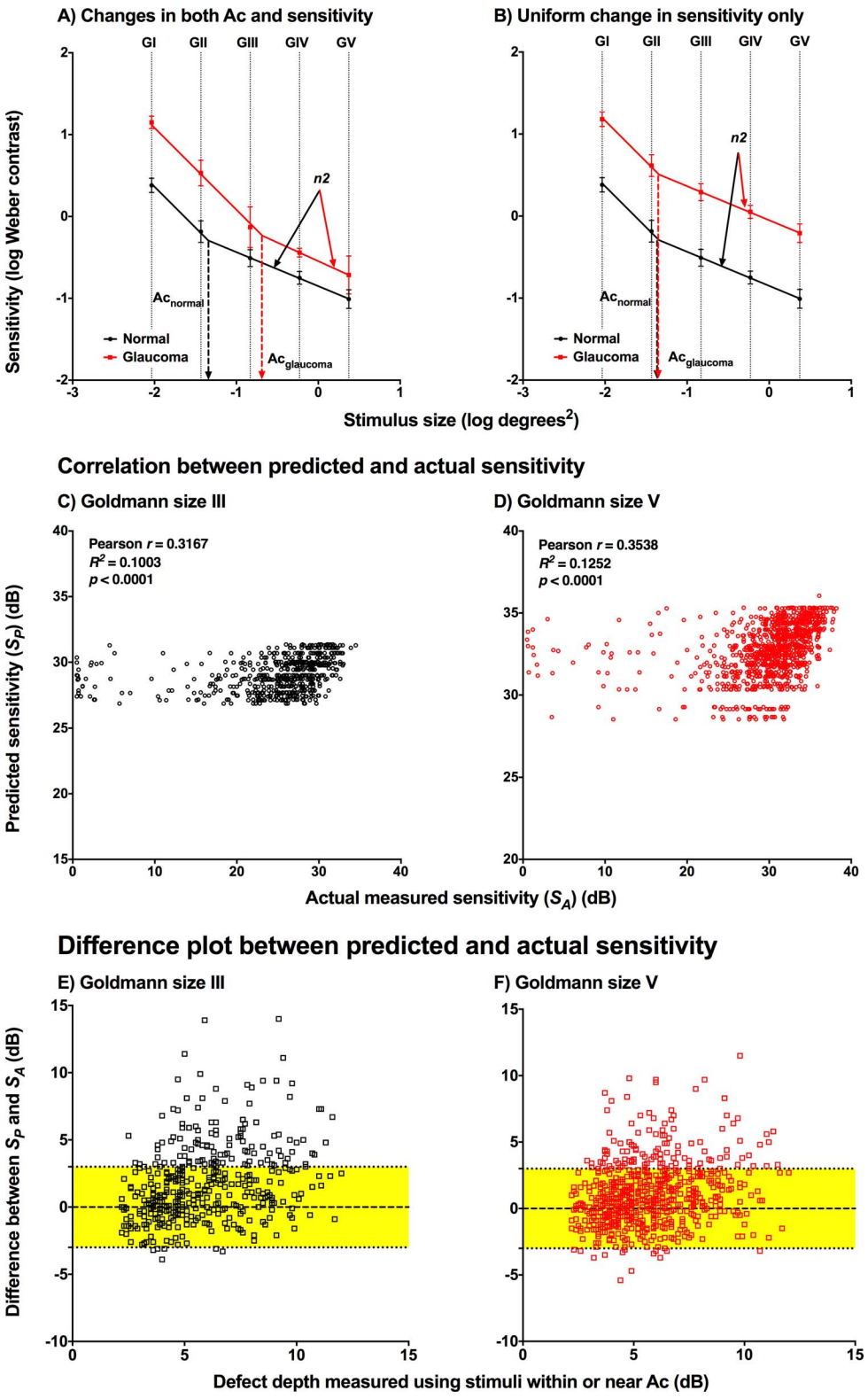

Figure 6. A comparison of spatial summation functions when using a model assuming both a change in sensitivity (a very slight upward shift in the function) and in Ac (predominantly rightward shift) (A) and a model assuming uniform progression rates (i.e., not explained by the change in Ac) across all stimulus sizes (B). A comparison of sensitivity predicted by the model (S_P) and actual measured sensitivities (S_A) for Goldmann size III (GIII, *black*) and Goldmann size V (GV, *red*) at locations where stimulus size within or near the size of Ac (SES) detected an 'event' after excluding 'events' that had reached the measurement floor when using SES, as per Figure 5.

This is in contrast to a model that assumes uniform progression across all stimulus sizes (Fig. 6B). Such a model should therefore display no benefit in using alternative stimulus sizes, given the narrower distribution limits found using larger stimulus sizes in comparison to smaller sizes.

We therefore tested this uniform progression model using the same predictive model as described above. Instead of scaling progression of sensitivities using *n2* for stimulus sizes larger than Ac, all stimulus sizes (SES, GIII, and GV) were modeled with the same progression rate (−0.8 dB/y) to obtain S_P once again. For the purposes of this analysis, points reaching the 19-dB floor were excluded. In contrast to the model assuming changes in both sensitivity and Ac, correlations were poorer for both GIII ($r = 0.3167$ vs. 0.4828) and GV ($r = 0.3538$ vs. 0.5597) for the uniform progression rate model (Figs. 6C, 6D). In addition, the difference plots showed a smaller proportion of points, which were within ±3 dB compared with the original model (GIII: 57.3% vs. 75.7%; GV: 67.2% vs. 84.0%). In combination, these results suggest that the model that assumes uniform progression rate does not reflect the actual progression of glaucoma patients as well as one that assumes changes in both sensitivity and Ac (i.e., the use of a scaled progression rate).

Are the Detection Benefits of Smaller Stimulus Sizes Related to Signal-to-Noise Ratio?

Rountree et al.[18] recently reported the use of SNR as a metric for evaluating the clinical utility of four different stimulus types for assessing glaucomatous VF defects. Although the present study was not designed to directly address this, we performed a similar calculation to determine whether this could account for the benefits in defect detection found using smaller stimulus sizes. For each stimulus size, SNR was calculated at locations with an 'event' (for that specific size) by dividing the magnitude of defect depth by the SD of healthy subjects (note that due to the design of the study, SD data for the glaucoma subjects could not be determined at an individual level) at that test location (i.e., sensitivity difference / SD). For all 'events', the SNRs (median, [IQR]) were as follows: SES, 2.53 [1.27, 4.04]; GI, 2.09 [1.38, 3.37]; GII, 2.55, [1.49, 3.86]; GIII, 1.91 [1.00, 2.79]; GIV, 1.43 [0.66, 3.02]; and GV, 1.35 [0.51, 2.45]. There were significant differences between SES and GIV ($P = 0.0233$) and GV ($P = 0.0045$), but not for GI ($P = 0.3637$), GII ($P = 0.7258$), or GIII ($P = 0.1766$). In comparing GIII and SES, the reason for no significant difference was likely due to the overlapping points; if those were excluded, then there is a significant difference between their respective SNRs (GIII, 1.80 [0.79, 4.45]; SES, 2.33 [1.04, 5.24]; $P = 0.0107$).

Discussion

In the present study, we present a model predicting lead time between using stimulus sizes that are within or near Ac across the 30-2 test grid compared with usual care, Goldmann size III, and the more recently suggested Goldmann size V stimulus sizes.[25,26] This model builds upon previous studies that have shown that stimulus sizes that are within or near the critical area of healthy subjects (i.e., the normal Ac) identify the greatest number of VF defects, and the greatest depth of defect, compared with stimulus sizes that are larger than Ac.[6–9,16] The present results corroborate previous suggestions that earlier detection of defects is possible using stimulus sizes tailored to the Ac of healthy subjects in a point-wise, location-specific fashion (which we refer to as SES), but further add, for the first time, a prediction which suggests a median lead time of approximately 4 years within a large clinical cohort.

Assumptions of the Predictive Model

The presented model makes a number of assumptions. First, it assumes that the two-line fit, and the subsequently derived spatial summation characteristics (Ac and *n2*), depicts the relationship between sensitivities measured using GI to V. Second, based on this relationship, it assumes that the rate of change in glaucoma may be scaled based on the slope *n2*. Finally, we used a uniform progression rate across the entirety of the 30-2 test grid. These are discussed below.

Whether to use a two-line (or multiply segmented) or a curvilinear fit is debated in the literature. Some studies have suggested the use of curvilinear fits to sensitivities measured using a range of stimulus sizes or durations.[15–17,42–44] For these curvilinear fits, Ac may be estimated from the point at which the slope first deviates from a value of −1. This method has also been used when no curve is fitted at all (e.g., Barlow[45]). Then, the ensuing curve represents the relationship between stimulus sizes within the region of partial summation. On the other hand, a number of

other studies have used two-line, or segmented, fits to describe the relationship between sensitivities measured using different stimulus sizes when using both clinical- and laboratory-based techniques.[6–9,12–14] Though individual pairwise relationships between stimulus sizes within Ac (e.g., GI and GII) and stimulus sizes outside Ac (e.g., GIII–GV) have been shown to be robust in previous studies[8] (Supplementary Figs. S3, S4), the junction of complete and partial summation using a two-line fit has not been well characterized or compared with a curvilinear fit. Future studies would be informative to examine the differences, and whether they are clinically meaningful.

Effect of Progression Rate on the Predictive Model

In the present study, we used an average progression rate of −0.8 dB/y,[38] and assumed its uniformity throughout the 30-2 test grid. This particular progression rate was originally reported in SITA-Standard VF results measured using GIII. In our study, we used this rate for stimulus sizes within Ac, and began at time = 0 (i.e., no perimetric defect) and modeled progression over time. The original study reporting this rate examined patients with established VF defects. The analogous "time point" within the present study at which GIII would also progress at a rate of −0.8 dB/y would be when it has established itself as a defect—but this would follow after a defect has been found using SES; hence, the use of −0.8 dB/y for SES, and a scaled rate for stimulus sizes larger than Ac, including GIII within the central VF.

Previous studies have shown that there are some test locations at which glaucomatous VF defects occur more frequently.[46] However, the change in rates of progression across the VF has not been well established, and future studies in this area would be informative. We used a generalized progression rate (e.g., the "average" progression rate reported by Heijl et al.[38]) that reflects progression across the entirety of the VF, combining locations that may have relatively faster and those with relatively slower rates; thus, the average lead time that we report is also reflective of this generalization. Although some locations may progress faster than others,[47,48] along with having different rates at different stages of the disease process, studies have been confounded by a number of factors, including different glaucoma phenotypes, severity of disease at presentation, treatment regimens, and individual variability—all of these affect not only the initial location of the VF defect, but also its progression rate. Another confounding element is the effect of eccentricity. Due to the greater variability of measurements in the periphery, criteria requiring a greater rate of change for statistical significance in those regions have been suggested.[49] When we applied a faster rate (doubled, i.e., −1.6 dB/y in the present study) in the two peripheral 'rings' of test locations (n = 44) in order to "match" the statistical significance of the central rate, we found similar benefits to using SES compared with both GIII and GV (Fig. 7). The present study considered each point to be independent, while glaucomatous defects may appear in clusters and progression typically occurs in the form of deepening and expansion within a cluster. Again, this model would serve as a useful foundation for integration of point-wise and cluster-based progression data in the future.

Different rates of change[50] would be expected to reveal different lead times (Table 3, also Fig. 7). For the slowest rate of progression that we tested, −0.5 dB/y, there was lead time of greater than 6 years for both GV and GIII. For the fastest rate of progression that we tested, −2 dB/y, there was still a lead time of over 1.5 years for stimulus sizes larger than the size of Ac. Thus, even for fast progressors, there appears to be a benefit in using SES for earlier detection of VF defects.

'Events' Reaching the Measurement Floor

Although the dynamic range of the instrument is 51 dB, the effective dynamic range may be much less, as sensitivity measurements are less reliable at relatively low sensitivities. In the present study, we considered 19 dB measured using GIII (and scaled for other sizes) as a "pseudo" measurement floor, as sensitivity results below this would likely be unreliable. Another reason for excluding sensitivity results below this value is because we specifically examined locations with early glaucomatous VF defects. Locations with defects near the measurement floor are unlikely to provide additional value for early detection, as the defects would be well established by that point. Once a VF defect is established, larger stimulus sizes could be recommended for ongoing monitoring, as small stimulus sizes, while better for detection, would reach the measurement floor due to their limited dynamic range.[22,23,27]

The use of SNR to assess clinical utility of perimetric stimuli (as reported in the Results section) also ties into evaluation of stimulus dynamic range. The present study emphasized a detection strategy;

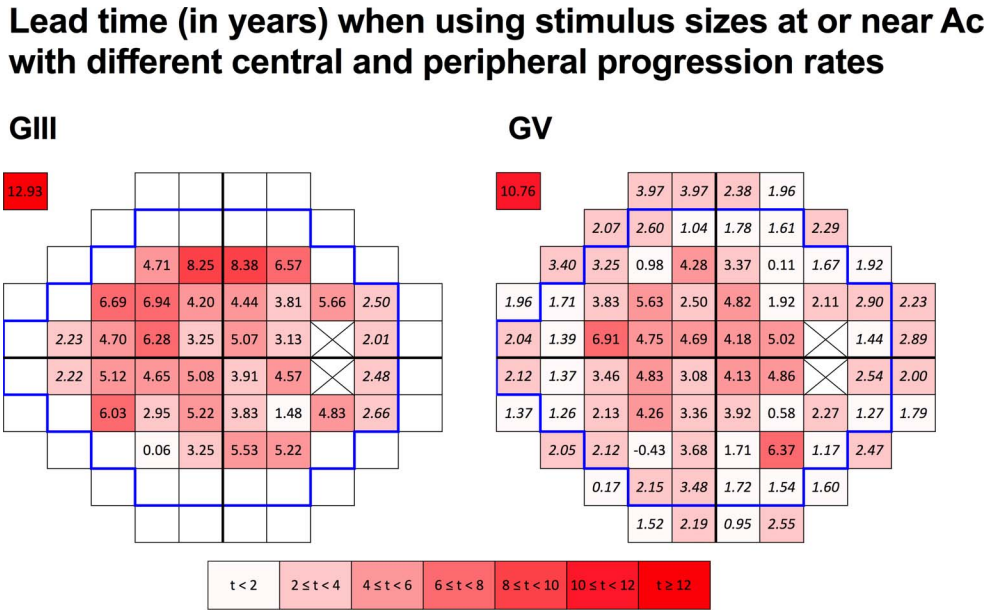

Figure 7. Lead time (in years) when using stimulus sizes at or near the size of Ac compared with GIII and GV across all test locations within the 30-2 test grid, plotted as per Figure 3, but for two different progression rates: −0.8 dB/y for the central points ($n = 31$), and −1.6 dB/y for the two outer concentric "rings" of test locations ($n = 44$, indicated by the italicized numbers).

hence, our utilization of a measurement floor. However, monitoring strategies would need to assess response variability. Although the SNR appeared to account for the benefits of SES and small test stimuli in the present study, it would be highly informative to then determine the response variability characteristics across disease strata to determine its useful dynamic range.[18] A stimulus with a low SNR achieved through high-signal fidelity and low-response variability is desirable.

One of the confounding factors in previous reports on measurement variability has been the use of instruments with fixed-stimulus step sizes (e.g., HFA).[12,27,29] Notably, the amount of energy delivered by different stimulus sizes is different, and these may not necessarily be scaled using conventional techniques. Future studies examining the utility of different Goldmann stimulus sizes would benefit from customizing the step sizes for each size to equate for energy delivered for a more robust comparison.

Enlargement of Ac in Glaucoma and the Predictive Model

Ac has been shown to enlarge in conditions of disease, such as glaucoma. In early stages of the disease, it is likely to remain smaller than the normal Ac value at the corresponding location. As mentioned, stimulus sizes within or near the Ac of healthy subjects (SES) identify a greater amount of sensitivity reduction in diseased states in comparison to stimulus sizes larger than Ac. This was also the reason for our use of a scaled progression rate (Fig. 6). With enlargement of Ac in diseased states, a stimulus size that is within the Ac of disease but larger than the healthy Ac would be expected to show a greater sensitivity reduction compared with a stimulus size that is larger than both critical areas (Fig. 1B). However, the extent by which Ac changes in different stages of disease has not been well characterized, as studies investigating spatial summation in glaucoma have focused on the early stages of the disease.

Because of the enlargement of Ac in glaucoma, we made two predictions regarding the relationship between S_P and S_A. The first prediction was that there would be a greater difference between S_P and S_A

Table 3. Lead Times Gained (in Years: Median and Interquartile Range) Using Different Rates of Progression When Comparing Use of Stimulus Sizes Within or Near the Size of Ac and GIII or GV

Rate of Progression	GIII	GV
−0.5 dB/y	7.7 (6.4, 9.0)	6.3 (4.4, 7.7)
−0.8 dB/y	4.8 (4.0, 5.6)	3.9 (2.8, 4.8)
−1.0 dB/y	3.9 (3.2, 4.5)	3.1 (2.2, 3.9)
−2.0 dB/y	1.9 (1.6, 2.3)	1.6 (1.1, 1.9)

at 'events' where the VF defect is more advanced. The second prediction was that the agreement between S_P and S_A would be worse with GIII compared with GV. As a corollary to the second prediction, we hypothesized that there would be a greater difference between S_P and S_A at a when using GIII compared with GV, as the GIII stimulus size is closer than GV to the size of Ac.

The results showed a greater amount of difference between S_P and S_A as the defect became more advanced, particularly for GIII. The bias of this relationship was that S_P had higher sensitivity estimates compared with S_A, especially at more advanced defects, as seen upon comparing Figures 4 and 5. This would be consistent with an enlargement of Ac,[6–9] as the actual sensitivity reduction could now reflect the measurement made by a stimulus size within the Ac of the glaucoma patient. The discordance was also found to be greater for GIII compared with GV, consistent with the fact that GIII would be closer to being within the size of Ac compared with GV.[6,8,12] The scaled progression rate model that we used out performed the uniform progression rate model through its better correlations and in terms of the proportion of points within ±3 dB, in line with studies showing the largely rightward shift in the spatial summation function in glaucoma.[6,8,9] Further detailed studies investigating the changes in Ac at various stages of glaucoma would be informative to describe the changing relationship between sensitivities measured using different stimulus sizes.

Clinical Implications

The potential benefits of using stimulus sizes within or near Ac have been demonstrated in previous studies.[6,8,9] The lead time suggested by the results of the present study have clinical implications in the choice of stimulus size for detecting early glaucomatous VF defects. In combination, the clinical benefits of SES reinforce the potential paradigm shift in selecting stimulus sizes that may be suitable for different stages of the disease in order to balance the ability for VF defect detection while minimizing measurement variability. However, as mentioned, further refinement for this technique includes the consideration of stimulus step sizes.[18] Our model also lends further support to the notion of modulating both the size and the contrast of the stimulus to optimize early detection of glaucomatous loss by scaling the stimulus to Ricco's area[6,8,9,18]: the apparent sensitivity loss appear to be largely accountable by changes to Ac, rather than by uniform changes in sensitivity across the different stimulus sizes, suggesting tailoring the stimulus size according to the stage of deficit. Interestingly, the lead time when using an average progression rate of −0.8 dB/y was similar to lead times using OCT compared with SAP.[51] It may be possible that the improvement in VF defect detection rate and lead time using SES indicates a potential to improve the structure–function relationship between VF and OCT results.

Study Limitations

Although the present study included 'events' of a wide variety of depths, it was still limited in its cross-sectional design. The inferences made when testing the model using the actual sensitivities measured in glaucoma patients were also based on an average progression rate among all participants.[38] In reality, they may have progressed differently. As the study design aimed to investigate lead time to detecting early glaucomatous defects, it was focused on patients with early stage disease. Future studies with a longitudinal design with patients with different stages of glaucoma would be informative. As suggested in the Results, longitudinal data would also be highly useful for analyzing differences in the rates of change using different stimulus sizes. Although we used a younger normative cohort, the method of age-correction to a 50-year-old equivalent was comparable to using an age-matched cohort, as noted in a recent study[29] and consistent with its clinical usage,[31] and thus, these results appear to be robust (see Supplementary Material). Although SNR appeared to account for the benefits of SES and small stimuli, one of the assumptions is that the relative difference in SD across stimulus sizes may be preserved between the healthy and early glaucoma cohorts, and hence, the SNR might not be expected to change significantly with SD data from the glaucoma patients. Examination of the response variability in glaucoma across different stimulus sizes would be informative to see if these benefits persist.

In summary, we demonstrated significant lead time to the detection of early glaucomatous defects when using stimulus sizes within or near the size of Ac. Using alternative stimulus sizes at different stages of glaucoma may be suggested: smaller sizes for initial defect detection, and larger sizes for later monitoring as sensitivities measured using smaller stimuli may reach the measurement floor. The results in the present study provide a framework for studying further methods for optimizing perimetric stimuli,

including the incorporation of modulations to test size, contrast, duration, and step size.

Acknowledgments

The authors thank Cornelia Zangerl for technical assistance. The authors also thank the anonymous reviewers and editor for their constructive and insightful comments.

Supported by grants from the National Health and Medical Research Council of Australia (NHMRC #1033224; Guide Dogs NSW/ACT are partners in the NHMRC grant); a PhD scholarship provided by Guide Dogs NSW/ACT and an Australian Government Research Training Program PhD Scholarship (JP); and an Australian Research Council Future Fellowship (BVB; FT130100338).

MK and SKK are named inventors on a patent involving the use of different Goldmann target sizes at different visual field locations for contrast sensitivity testing (International Publication Number WO 2014/094035 A1 [USA] and European Patent Number 13865419.9). JP and BVB have no other disclosures.

Disclosure: **J. Phu**, None; **S.K. Khuu**, (P); **B.V. Bui**, None; **M. Kalloniatis**, (P)

References

1. Jampel HD, Singh K, Lin SC, Chen TC, Francis BA, Hodapp E, Samples JR, Smith SD. Assessment of visual function in glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2011;118:986–1002.
2. Phu J, Khuu SK, Yapp M, Assaad N, Hennessy MP, Kalloniatis M. The value of visual field testing in the era of advanced imaging: clinical and psychophysical perspectives. *Clin Exp Optom*. 2017;100:313–332.
3. Lisboa R, Leite MT, Zangwill LM, Tafreshi A, Weinreb RN, Medeiros FA. Diagnosing preperimetric glaucoma with spectral domain optical coherence tomography. *Ophthalmology*. 2012;119:2261–2269.
4. Medeiros FA, Zangwill LM, Bowd C, Mansouri K, Weinreb RN. The structure and function relationship in glaucoma: implications for detection of progression and measurement of rates of change. *Invest Ophthalmol Vis Sci*. 2012;53:6939–6946.
5. Anderson RS. The psychophysics of glaucoma: improving the structure/function relationship. *Prog Retin Eye Res*. 2006;25:79–97.
6. Kalloniatis M, Khuu SK. Equating spatial summation in visual field testing reveals greater loss in optic nerve disease. *Ophthalmic Physiol Opt*. 2016;36:439–452.
7. Mulholland PJ, Redmond T, Garway-Heath DF, Zlatkova MB, Anderson RS. Spatiotemporal summation of perimetric stimuli in early glaucoma. *Invest Ophthalmol Vis Sci*. 2015;56:6473–6482.
8. Phu J, Khuu SK, Zangerl B, Kalloniatis M. A comparison of Goldmann III, V and spatially equated test stimuli in visual field testing: the importance of complete and partial spatial summation. *Ophthalmic Physiol Opt*. 2017;37:160–176.
9. Redmond T, Garway-Heath DF, Zlatkova MB, Anderson RS. Sensitivity loss in early glaucoma can be mapped to an enlargement of the area of complete spatial summation. *Invest Ophthalmol Vis Sci*. 2010;51:6540–6548.
10. Ricco A. Relazione fra il minimo angolo visuale e l'intensita luminosa. *Annali di Ottalmologia*. 1877;6:373–479.
11. Choi AYJ, Nivison-Smith L, Khuu SK, Kalloniatis M. Determining spatial summation and its effect on contrast sensitivity across the central 20 degrees of visual field. *PloS One*. 2016;11.
12. Khuu SK, Kalloniatis M. Standard automated perimetry: determining spatial summation and its effect on contrast sensitivity across the visual field. *Invest Ophthalmol Vis Sci*. 2015;56:3565–3576.
13. Khuu SK, Kalloniatis M. Spatial summation across the central visual field: implications for visual field testing. *J Vis*. 2015;15(1): 15.1.6.
14. Owen WG. Spatio-temporal integration in the human peripheral retina. *Vision Res*. 1972;12:1011–1026.
15. Wilson ME. Invariant features of spatial summation with changing locus in the visual field. *J Physiol*. 1970;207:611–622.
16. Sloan LL, Brown DJ. Area and luminance of test object as variables in projection perimetry: clinical studies of photometric dysharmony. *Vision Res*. 1962;2:527–541.
17. Wilson ME. Spatial and temporal summation in impaired regions of the visual field. *J Physiol*. 1967;189:189–208.

18. Rountree L, Mulholland PJ, Anderson RS, Garway-Heath DF, Morgan JE, Redmond T. Optimising the glaucoma signal/noise ratio by mapping changes in spatial summation with area-modulated perimetric stimuli. *Sci Rep*. 2018;8:2172.
19. Gilpin LB, Stewart WC, Hunt HH, Broom CD. Threshold variability using different Goldmann stimulus sizes. *Acta Ophthalmol (Copenh)*. 1990; 68:674–676.
20. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Am J Ophthalmol*. 2000; 130:689.
21. Gardiner SK, Demirel S, Johnson CA. Modeling the sensitivity to variability relationship in perimetry. *Vision Res*. 2006;46:1732–1745.
22. Gardiner SK, Swanson WH, Demirel S. The effect of limiting the range of perimetric sensitivities on pointwise assessment of visual field progression in glaucoma. *Invest Ophthalmol Vis Sci*. 2016;57:288–294.
23. Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. *Ophthalmology*. 2014;121:1359–1369.
24. Phu J, Kalloniatis M, Khuu SK. Reducing spatial uncertainty through attentional cueing improves contrast sensitivity in regions of the visual field with glaucomatous defects. *Transl Vis Sci Technol*. 2018;7(2):8.
25. Wall M, Doyle CK, Eden T, Zamba KD, Johnson CA. Size threshold perimetry performs as well as conventional automated perimetry with stimulus sizes III, V, and VI for glaucomatous loss. *Invest Ophthalmol Vis Sci*. 2013;54:3975–3983.
26. Wall M, Doyle CK, Zamba KD, Artes P, Johnson CA. The repeatability of mean defect with size III and size V standard automated perimetry. *Invest Ophthalmol Vis Sci*. 2013;54: 1345–1351.
27. Wall M, Woodward KR, Doyle CK, Zamba G. The effective dynamic ranges of standard automated perimetry sizes III and V and motion and matrix perimetry. *Arch Ophthalmol*. 2010;128: 570–576.
28. Garway-Heath DF, Caprioli J, Fitzke FW, Hitchings RA. Scaling the hill of vision: the physiological relationship between light sensitivity and ganglion cell numbers. *Invest Ophthalmol Vis Sci*. 2000;41:1774–1782.
29. Phu J, Khuu SK, Nivison-Smith L, et al. Pattern recognition analysis reveals unique contrast sensitivity isocontours using static perimetry thresholds across the visual field. *Invest Ophthalmol Vis Sci*. 2017;58:4863–4876.
30. Heijl A, Lindgren G, Olsson J. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol*. 1987;105: 1544–1549.
31. Bengtsson B, Olsson J, Heijl A, Rootzen H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand*. 1997;75:368–375.
32. Phu J, Bui BV, Kalloniatis M, Khuu SK. How many subjects are needed for a visual field normative database? A comparison of ground truth and bootstrapped statistics. *Transl Vis Sci Technol*. 2018;7(2):1.
33. Mills RP, Budenz DL, Lee PP, et al. Categorizing the stage of glaucoma from pre-diagnosis to end-stage disease. *Am J Ophthalmol*. 2006;141:24–30.
34. Brown B, Peterken C, Bowman KJ, Crassini B. Spatial summation in young and elderly observers. *Ophthalmic Physiol Opt*. 1989;9:310–313.
35. Dannheim F, Drance SM. Studies of spatial summation of central retinal areas in normal people of all ages. *Can J Ophthalmol*. 1971;6:311–19.
36. Redmond T, Zlatkova MB, Garway-Heath DF, Anderson RS. The effect of age on the area of complete spatial summation for chromatic and achromatic stimuli. *Invest Ophthalmol Vis Sci*. 2010;51:6533–6539.
37. Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci*. 2002;43:2654–2659.
38. Heijl A, Buchholz P, Norrgren G, Bengtsson B. Rates of visual field progression in clinical glaucoma care. *Acta Ophthalmol*. 2013;91:406–412.
39. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol*. 1989;108:130–135.
40. Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. 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.
41. Motulsky HJ, Brown RE. Detecting outliers when fitting data with nonlinear regression - a new method based on robust nonlinear regression

and the false discovery rate. *BMC Bioinformatics*. 2006;7:123.
42. Graham CH, Margaria R. Area and the intensity-time relation in the peripheral retina. *Am J Physiol*. 1935;113:299–305.
43. Latham K, Whitaker D, Wild JM. Spatial summation of the differential light threshold as a function of visual field location and age. *Ophthalmic Physiol Opt*. 1994;14:71–78.
44. Sloan LL. Area and luminance of test object as variables in examination of the visual field by projection perimetry. *Vision Res*. 1961;1:121–138.
45. Barlow HB. Temporal and spatial summation in human vision at different background intensities. *J Physiol*. 1958;141:337–350.
46. Schiefer U, Papageorgiou E, Sample PA, et al. Spatial pattern of glaucomatous visual field loss obtained with regionally condensed stimulus arrangements. *Invest Ophthalmol Vis Sci*. 2010; 51:5685–5689.
47. Shon K, Wollstein G, Schuman JS, Sung KR. Prediction of glaucomatous visual field progression: pointwise analysis. *Curr Eye Res*. 2014;39: 705–710.
48. Su D, Park SC, Simonson JL, Liebmann JM, Ritch R. Progression pattern of initial parafoveal scotomas in glaucoma. *Ophthalmology*. 2013;120: 520–527.
49. Viswanathan AC, Fitzke FW, Hitchings RA. Early detection of visual field progression in glaucoma: a comparison of PROGRESSOR and STATPAC 2. *Br J Ophthalmol*. 1997;81:1037–1042.
50. Chauhan BC, Garway-Heath DF, Goni FJ, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol*. 2008;92:569–573.
51. Kuang TM, Zhang C, Zangwill LM, Weinreb RN, Medeiros FA. Estimating lead time gained by optical coherence tomography in detecting glaucoma before development of visual field defects. *Ophthalmology*. 2015;122:2002–2009.