Clinical Evaluation of Swedish Interactive Thresholding Algorithm–Faster Compared With Swedish Interactive Thresholding Algorithm–Standard in Normal Subjects, Glaucoma Suspects, and Patients With Glaucoma

JACK PHU, SIEU K. KHUU, ASHISH AGAR, AND MICHAEL KALLONIATIS

• PURPOSE: To compare the visual fields results obtained using the Swedish interactive thresholding algorithm-Standard (SS) and the Swedish interactive thresholding algorithm-Faster (SFR) in normal subjects, glaucoma suspects, and patients with glaucoma and to quantify potential time-saving benefits of the SFR algorithm.

• DESIGN: Prospective, cross-sectional study.

• METHODS: One randomly selected eye from 364 patients (77 normal subjects, 178 glaucoma suspects, and 109 patients with glaucoma) seen in a single institution underwent testing using both SS and SFR on the Humphrey Field Analyzer. Cumulative test time using each algorithm was compared after accounting for different rates of test reliability. Pointwise and cluster analysis was performed to determine whether there were systematic differences between algorithms.

• RESULTS: Using SFR had a greater rate of unreliable results (29.3%) compared with SS (7.7%, $P < .0001$). This was mainly because of high false positive rates and seeding point errors. However, modeled test times showed that using SFR could obtain a greater number of reliable results within a shorter period of time. SFR resulted in higher sensitivity values (on average 0.5 dB for patients with glaucoma) that was greater under conditions of field loss (< 19 dB). Cluster analysis showed no systematic patterns of sensitivity differences between algorithms.

• CONCLUSIONS: After accounting for different rates of test reliability, SFR can result in significant time savings compared with SS. Clinicians should be cognizant of false positive rates and seeding point errors as common sources of error for SFR. Results between algorithms are not directly interchangeable, especially if there is a visual field deficit < 19 dB. (Am J Ophthalmol 2019;208: 251–264.

AJO.com Supplemental Material available at AJO.com.

Accepted for publication Aug 19, 2019.

From the Centre for Eye Health (J.P., A.A., M.K.) and the School of Optometry and Vision Science (J.P., S.K.K., M.K.), University of New South Wales, Kensington, and the Department of Ophthalmology (A.A.), Prince of Wales Hospital, Randwick, New South Wales.

Inquiries to Jack Phu, Centre for Eye Health, Rupert Myers Building South Wing, University of New South Wales, UNSW Sydney 2052, New South Wales, Australia; e-mail: jphu@cfeh.com.au

PERIMETRIC TESTING IN CLINICAL PRACTICE[1,2] IS AN essential part of the diagnosis and management of glaucoma, the most common optic neuropathy and the leading cause of irreversible blindness. One of the most commonly used automated perimeters for glaucoma assessment is the Humphrey Field Analyzer (Carl Zeiss Meditec, Dublin, California, USA),[3–7] and algorithms that have been available on the instrument include the full threshold, Swedish interactive thresholding algorithm (SITA)-Standard (SS), SITA-Fast, and Fastpac paradigms.[8] Specifically, SS has garnered widespread use in clinical practice for glaucoma assessments because of its shorter test duration relative to full threshold[9] while generally preserving the fidelity of the resultant estimated sensitivity result.[10–13]

One significant problem with clinical perimetry is the variability in sensitivity measurement,[14] particularly in conditions of disease[15–17] that may stem from a number of causes.[18–22] In the face of known issues with variability within and between tests, there has been a paradigm shift towards conducting more visual field tests.[23–25] Specifically, the concept of frontloading visual fields has been recommended at a baseline examination and at follow-up in order to minimize variability to maximize the detection of progressive loss.[23] In this way, a large amount of perimetric data can be collected to increase the likelihood of identifying patients exhibiting different rates of disease progression.[26,27]

However, the implementation of recommendations for increased frequency of testing has been challenging, with issues of timing and availability of resources being cited as barriers, despite long-term cost savings.[28] Alternative approaches, such as the use of previous sensitivity information,[29] adaptation to regions of interest, such as scotomata,[30] censoring of unreliable test points to reduce the number of stimulus presentations,[31] and the use of

structural information for targeting specific regions of the visual field[32,33] remain theoretical at present and have not yet been widely implemented in practice.

One recent suggestion has been the use of SITA-Faster (SFR),[34] an updated version of the SITA family of threshold-estimating paradigms, that has been shown to reduce testing time by approximately 50% compared with SS. Numerous modifications to SS and SITA-Fast leading to SFR have enabled this reduction in test time, including eliminating stimulus presentation delay and modifications to the starting stimulus intensity and staircase reversals to arrive at the sensitivity threshold at the primary "seeding" points (the 4 points at the 9°:9° position in each quadrant that provide an initial estimate of sensitivity in adjacent test locations). In a comparison of the 3 SITA paradigms, Heijl and associates[34] demonstrated that SITA-Fast and SFR produced almost identical results in terms of sensitivity and probability scores, and small differences between SFR and SS. The logic was therefore extended to potentially replacing SS with SFR in order to meet an increasing demand of more perimetric tests in clinical practice. Before this can be done, more evidence is required to determine the clinical utility of the SFR algorithm compared with the currently commonly used SS.

The goal of the present study was to compare the use of SS and SFR to address 3 practical questions related to their clinical use, adding to the work already presented by Heijl and associates.[34] First, we determined the number of unreliable visual field test results and the reasons for their occurrence, and then related these back to the modifications of SFR. These results were used to address the second aim, which was to determine whether SFR results in an overall saving of clinical testing time, after factoring in retesting of unreliable results. There may be a critical point where net test time taken to arrive at the same number of reliable visual field test results are equated when using SS and SFR. Third, we used a pointwise approach to ascertain location-specific differences in sensitivity or probability score values between the 2 paradigms and confirmed these using cluster analysis.

METHODS

• **SUBJECTS:** This cross-sectional study adhered to the Declaration of Helsinki and ethics approval was provided by the Human Research Ethics Committee of the University of New South Wales. Patients seen at the Centre for Eye Health between June 1, 2018 and June 30, 2019 were examined for their suitability for the study. These patients consisted of those who were referred to the Centre for Eye Health for comprehensive glaucoma assessment. Eligibility criteria for inclusion were: age 18 years or over; patients consenting to use of their clinical data for research and teaching; no ocular, systemic or neurological comorbidities that would confound the visual field test result; no history of ocular surgery aside from unremarkable selective laser trabeculoplasty, peripheral iridotomy or cataract surgery and intraocular lens implantation (all within <2 years); no history of ocular trauma; spherical equivalent refractive error between +8.00 diopters (D) and −8.00 D (with no evidence of trial lens rim artefacts); and the inability to conduct a perimetric test.

Patients formed 3 diagnostic categories, the first 2 of which were similar to the criteria presented by Heijl and associates.[34] First, patients with glaucoma were diagnosed as per current glaucoma guidelines,[35,36] including clear characteristic glaucomatous structural anomalies at the optic nerve head (including, but not limited to: increased cup-to-disc ratio, cup-to-disc asymmetry, and neuroretinal rim thinning or notching) and/or retinal nerve fiber layer (corresponding to the aforementioned neuroretinal rim changes), with or without corresponding visual field loss and with or without elevated intraocular pressure. Second, glaucoma suspects included those being followed on the basis of clinical features suspicious for but not conclusive for glaucoma, including elevated intraocular pressure (defined as >21 mm Hg on applanation tonometry), suspicious optic disc or retinal nerve fiber layer changes, significant systemic, ocular, or family risk factors for glaucoma, or suspicious visual field results. By definition, none of the glaucoma suspects were receiving treatment for glaucoma; treated ocular hypertensive patients were excluded. Third, normal subjects consisted of patients with completely normal ophthalmic findings: no evidence of optic nerve pathology, no increased intraocular pressure, no other retinal or neurologic diseases affecting the visual pathway, and visual fields that did not have defects expected in the presence of pathology. All patients had previous experience performing perimetric testing. Clinical determinations were made by at least 1 (generally 2) examining clinicians, and the results were subsequently also reviewed by at least 2 clinicians for agreement.

• **VISUAL FIELD TESTING:** All patients underwent automated perimetric testing using the SS and SFR paradigms on the Humphrey Field Analyzer 24-2 test grid. Testing for both was conducting at the same clinical visit, and the test order (SS or SFR first) was randomized. One eye was randomly selected for testing for all subjects. Pointwise sensitivity and probability score data were flipped to right eye orientation for all subjects for analysis. Randomization of the eye was performed before the clinic visit. The perimetrist only measured visual acuities before perimetric testing, and all other routine clinical procedures were performed after completion.

The perimetrist administered the test (unmasked to the algorithm) as per the instructions presented on the screen by the instrument before test commencement, as per the methods of Heijl and associates.[34] The perimetrist monitored the patient carefully throughout the test, with proper head, eye, and lens positioning maintained to avoid lens

rim artifacts. If during the test there were signs of unreliability, such as gaze instability, high false positive rates, inattentiveness, or others, the perimetrist was permitted to stop the test, to reiterate the instructions to the patients, and to restart once.[34] After a second failed attempt, the patient was excluded from the study. Once the second attempted of testing was completed, the result was accepted for this study, regardless of the reliability outcome.

• **AIM 1: EVALUATION OF TEST RELIABILITY AND TEST DURATION:** The first aim of the study was to evaluate test reliability. SS has a number of indices for test reliability: fixation losses determined using the Heijl–Krakau method, false positive catch trials, false negative catch trials, and the gaze tracker.[37–39] These are commonly reported in the literature as criteria for result reliability, and therefore 2 specific criteria of reliability for a SS result were: 1) fixation losses >20% in conjunction with significant deviations on the gaze tracker (defined as >6° of eye movement >20% of the time) and 2) a >15% false positive rate. The gaze tracker results are presented as scalar but not vector units, so directional movements cannot be extrapolated. There must also be no head tracker deviations triggering instrument warnings. As per more recent studies,[40,41] we did not use false negative rate as an index of poor reliability because it is known to be elevated in disease. SFR however only uses 2 of these indices: the false positive catch trials and the gaze tracker, and does not use fixation losses found using the Heijl-Krakau method.[34] We therefore only used the criteria of a >15% false positive rate and >6° of eye movement >20% of the time as cutoffs for reliability for a SFR result, and disregarded the Heijl-Krakau blind spot monitoring result. We compared the performance of both algorithms head-to-head and therefore used the same criteria for reliability for assessing both field results: the false positive cutoff and gaze tracker errors (also in line with recent usage for SS).

Using the above, we ascertained the number of unreliable results using each paradigm, and for the final pointwise and cluster analysis only included those patients who had reliable results for both tests. For the unreliable results, we examined the causes for low test reliability.

• **AIM 2: MODELING TEST DURATION DIFFERENCES AFTER EVALUATING OVERALL TEST RELIABILITY:** Our second aim was to examine overall differences in test duration given scenarios of reliable and unreliable test results. We modeled these using 2 methods. The first model was to determine the time it would take to obtain x number of reliable visual field results, for example within an individual patient. We factored in possible occurrences of unreliable visual fields in between the reliable results. Because only permutations where there were 0, 1, or 2 unreliable visual fields were considered, the probabilities of these events occurring given the need for x number of reliable visual fields were normalized. These were done using the following equations:

$$p_{y=0} = p_R^x \tag{1}$$

$$p_{y=1} = x \times p_R^x \times p_U \tag{2}$$

$$p_{y=2} = \left(x \times \frac{x+1}{2}\right) \times p_R^x \times p_U^2 \tag{3}$$

The sum of these was p_y, which represented the total of all probabilities within that set of conditions. Therefore, each normalized probability was $p_{y=0}/p_y$, $p_{y=1}/p_y$, and $p_{y=2}/p_y$.

From this, we determined the average net time needed to obtain x number of reliable visual fields. The probability of the reliable visual field occurring was multiplied by the time taken to complete the field test, and the probabilities of the unreliable fields tests were multiplied by the test duration of an unreliable result. Given the possible permutations, the net time for x reliable fields is given by the sum of time for combinations with no unreliable results (equation 4, t_0), 1 unreliable result (equation 5, t_1), or 2 unreliable results (equation 6, t_2):

$$t_0 = \frac{p_{y=0}}{p_y} \times (t_R) \tag{4}$$

$$t_1 = \frac{p_{y=1}}{p_y} \times x \times \left(t_R^x + t_U\right) \tag{5}$$

$$t_2 = \frac{p_{y=2}}{p_y} \times \left(x \times \frac{x+1}{2}\right) \times \left(t_R^x + t_U^2\right) \tag{6}$$

Where p_R and p_U were the probabilities of obtaining a reliable or unreliable result, respectively, and t_R and t_U were the test durations for the reliable or unreliable results, respectively. This net result of cumulative test time was then plotted as a function of number of desired total number of reliable tests, x. As we had collected these data prospectively and consecutively, we made no predictions regarding the potential crossover points of levels of x or y between SFR and SS.

The second model modeled the time required to obtain at least 1 reliable result across a patient cohort of a certain size. We had 3 conditions: no retest allowed, 1 retest allowed, and 2 retests allowed. To do this, we multiplied the number of subjects within the cohort x with the probabilities of obtaining a reliable (p_R) or unreliable (p_U) visual field result. These were then multiplied by the time taken to obtain a reliable (t_R) or unreliable (t_U) visual field result. The cumulative times were summed to obtain the total time given a set of x patients. For conditions where retests were allowed, the same principle was applied, with a smaller set of x x p_R, representing a retest of those with an initial unreliable result.

For the purposes of this study, both models assumed that the likelihood of rates of low reliability were fixed, though clinically it may be affected by factors such as learning effects and stage of disease.

• AIM 3: POINTWISE DIFFERENCES IN SENSITIVITY AND PROBABILITY SCORES USING CLUSTER ANALYSIS: Our third aim was to examine whether pointwise and systematic differences exist in terms of sensitivity and probability score results. For the sensitivity (in dB) results, we compared the values that were directly extracted from the visual field printouts. The probability scores, typically expressed in grayscale, from $P < .05$ to $P < .005$, were extracted from the pattern deviation map. We assigned an ordinal score to this scale, as per our recent work[42]: $P < .05 = 1$, $P < .02 = 2$, $P < .01 = 3$, and $P < .005 = 4$. A point with no statistically significant reduction was given a score of 0. Though the differences between the probability levels are not equal, the use of an ordinal score can be more relatable to clinical practice because these present a fixed visual cue for the likelihood of abnormality.

Cluster analysis was then used as a complementary technique to determine if there were locations within the visual field that differed in the same fashion between the algorithms. We have previously reported on cluster analysis across the visual field for highlighting contrast sensitivity isocontours,[43,44] which is slightly different to other approaches seeking to characterize archetypes of deficit patterns.[45] Additional details regarding this analysis are provided in the Supplemental Material (available at AJO.com).

• STATISTICAL ANALYSES: All data were extracted directly from the Humphrey Field Analyzer printout, and clinical information relevant to the present study was also similarly obtained. From the printout, we extracted test time, global indices (mean deviation and pattern standard deviation), pointwise sensitivity values and probability scores (as described above), and reliability indices. Differences were compared intraindividually, with the resultant data pooled across diagnostic groups (normal subjects, glaucoma suspects, and patients with glaucoma).

GraphPad Prism software (version 7; GraphPad, La Jolla, California, USA) was used to perform statistical analysis. A D'Agostino and Pearson omnibus normality test ($\alpha = 0.05$) was significant ($P < .05$) for differences in mean deviation, pattern standard deviation, and false positive rate were not normally distributed. Normally distributed data were analyzed using parametric statistics, and nonnormally distributed data were analyzed using nonparametric statistics. Proportion data were compared using the Fisher exact test. Correlations were performed using the Pearson r for parametric data and Spearman rho for nonparametric data.[46]

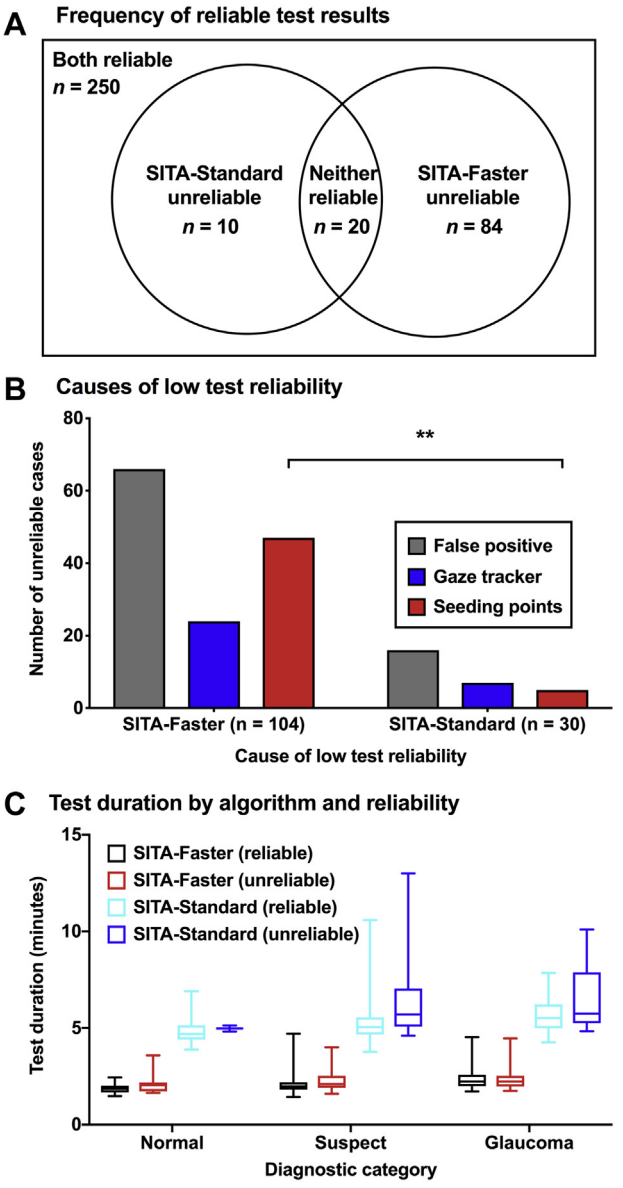

FIGURE 1. (A) Venn diagram of the frequency of reliable visual field test results for each test paradigm. (B) Causes of cases of low test reliability for each test paradigm. The asterisk indicates a significant difference between the Swedish interactive thresholding algorithm (SITA)-Standard and SITA-Faster at $P < .01$. Note that some patients had multiple sources of error. (C) Test duration (in minutes) for each diagnostic category, divided into SITA-Faster and SITA-Standard and whether or not the test result was reliable. The line, boxes, and tails represent the median, quartiles, and range, respectively.

RESULTS

• AIM 1: EVALUATION OF TEST RELIABILITY AND TEST DURATION: During the study period, 364 patients completed visual field testing using both SS and SFR. The demographic characteristics of the final sample of 250 subjects with reliable results on both algorithms (Figure 1A) are shown in Table 1. The relatively younger cohort of the normal subjects was likely a product of the consecutive nature of subject testing. The overall distribution of subjects showed a generally demographically diverse cohort. Glaucoma patients were mostly early stage

TABLE 1. Demographic Characteristics of the Patients With Reliable Visual Field Results Using Both Swedish Interactive Thresholding Algorithm-Standard and Swedish Interactive Thresholding Algorithm-Faster

	Normal Subjects (n = 54)	Glaucoma Suspects (n = 120)	Patients With Glaucoma (n = 76)
Mean age, y (SD)[a]	48.2 (11.9)	59.1 (11.8)	63.5 (14.1)
Males, n (%)[b]	22 (40.7)	70 (58.3)	43 (56.6)
Ethnicity, n (%)[c]			
White	32 (59.3)	85 (70.8)	46 (60.5)
Asian	20 (37.0)	27 (22.5)	23 (30.3)
Indian or Pakistani	1 (1.9)	5 (4.2)	5 (6.4)
Hispanic	1 (1.9)	2 (1.7)	0 (0)
African	0 (0)	0 (0)	2 (2.6)
Aboriginal	0 (0)	1 (0.8)	0 (0)
Median SITA-Standard mean deviation, dB (IQR)[a]	−0.14 (−0.75 to 0.48)	−0.21 (−1.5 to 0.58)	−2.02 (−4.62 to −0.84)
Median SITA-Standard pattern standard deviation, dB (IQR)[a]	1.35 (1.16–1.57)	1.58 (1.37–1.84)	2.58 (2.00–4.17)
Median spherical equivalent refractive error, Diopters (IQR)[d]	−0.25 (−1.25 to 0.25)	0 (−1.25 to +1.00)	0 (−2.31 to +1.31)

IQR = interquartile range; SD = standard deviation; SITA = Swedish interactive thresholding algorithm.
[a]Kruskal-Wallis test showed significant differences between the groups ($P < .0001$).
[b]χ^2 test showed no significant difference between the groups ($P = .1640$).
[c]χ^2 test showed no significant difference between the groups ($P = .1089$).
[d]Kruskal-Wallis test showed no significant difference between groups ($P = .6559$).

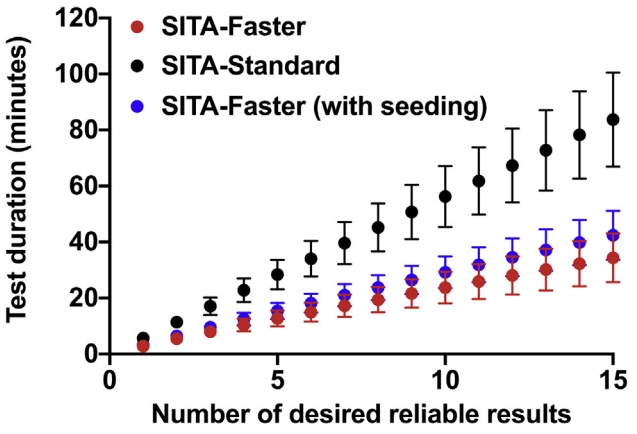

FIGURE 2. Cumulative test duration (in minutes) as a function of number of desired reliable visual field test results modeled using a condition of 2 maximum unreliable test results obtained during a representative test period of 5 years (3 fields per year) for Swedish interactive thresholding algorithm (SITA)-Standard (red) and SITA-Faster (black). We also added a new model, SITA-Faster with seeding points, which was the same as SITA-Faster, plus 30 seconds to test duration and reducing the probability of unreliability (blue). The error bars indicate 1 standard deviation.

glaucoma, with a number (n = 16/76, 21.1%) of moderate and advanced stage glaucoma patients,[47] based on their SS mean deviation scores (median −2.02 dB [interquartile range −4.62 to −0.84 dB; full range −30.62 to +1.08 dB]).

There was no significant difference in the proportion with reliable results between normal subjects (70.1%; 54/77), glaucoma suspects (68.0%; 120/178), and patients with glaucoma (69.7%; 76/109) ($P = .9084$). There were significantly fewer instances where SS was unreliable compared with SFR (8.2% vs 28.6%, $P < .0001$). When SFR was performed first (n = 181), there was a small increase in the number of overall unreliable visual fields results (34.3% vs 25.2% when SS was performed first; n = 183; $P = .0121$). The relative proportion of times where SFR and SS fields were unreliable remained similar (9.4% vs 34.3% [SFR was unreliable 3.6 times more often than SS] when SFR was first and 7.1% vs 23.0% [3.2 times] when SS was first). However, given the unequal number of overall reliable results and distribution of subjects depending on the algorithm used for initial testing, we combined the results for subsequent analysis.

The most common cause of low test reliability across both algorithms was a false positive rate >15% (SFR 66/104 [63.5%]; SS 16/30 [53.3%]), occurring at a similar rate across both tests (Figure 1B; $P = .3159$). There were significantly more instances of seeding point errors (defined as any combination of the 4 points tested during the initial phase of the test as having a combined statistically significant reduction of $P < .0001$ [multiplication of P values] or less in the absence of other pathologic depressions or correlating structural loss, such as cataracts or established scotomata) that had led to poor reliability in SFR (47/104, 45.2%) compared with SS (5/30, 16.7%) ($P = .0047$).

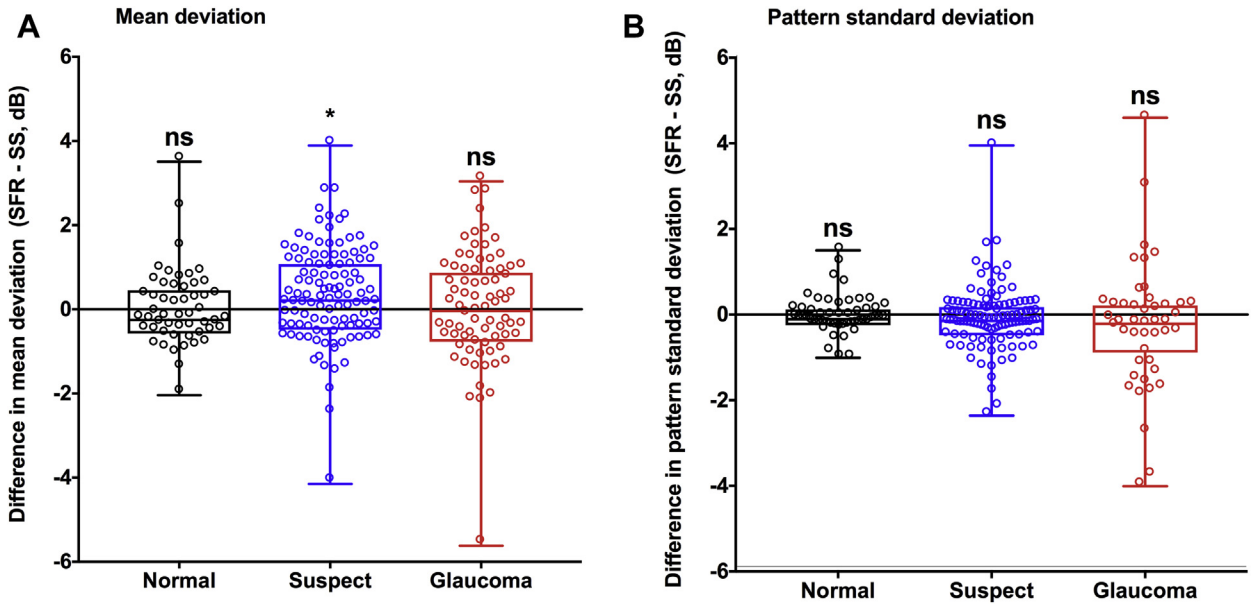

FIGURE 3. Difference between Swedish interactive thresholding algorithm (SITA)-Faster (SFR) and SITA-Standard (SS) in mean deviation (A) and pattern standard deviation (B) in dB for normal subjects, glaucoma suspects, and patients with glaucoma. A positive value on the y axis represents a more positive value obtained using SITA-Faster. The solid black horizontal line indicates no difference between the algorithms. The asterisk indicates a significant difference from 0 using a 1-sample *t* test ($P < .05$). ns = no significance.

There was no difference between groups for the gaze tracker errors (SFR 24/104 [22.1%]; SS 7/30 [23.3%], $P = .9766$).

Test durations were shorter for all 3 groups when using SFR compared with SS when considering reliable results (median difference 182 seconds [interquartile range 161–207.3 seconds], $P < .0001$). Two-way analysis of variance revealed a main effect of diagnostic group on the overall test time, with patients with glaucoma specifically exhibiting a greater relative difference in test duration between the SITA algorithms and different levels of reliability compared with normal and glaucoma suspects ($F_{2,718} = 11.61$, $P < .0001$; Figure 1C). On average, there was a 9.7% increase in test duration for an unreliable result using SFR and a 20.4% increase for SS.

• **AIM 2: MODELING TEST TIME BASED ON OVERALL EVALUATION OF VISUAL FIELD RELIABILITY:** After identifying patients supplying reliable and unreliable data in aim 1, we sought to model test time to determine if net time savings using SFR can be achieved. The modeled cumulative test times were based on 262 reliable SFR (mean, standard deviation [SD] 126.2 ± 28.2 seconds), 334 reliable SS (315.8 ± 57.2 seconds), 104 unreliable SFR (138.9 ± 38.3 seconds), and 30 unreliable SS (381.4 ± 113.4 seconds) results, and the probability for reliable and unreliable occurrences found using the data presented in Figure 1. Specifically leveraging the data presented in Figure 1, we also modeled an additional theoretical "algorithm," SFR "with seeding," which includes an additional 30 seconds that is typically used by SS or full threshold for carefully establishing the initial 4 seeding points. This is notably one of the modifications applied to SFR to shorten testing times: the starting stimulus intensity has a lower luminance, and only 1 reversal is performed, which we have demonstrated to lead to cases of low test reliability in Figure 1. Based on the results of Figure 1, this modification is speculated to reduce the number of unreliable cases of SFR by 23, with a new unreliability rate of 17.8% (a 37.3% reduction).

For this model, average total predicted test times were modeled for conditions where ≤2 unreliable visual field results were accepted in ≤15 tests (a 5-year period if 3 reliable results are desired per year); because the results for 3 or 4 unreliable visual field conditions were similar, these are not shown for clarity (Figure 2). The lower rate of unreliable test results when using SS compared with SFR did not appear to cause a convergence in overall time to obtain ≤15 visual reliable fields tests when modeled in this manner. This was still true when we modelled an alternative SFR algorithm with seeding points. SFR therefore still results in overall time saving (Supplemental Figure 1A-C; Supplemental Material available at AJO.com).

• **AIM 3: GLOBAL AND POINTWISE DIFFERENCES BETWEEN SS AND SFR SENSITIVITIES:** Aim 3 was assessed solely with the data from subjects with reliable results on both algorithms. In brief, for normal subjects, 1-sample *t* test comparing the difference between SS and SFR was not significantly different to 0 for mean deviation ($P = .6593$) and for pattern standard deviation ($P = .4597$; Figure 3). Patients with glaucoma also showed no

significant difference between algorithms for mean deviation ($P = .8974$) or pattern standard deviation ($P = .1019$). The mean deviation score was slightly higher for glaucoma suspects when using SFR (mean difference 0.26, $P = .0175$) but not for pattern standard deviation ($P = .0508$; Figure 3).

Pointwise differences in sensitivity across each visual field test location, arranged into near-concentric rings from fixation, are shown in Figure 4 for each diagnostic group. Based on the pointwise variability of test locations across the visual field, 6.5%, 11.8%, and 13.4% of all points were outside 2 SDs for normal subjects, glaucoma suspects, and patients with glaucoma, respectively. Differences in probability score were not plotted for clarity because of the interval scale and overlapping data.

One-sample *t* tests were applied for each "ring" to determine if there was a systematic bias in sensitivity difference. In general, there was a bias toward a higher sensitivity result found using SFR compared with SS (Table 2).

We then determined the correlation between the difference in SFR and SS sensitivities and the baseline SS sensitivity, which was used as the reference standard (Figure 5). The correlations were approximately moderate[46] across all diagnostic groups (normal subjects, $r = -0.46$; glaucoma suspects, $r = -0.48$; patients with glaucoma, $r = -0.37$). Linear regression analysis revealed an x-intercept (ie, a sensitivity difference of 0) of 31.1, 30.7, and 28.4 dB for normal subjects, glaucoma suspects, and patients with glaucoma, respectively. Therefore, SFR results in higher sensitivity values when SS sensitivities were lower, and vice versa for higher SS results. Because of the skews induced by sensitivities of low reliability, we also performed separate analyses for when sensitivities <19 dB were censored (Figure 6).[18] Notably, the relationship worsened significantly for patients with glaucoma; though statistically significant, the coefficient of determination was poor at 0.0182. This demonstrated that sensitivities are not predictable for glaucoma patients at levels of <19 dB.

Difference in probability score data was examined as a function of SS probability score (Table 3). A Wilcoxon signed rank test (difference to 0) showed significant differences across all groups except for a probability score of 4 for normal subjects (possibly related to the small sample size). Where SS showed a probability score of 0, all subjects tended to have a significant number of SFR results with a probability score >1. Conversely, when the probability score on SS was 1–4, the SFR score tended to be lower, whereby there was a relatively lower level of statistical significance of visual field defects. This mirrored the relationship identified in the difference in sensitivity, where more significant defects in SS were underestimated when using SFR.

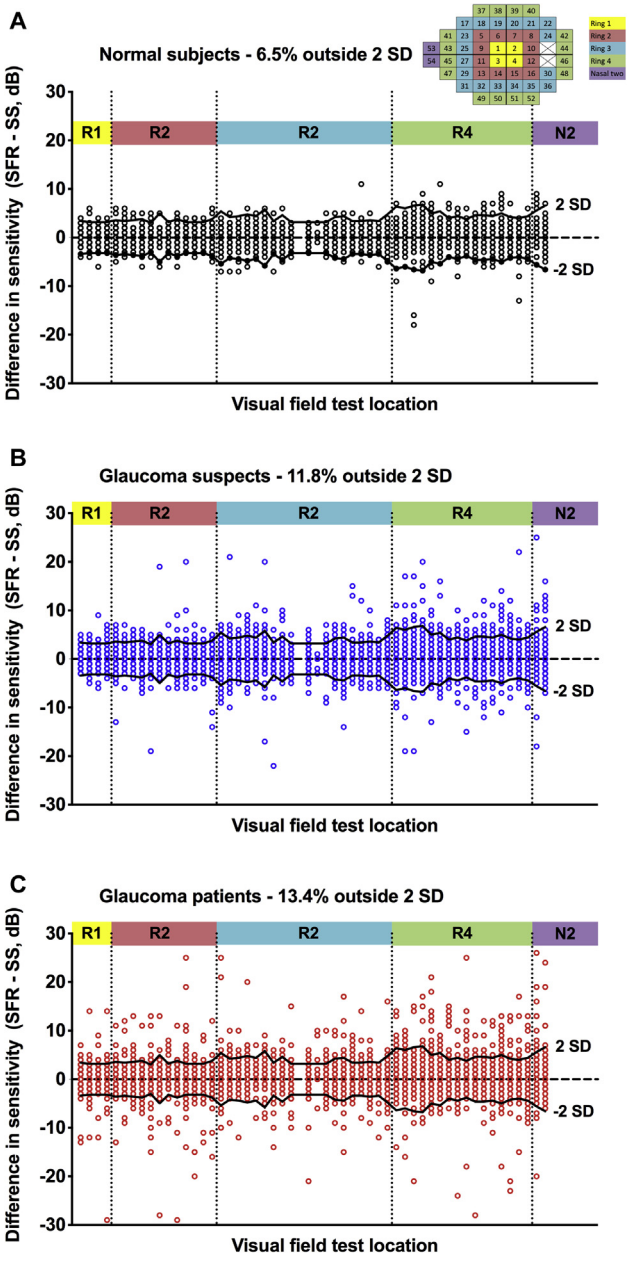

FIGURE 4. Difference in sensitivity between Swedish interactive thresholding algorithm (SITA)-Faster (SFR) and SITA-Standard (SS) in dB, plotted as a function of visual field test location for normal subjects, glaucoma suspects, and patients with glaucoma. Each datum point indicates the result from 1 observer at that location. The horizontal dashed line indicates no difference. The vertical dotted lines separate the approximate concentric "rings" as depicted in the upper right inset. The solid symmetrical lines indicate 2 standard deviations (SDs) at each location across the visual field according to Phu and associates.[53] The proportion of points outside 2 SDs for each group is shown in each figure label. N2 = nasal 2 points; R1 = ring 1; R2 = ring 2; R3 = ring 3; R4 = ring 4.

TABLE 2. One-Sample *t* Test Results for Comparing the Differences Between Swedish Interactive Thresholding Algorithm-Faster and Swedish Interactive Thresholding Algorithm-Standard Sensitivity (dB) and Probability Score Results

	Normal Subjects	Glaucoma Suspects	Patients With Glaucoma
Sensitivity			
Ring 1	0.28 (0.06–0.50)[a]	0.20 (0.03–0.38)[a]	−0.17 (−0.53 to 0.19)
Ring 2	0.25 (0.11–0.39)[b]	0.38 (0.25–0.51)[c]	0.06 (−0.17 to 0.29)
Ring 3	0.19 (0.07–0.31)[d]	0.41 (0.29–0.53)[c]	0.45 (0.29–0.62)[c]
Ring 4	0.41 (0.26–0.55)[c]	0.66 (0.53–0.79)[c]	0.71 (0.51–0.91)[c]
Nasal 2	0.54 (0.06–1.03)[a]	0.92 (0.39–1.46)[b]	1.85 (0.98–2.71)[c]
Average	0.30 (0.22–0.38)[c]	0.50 (0.43–0.58)[c]	0.48 (0.36–0.59)[c]
Probability score			
Ring 1	−0.01 (−0.07 to 0.04)	−0.08 (−0.15 to −0.02)[a]	−0.19 (−0.30 to −0.08)[b]
Ring 2	0.04 (−0.02 to 0.09)	−0.03 (−0.08 to 0.02)	−0.06 (−0.13 to 0.02)
Ring 3	−0.01 (−0.04 to 0.03)	−0.03 (−0.07 to 0.00)	−0.07 (−0.13 to 0.01)[a]
Ring 4	−0.03 (−0.04 to 0.04)	−0.03 (−0.07 to 0.01)	−0.10 (−0.17 to −0.04)[b]
Nasal 2	0.16 (0.04–0.28)[a]	−0.00 (−0.11 to 0.11)	−0.21 (−0.42 to 0.01)[a]
Average	0.01 (−0.01 to 0.03)	−0.03 (−0.06 to 0.01)[d]	−0.09 (−0.13 to −0.06)[c]

Mean differences with 95% confidence intervals of the difference are shown. A positive value indicates that the Swedish interactive thresholding algorithm-Faster result was higher than the Swedish interactive thresholding algorithm-Standard result. Results have been arranged in approximate rings from fixation, as per Figure 3.

[a]$P < .05$, [b]$P < .001$, [c]$P < .0001$, and [d]$P < .01$ indicate the level of significance from the 1-sample *t* test comparing the Swedish interactive thresholding algorithm-Faster and Swedish interactive thresholding algorithm-Standard difference from a mean of 0.

• CLUSTER ANALYSIS OF SENSITIVITY DIFFERENCES BETWEEN ALGORITHMS: To complement the analysis based on approximate distance shown in Figure 4, cluster analysis was performed to examine for pointwise systematic locations exhibiting the same change. The resultant clusters demonstrated no systematic pattern. Supplemental Figure 2 (Supplemental Material available at AJO.com) shows an example of a step before the final complete merge when using a $d' > 1$ criterion. The inability to resolve a pattern using this was caused by the significant interindividual variation.

DISCUSSION

IN THE PRESENT STUDY, WE ADDRESSED 3 KEY QUESTIONS pertaining to the application of SFR in clinical practice. First, we confirmed shorter test times obtained using SFR compared with SS and identified errors that contributed to differences in reliability rates between algorithms. This was then used to address the second aim, which was to model cumulative test times. This analysis illustrated the time saving benefits of SFR, even when lower rates of test reliability were considered. Below we propose alterations to further improve this while maintaining its benefits. Third, we found that SFR generally resulted in higher sensitivity values, and differences were more pronounced in patients with glaucoma and regions with greater visual field loss. However, no systematic difference was found on cluster analysis because of inter- and intraindividual variation.

• MODIFICATIONS TO SFR AND LOW TEST RELIABILITY: We identified 2 modifications used to reduce test time in SFR that may have contributed to lower test reliability relative to SS.[34] Starting with an age-corrected sensitivity value (rather than 25 dB) at the 4 seeding points introduces significant uncertainty.[21,22,48] Consequently, this may alter response criterion, such that a more intense stimulus is required before the subject indicates a response. This is further complicated by having only 1 reversal for the staircase, which finally manifests as higher seeding point errors using SFR because there is no opportunity to reorient to the actual sensitivity.[49]

As described by Heijl and associates,[34] a higher false positive rate is expected in SITA-Fast and SFR compared with SS. This has been attributed to the more difficult testing condition of SITA-Fast and SFR where presentation intensities are near to threshold at initial presentation,[13] which is thought to lead to a more liberal response criterion. In comparison, SS typically uses more subthreshold intensities as starting points, creating a positive bias (higher sensitivity) in responses. In addition, SFR may also have a greater rate of false positive errors compared with SS because of the shorter timing between stimulus presentations.[34] Reaction time varies considerably across individuals, test location, and threshold level,[50,51] and this may result in greater false positive rates.

• TRADING OFF TEST RELIABILITY FOR TEST TIME: At present, based on our results, there is a high level of low test reliability found using SFR compared with SS within a subset of patients, even after retesting. However, does

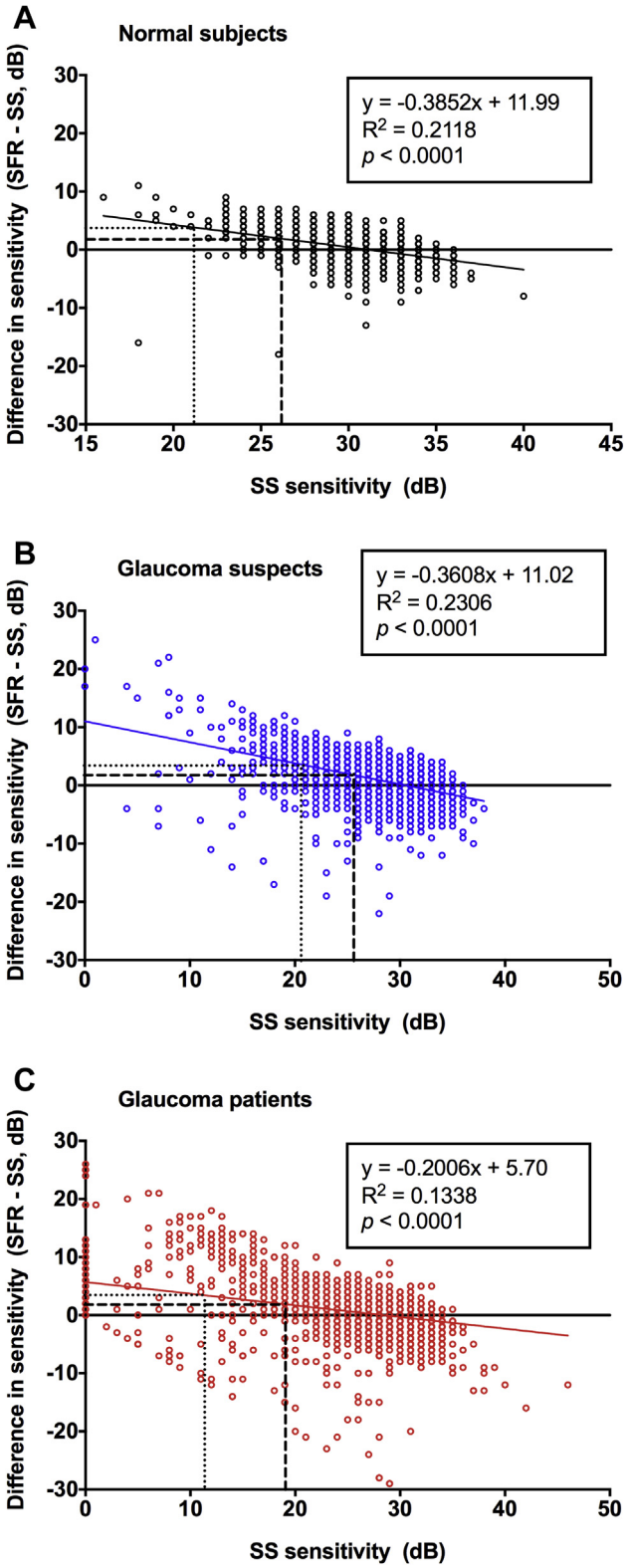

FIGURE 5. Difference in sensitivity between Swedish interactive thresholding algorithm (SITA)-Faster (SFR) and SITA-Standard (SS) as a function of SS sensitivity value for each diagnostic group. Each datum point represents the result from 1 observer at that particular SS sensitivity value. The horizontal solid line indicates no difference between algorithms. The results of the linear regression analysis are also shown. The dashed and dotted lines indicate the difference in sensitivity when a SS difference of 5 dB and 10 dB from the *x*(*y* = 0) value, respectively.

this mean that these cancel out the benefits obtained with shorter test duration, such as the ability to conduct more tests?

Our models suggest that the overall smaller increase in test duration as a product of low test reliability when using SFR may still outweigh the increase in frequency of unreliable results. This is a product of unreliable results obtained with SS having a greater percentage increase in test duration. This indicates that increasing the frequency of testing with SFR may still provide practical benefits in some cases.

Our results suggest that a number of the modifications made to shorten the test duration for SFR may also require alterations to patient instruction to reduce the rate of low test reliability, with an emphasis on correct measurement of the seeding points, and could enable one to pay specific attention to the onset of the first 4 stimuli. We also postulate that one simple modification to seeding point determination in SFR may assist in retaining the benefits of short test duration while improving the overall reliability and confidence in clinical interpretation. At this point, this suggestion remains speculative, and further experimental testing using a different staircase procedure at the seeding points would be informative. Furthermore, these conclusions were based primarily upon a cohort of patients with predominantly early to moderate glaucoma; a larger study specifically in later stage disease would be useful.

Another finding from our study was the difference in reliable results, whereby SFR performed first led to a greater rate of unreliability overall. As a result, the analyses could not be performed separately because of different distributions of reliable and unreliable subjects. This result, however, raises the question of whether the learning effect is also dependent on algorithm[52]: had the modifications leading to SFR created a new learning curve for perimetric testing? Importantly, does this also mean that alternating between test paradigms may not be recommended because of different response behaviors? A different study design would be required to address these interesting clinical questions.

• **CORRELATIONS BETWEEN SENSITIVITY MEASUREMENTS USING SFR AND SS:** Given the interest in using SFR in lieu of SS in clinical practice, we aimed to determine if there were systematic differences between algorithm results. The overall differences in sensitivity between algorithms were low, and at face value, correlations between algorithms suggest that they may be interchangeable, or at least some correction factor could be applied to render the results comparable.

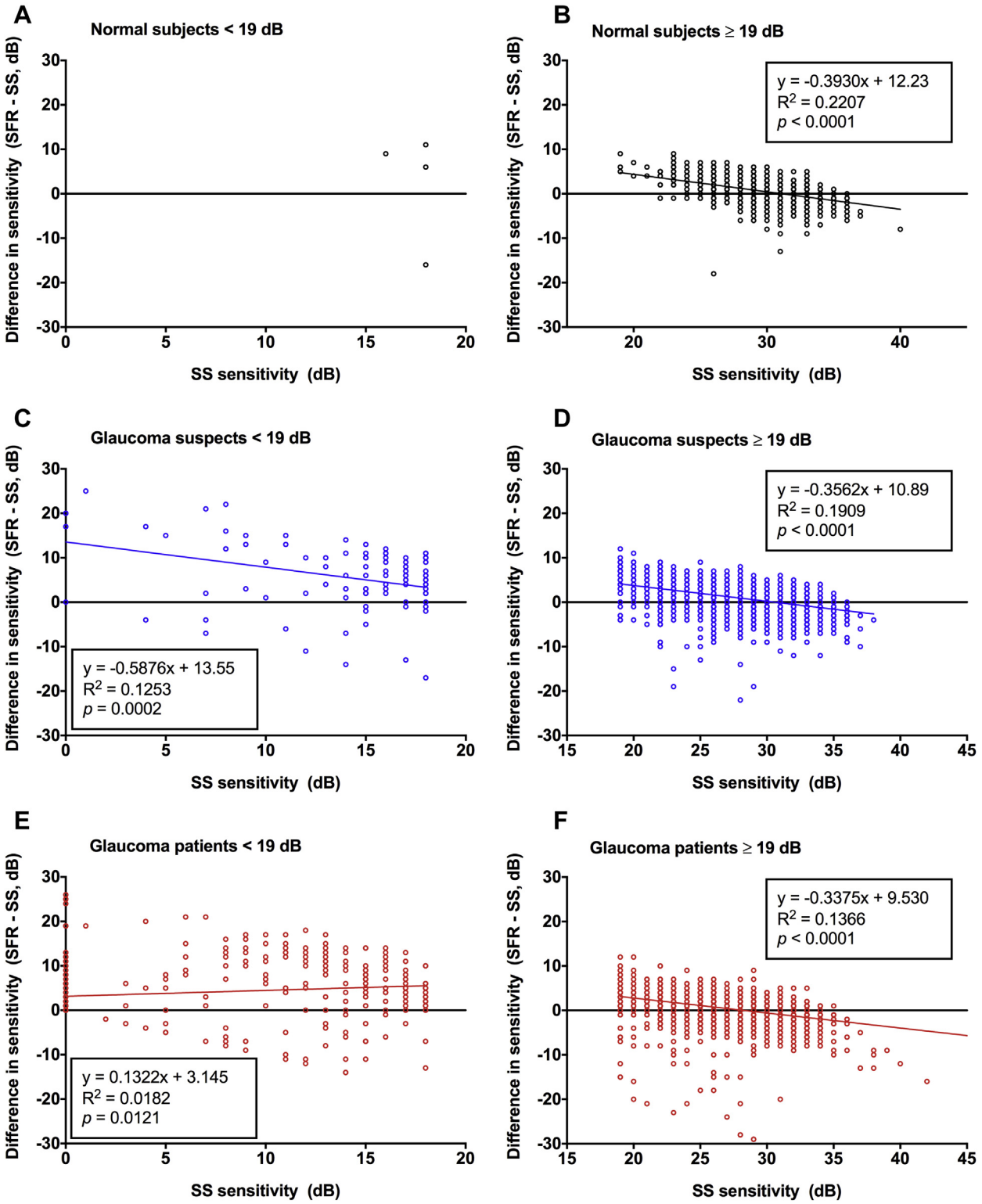

FIGURE 6. Difference in sensitivity between Swedish interactive thresholding algorithm (SITA)-Faster (SFR) and SITA-Standard (SS) as a function of SS sensitivity value for each diagnostic group separated into 2 SS sensitivity groups: < 19 dB (left) and ≥19 dB (right). Each datum point represents the result from 1 observer at that particular SS sensitivity value. The horizontal solid line indicates no difference between algorithms. The results of the linear regression analysis are also shown, except for normal subjects < 19 dB because of the few data points available.

TABLE 3. Proportion of Different Levels of Difference Between Swedish Interactive Thresholding Algorithm-Faster and Swedish Interactive Thresholding Algorithm-Standard (Ordinal Units) Occurring for Each Swedish Interactive Thresholding Algorithm-Standard Probability Score for Each Diagnostic Group

	Difference Between SITA-Faster and SITA-Standard									
SITA-Standard Probability Score	−4	−3	−2	−1	0	1	2	3	4	*P* Value
Normal subjects										
0 (n = 2626)	0.0	0.0	0.0	0.0	94.2	3.5	1.1	0.8	0.3	<.0001
1 (n = 106)	0.0	0.0	0.0	89.6	5.7	1.9	1.9	0.9	0.0	<.0001
2 (n = 48)	0.0	0.0	83.3	8.3	6.3	0	2.1	0.0	0.0	<.0001
3 (n = 23)	0.0	65.2	17.4	8.7	4.3	4.3	0.0	0.0	0.0	<.0001
4 (n = 5)	60.0	0.0	0.0	20.0	20.0	0.0	0.0	0.0	0.0	.1250
Glaucoma suspects										
0 (n = 5396)	0.0	0.0	0.0	0.0	89.6	6.2	2.3	1.4	0.5	<.0001
1 (n = 404)	0.0	0.0	0.0	76.5	13.4	5.9	2.7	1.5	0.0	<.0001
2 (n = 167)	0.0	0.0	64.1	15.6	9.0	7.8	3.6	0.0	0.0	<.0001
3 (n = 162)	0.0	66.7	11.7	7.4	7.4	6.8	0.0	0.0	0.0	<.0001
4 (n = 111)	39.6	10.8	12.6	9.9	27.0	0.0	0.0	0.0	0.0	<.0001
Patients with glaucoma										
0 (n = 2881)	0.0	0.0	0.0	0.0	85.0	7.5	3.4	2.9	1.2	<.0001
1 (n = 321)	0.0	0.0	0.0	68.8	13.4	5.9	4.7	7.2	0.0	<.0001
2 (n = 177)	0.0	0.0	63.3	11.9	8.5	10.2	6.2	0.0	0.0	<.0001
3 (n = 151)	0.0	55.6	11.3	9.3	11.3	12.6	0.0	0.0	0.0	<.0001
4 (n = 422)	19.4	9.2	5.9	9.7	55.7	0.0	0.0	0.0	0.0	<.0001

P values indicate the result of a Wilcoxon signed rank test examining for a difference to 0.
SITA = Swedish interactive thresholding algorithm.

Although areas in some patients seem to exhibit a greater magnitude of difference between algorithms, the interindividual variability in sensitivity manifested as low separability scores in the cluster analysis for both normal subjects and patients with glaucoma. The absolute sensitivity difference between classes identified on cluster analysis was small and was lower than the variance. One reason for this may be that the probability density functions are intrinsically different between SFR and SS.[13] Another reason that is supported by our data is that the overall sensitivity differences were low, such that larger disparities may be masked by the variability of the grouped data. Most differences were still within the approximate test–retest variability using the SS algorithm.[53] The probability score results also appear to support this as the majority of points demonstrated no statistically significant change in probability score. Again, this suggested that large differences between algorithms were the exception.

• **ARE THE SITA ALGORITHMS INTERCHANGEABLE?:** In isolation, the results of the cluster analysis suggest that there is a uniform difference between SS and SFR in sensitivity across the visual field. However, knowing the ground truth sensitivity differences, the merging of all clusters occurred as the SD values at all locations exceeded that of the average sensitivity differences. Therefore, the lack of statistically separable clusters found on cluster analysis did not reflect a systematic and uniform difference across the visual field. Instead, the high variance at each test location meant that the differences between SS and SFR are less predictable.

Higher variance at locations with greater defect depth has been demonstrated by a number of other studies,[54,55] and there have also been suggestions for the censoring of perimetric data below certain sensitivity levels (15–19 dB).[31,56] When we censored our data below 19 dB, there was no difference in the slope values or goodness of fit, suggesting that moderate strength of correlation was not related to high variance at low sensitivities but rather was likely driven by algorithm differences.

Importantly, we confirm the findings of Heijl and associates[34] that the difference in sensitivity between algorithms is dependent upon the underlying SS sensitivity (the reference value in the present study). The overestimation of sensitivity by SFR also manifested as an underestimation of the probability score on the pattern deviation map (the significance of the defect). This has implications for determination of patterns of loss using grading systems, such as the Glaucoma Staging Score.[47] The effect of these differences in measurement of longitudinal sensitivity change requires additional investigation. Based on these results, we suggest that the SITA algorithms are not

necessarily interchangeable unless a postprocessing step is added to account for these differences. In addition, our results further contribute to the literature by presenting a different relationship between SS and SFR at different sensitivity brackets (<19 dB and ≥19 dB). The significant variability in relationship at unreliable sensitivity values, in combination with the discordance in probability scores, further suggests that algorithms are not interchangeable, especially in more severe field loss. This will likely have ramifications in long-term monitoring, which requires further investigation.

• **LIMITATIONS:** A limitation of this study was that we only measured visual fields once. Although all patients enrolled in this study had previously undergone perimetric testing, there was no method to measure intraobserver retest variability using each algorithm to confirm the results of Heijl and associates.[34] Further validation of test–retest variability is important crucial to understanding the utility of SFR in longitudinal settings. We also did not control the level of perimetric experience across each subject, although we attempted to reduce these potential effects on the comparisons of algorithms by randomizing test order and allowing 1 test restart. Our model is also conservative in terms of the number of restarts of visual fields and did not account for a potential need for bringing patients back on another clinic visit, which can add considerable time. Under those circumstances, the time-saving advantage may be reduced.

Although the cross-sectional consecutive study design has an advantage of reducing spectrum or selection bias, the overall representation of moderate or worse glaucoma cases was relatively limited. Our cohort, although much larger than the last report on SFR,[34] also consisted of patients within a single referral-only clinic—comprised predominantly of white and East Asian patients—and therefore may not be directly extrapolated to other clinics. The decay in relationship specifically for patients with glaucoma with advanced field loss reinforces the poor reliability of perimetric testing within the sensitivity range of <19 dB when SFR is performed[18,56]; however, another study in a clinic with more advanced cases would be useful to confirm these findings.

CONCLUSIONS

OUR RESULTS INDICATE THAT SFR PROVIDES BENEFITS IN test time, even with higher rates of low test reliability compared with SS. Common sources of error when using SFR could be addressed through careful patient instruction and/or future subtle modifications to the algorithm to improve the measurement of seeding points. Differences in sensitivity and probability scores between SFR and SS were small, with no clear evidence of systematic differences. SFR could therefore be considered if a short test duration or more data are desired. Its use under conditions with greater visual field loss should be judicious because of potentially larger differences in output sensitivity or probability score results: the algorithms are specifically not interchangeable once more severe vision loss occurs.

ALL AUTHORS HAVE COMPLETED AND SUBMITTED THE ICMJE FORM FOR DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST. Funding/Support: Publication of this article was supported in part by National Health and Medical Research Council grant 1033224 (to M.K.). Guide Dogs New South Wales/Australia Capital Territory are partners on the grant. The funding organization had no role in the design or conduct of this research. Financial Disclosures: No conflicting relationship exists for all authors. All authors attest that they meet the current ICMJE criteria for authorship. We thank the technicians and clinical staff at the Centre for Eye Health, University of New South Wales Sydney, Australia, for their technical assistance.

REFERENCES

1. 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.
2. Jampel HD, Singh K, Lin SC, et al. Assessment of visual function in glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology* 2011;118:986–1002.
3. Choudhari NS, Pathak-Ray V, Kaushik S, Vyas P, George R. Prevalent practice patterns in glaucoma: poll of Indian ophthalmologists at a national conference. *Indian J Ophthalmol* 2016;64:715–721.
4. Gaskin BJ, Carroll SC, Gamble G, Goldberg I, Danesh-Meyer HV. Glaucoma management trends in Australia and New Zealand. *Clin Exp Ophthalmol* 2006;34:208–212.
5. Gordon-Bennett PS, Ioannidis AS, Papageorgiou K, Andreou PS. A survey of investigations used for the management of glaucoma in hospital service in the United Kingdom. *Eye* 2008;22:1410–1418.
6. Linden C, Bengtsson B, Alm A, Calissendorff B, Eckerlund I, Heijl A. Glaucoma management in Sweden–results from a nationwide survey. *Acta Ophthalmol* 2013;91:20–24.
7. National Health and Medical Research Council. Guidelines for the screening, prognosis, diagnosis, management and prevention of glaucoma. Commonwealth of Australia; 2010. Available at: https://nhmrc.gov.au/about-us/publications/guidelines-screening-prognosis-diagnosis-management-and-prevention-glaucoma. Accessed October 9, 2018.
8. McKendrick AM. Recent developments in perimetry: test stimuli and procedures. *Clin Exp Optom* 2005;88:73–80.

9. Shirato S, Inoue R, Fukushima K, Suzuki Y. Clinical evaluation of SITA: a new family of perimetric testing strategies. *Graefes Arch Clin Exp Ophthalmol* 1999;237:29–34.
10. Bengtsson B, Heijl A. Evaluation of a new perimetric threshold strategy, SITA, in patients with manifest and suspect glaucoma. *Acta Ophthalmol Scand* 1998;76:268–272.
11. Bengtsson B, Heijl A. Inter-subject variability and normal limits of the SITA Standard, SITA Fast, and the Humphrey Full Threshold computerized perimetry strategies, SITA STATPAC. *Acta Ophthalmol Scand* 1999;77:125–129.
12. Bengtsson B, Heijl A. Comparing significance and magnitude of glaucomatous visual field defects using the SITA and Full Threshold strategies. *Acta Ophthalmol Scand* 1999;77:143–146.
13. Bengtsson B, Heijl A. SITA Fast, a new rapid perimetric threshold test. Description of methods and evaluation in patients with manifest and suspect glaucoma. *Acta Ophthalmol Scand* 1998;76:431–437.
14. Heijl A, Lindgren G, Olsson J. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol* 1987;105:1544–1549.
15. 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.
16. Chauhan BC, Tompkins JD, LeBlanc RP, McCormick TA. Characteristics of frequency-of-seeing curves in normal subjects, patients with suspected glaucoma, and patients with glaucoma. *Invest Ophthalmol Vis Sci* 1993;34:3534–3540.
17. Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci* 2000;41:417–421.
18. 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.
19. 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.
20. Stewart WC, Hunt HH. Threshold variation in automated perimetry. *Surv Ophthalmol* 1993;37:353–361.
21. Phu J, Kalloniatis M, Wang H, Khuu SK. Differences in static and kinetic perimetry results are eliminated in retinal disease when psychophysical procedures are equated. *Transl Vis Sci Technol* 2018;7:22.
22. 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;7:8.
23. Crabb DP, Garway-Heath DF. Intervals between visual field tests when monitoring the glaucomatous patient: wait-and-see approach. *Invest Ophthalmol Vis Sci* 2012;53:2770–2776.
24. 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.
25. Ledolter J, Kardon RH. Does testing more frequently shorten the time to detect disease progression? *Transl Vis Sci Technol* 2017;6:1.
26. Anderson AJ. Estimating the true distribution of visual field progression rates in glaucoma. *Invest Ophthalmol Vis Sci* 2015;56:1603–1608.
27. Heijl A, Buchholz P, Norrgren G, Bengtsson B. Rates of visual field progression in clinical glaucoma care. *Acta Ophthalmol* 2013;91:406–412.
28. Boodhna T, Crabb DP. More frequent, more costly? Health economic modelling aspects of monitoring glaucoma patients in England. *BMC Health Serv Res* 2016;16:611.
29. Turpin A, Jankovic D, McKendrick AM. Retesting visual fields: utilizing prior information to decrease test-retest variability in glaucoma. *Invest Ophthalmol Vis Sci* 2007;48:1627–1634.
30. Chong LX, McKendrick AM, Ganeshrao SB, Turpin A. Customized, automated stimulus location choice for assessment of visual field defects. *Invest Ophthalmol Vis Sci* 2014;55:3265–3274.
31. Pathak M, Demirel S, Gardiner SK. Reducing variability of perimetric global indices from eyes with progressive glaucoma by censoring unreliable sensitivity data. *Transl Vis Sci Technol* 2017;6:11.
32. Denniss J, McKendrick AM, Turpin A. Towards patient-tailored perimetry: automated perimetry can be improved by seeding procedures with patient-specific structural information. *Transl Vis Sci Technol* 2013;2:3.
33. Ganeshrao SB, McKendrick AM, Denniss J, Turpin A. A perimetric test procedure that uses structural information. *Optom Vis Sci* 2015;92:70–82.
34. Heijl A, Patella VM, Chong LX, et al. A new SITA perimetric threshold testing algorithm; construction and a multicenter clinical study. *Am J Ophthalmol* 2018.
35. Casson RJ, Chidlow G, Wood JP, Crowston JG, Goldberg I. Definition of glaucoma: clinical and experimental concepts. *Clin Exp Ophthalmol* 2012;40:341–349.
36. Prum BE Jr, Rosenberg LF, Gedde SJ, et al. Primary open-angle glaucoma preferred practice pattern((R)) guidelines. *Ophthalmology* 2016;123:P41–P111.
37. Heijl A, Krakau CE. An automatic perimeter for glaucoma visual field screening and control. Construction and clinical cases. *Albrecht Von Graefes Arch Klin Exp Ophthalmol* 1975;197:13–23.
38. Olsson J, Bengtsson B, Heijl A, Rootzen H. An improved method to estimate frequency of false positive answers in computerized perimetry. *Acta Ophthalmol Scand* 1997;75:181–183.
39. Ishiyama Y, Murata H, Asaoka R. The usefulness of gaze tracking as an index of visual field reliability in glaucoma patients. *Invest Ophthalmol Vis Sci* 2015;56:6233–6236.
40. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Invest Ophthalmol Vis Sci* 2000;41:2201–2204.
41. Yohannan J, Wang J, Brown J, et al. Evidence-based criteria for assessment of visual field reliability. *Ophthalmology* 2017;124:1612–1620.
42. Phu J, Wang H, Miao S, Zhou L, Khuu SK, Kalloniatis M. Neutralizing peripheral refraction eliminates refractive scotomata in tilted disc syndrome. *Optom Vis Sci* 2018;95:959–970.
43. Phu J, Khuu SK, Bui BV, Kalloniatis M. Application of pattern recognition analysis to optimize hemifield asymmetry patterns for early detection of glaucoma. *Transl Vis Sci Technol* 2018;7:3.

44. 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.
45. Cai S, Elze T, Bex PJ, Wiggs JL, Pasquale LR, Shen LQ. Clinical correlates of computationally derived visual field defect archetypes in patients from a glaucoma clinic. *Curr Eye Res* 2017;42:568–574.
46. Hinkle DE, Weirsma W, Jurs SG. Applied Statistics for the Behavioral Sciences. 5th ed. Boston, MA: Houghton Mifflin; 2003.
47. 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.
48. Phu J, Kalloniatis M, Khuu SK. The effect of attentional cueing and spatial uncertainty in visual field testing. *PLoS One* 2016;11:e0150922.
49. Garcia-Perez MA. Optimal setups for forced-choice staircases with fixed step sizes. *Spat Vis* 2000;13:431–448.
50. McKendrick AM, Denniss J, Turpin A. Response times across the visual field: empirical observations and application to threshold determination. *Vision Res* 2014; 101:1–10.
51. Wall M, Kutzko KE, Chauhan BC. The relationship of visual threshold and reaction time to visual field eccentricity with conventional automated perimetry. *Vision Res* 2002;42:781–787.
52. Capris P, Autuori S, Capris E, Papadia M. Evaluation of threshold estimation and learning effect of two perimetric strategies, SITA Fast and CLIP, in damaged visual fields. *Eur J Ophthalmol* 2008;18:182–190.
53. 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;7:1.
54. Wall M, Woodward KR, Doyle CK, Artes PH. Repeatability of automated perimetry: a comparison between standard automated perimetry with stimulus size III and V, matrix, and motion perimetry. *Invest Ophthalmol Vis Sci* 2009;50:974–979.
55. Russell RA, Crabb DP, Malik R, Garway-Heath DF. The relationship between variability and sensitivity in large-scale longitudinal visual field data. *Invest Ophthalmol Vis Sci* 2012;53: 5985–5990.
56. 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.