False Positive Responses in Standard Automated Perimetry

ANDERS HEIJL, VINCENT MICHAEL PATELLA, JOHN G. FLANAGAN, AIKO IWASE, CHRISTOPHER K. LEUNG, ANJA TUULONEN, GARY C. LEE, THOMAS CALLAN, AND BOEL BENGTSSON

• PURPOSE: To analyze the relationship between rates of false positive (FP) responses and standard automated perimetry results.
• DESIGN: Prospective multicenter cross-sectional study.
• METHODS: One hundred twenty-six patients with manifest or suspect glaucoma were tested with Swedish Interactive Thresholding Algorithm (SITA) Standard, SITA Fast, and SITA Faster at each of 2 visits. We calculated intervisit differences in mean deviation (MD), visual field index (VFI), and number of statistically significant test points as a function of FP rates and also as a function of general height (GH).
• RESULTS: Increasing FP values were associated with higher MD values for all 3 algorithms, but the effects were small, 0.3 dB to 0.6 dB, for an increase of 10 percentage points of FP rate, and for VFI even smaller (0.6%-1.4%). Only small parts of intervisit differences were explained by FP (r^2 values 0.00-0.11). The effects of FP were larger in severe glaucoma, with MD increases of 1.1 dB to 2.0 dB per 10 percentage points of FP, and r^2 values ranging from 0.04 to 0.33. The numbers of significantly depressed total deviation points were affected only slightly, and pattern deviation probability maps were generally unaffected. GH was much more strongly related to perimetric outcomes than FP.
• CONCLUSIONS: Across 3 different standard automated perimetry thresholding algorithms, FP rates showed only weak associations with visual field test results, except in severe glaucoma. Current recommendations regarding acceptable FP ranges may require revision. GH or other analyses may be better suited than FP rates for identifying unreliable results in patients who frequently press the response button without having perceived stimuli. (Am J Ophthalmol 2021;233: 180–188.

WITH THE INTRODUCTION OF COMPUTERIZED perimeters in the 1970s, 3 so-called "reliability parameters" were implemented with the hope of helping users judge whether test results were reliable and useful. These parameters were fixation losses (FLs), false negative (FN) responses, and false positive (FP) responses.[1-3] FL responses are obtained using a method described in 1974 in which test stimuli are presented at the expected location of the physiologic blind spot of the tested eye.[1] The method was originally designed to give a qualitative idea about fixation in an early computerized perimeter, where the operator could not see the tested eye. The method has been widely used in many or most automated perimeters, but has well-known shortcomings, especially in eyes where the blind spot is not situated in the assumed location. Today, various methods for gaze tracking can be considered superior to the blind spot technique, and at least one new testing algorithm relies by default upon gaze tracking and not FL estimates based on the blind spot method.[4]

FN responses were intended to be an index of patient vigilance. FN rates usually are measured by displaying stimuli that should be easily visible, based upon threshold sensitivity measurements made at the chosen locations earlier in the test. However, in the 1980s it was reported that the percentage of FN answers depended more on the level of visual field damage than on patient vigilance.[5] In Bengtsson and Heijl,[6] this shortcoming was clearly demonstrated by testing both eyes of patients having unilateral glaucoma. It is now recognized that test results should not be discarded solely on the basis of elevated FN response rates.

While FL and FN rates have been considered decreasingly important over time, this has not been the case thus far for FP response rates. FP rate estimates are meant to identify "trigger-happy" testing behavior, ie, examinations in which patients too frequently pressed the perimeter's response button without having perceived a stimulus. Classic "trigger-happy" fields, with very high-threshold sensitivity values and white patches in the grayscale maps, often have high percentages of FP answers, but this is not always the case (Figure 1). FP rates were originally esti-

AJO.com Supplemental Material available at AJO.com.
Accepted for publication June 25, 2021.

From Ophthalmology Research Unit, Department of Clinical Sciences Malmö, Lund University (A.H. and B.B.); Department of Ophthalmology, Skåne University Hospital, Malmö, Sweden (A.H.); Department of Ophthalmology (V.M.P.), University of Iowa, Iowa City, Iowa, USA; School of Optometry and Vision Science Program (J.G.F.), University of California, Berkeley, Berkeley, USA; Tajimi Iwase Eye Clinic (A.I.), Tajimi, Japan; Department of Ophthalmology and Visual Sciences (C.K.L.), Chinese University of Hong Kong, Hong Kong, China; Tays Eye Centre (A.T.), Tampere University Hospital, Tampere, Finland; Carl Zeiss Meditec, Inc. (G.C.L., T.C.), Dublin, California, USA

Inquiries to Anders Heijl, Department of Ophthalmology, Skåne University Hospital, Jan Waldenströms Gata 24, SE-20502, Malmö, Sweden.; e-mail: anders.heijl@med.lu.se

0002-9394/$36.00
https://doi.org/10.1016/j.ajo.2021.06.026

mated using catch trials in which no stimulus was presented, noting if the patient erroneously pressed the response button.[7] More recently, Swedish Interactive Thresholding Algorithm (SITA) testing programs have incorporated a different method of estimating FP rates that is based upon detection of patient responses during times when it is impossible or unlikely that a stimulus was seen.[8]

The reason that high FP response rates are of interest is that they are expected to be associated with artifactually elevated threshold sensitivity values, with higher FP rates being associated with higher mean deviation (MD) values, both in perimetry-naïve normal subjects[9] and in patients with glaucoma.[10] In the first of these studies, the analysis was based on just a single visual field test per normal subject, and in the latter study the results were based on differences between predicted and observed MD values in eyes with suspect or manifest glaucoma. However, FP rates have also been reported to have almost no correlation to measurement variability in a cohort of patients with suspect or manifest glaucoma who underwent threshold visual field testing twice within approximately 1 week.[11]

Recommended limits for clinically "acceptable" FP rates have evolved over time. In the 1980s, we used an arbitrary limit of 33%, which simply was the limit we had chosen

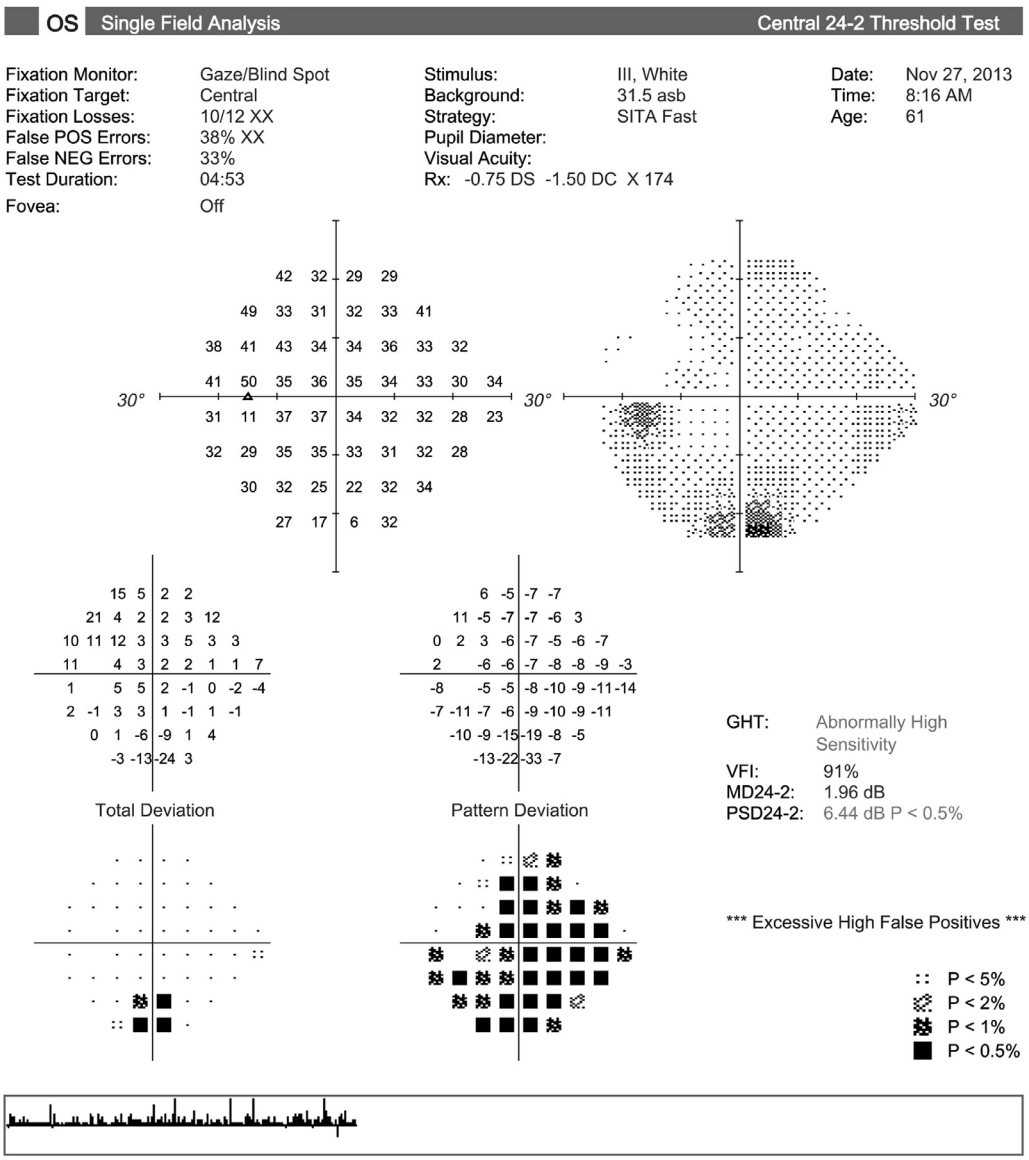

FIGURE 1. False positive rates in "trigger-happy fields": these 2 fields both show typical features of trigger-happy fields, including abnormally high threshold sensitivity values, "white scotomas," "reversed cataract pattern," many more significant test point locations in pattern deviation probability maps than in total deviation maps, and GHT classifications of "abnormally high sensitivity." One field (A) shows a high rate of false positive responses (35%) while the other (B) has a false positive rate of 0%.

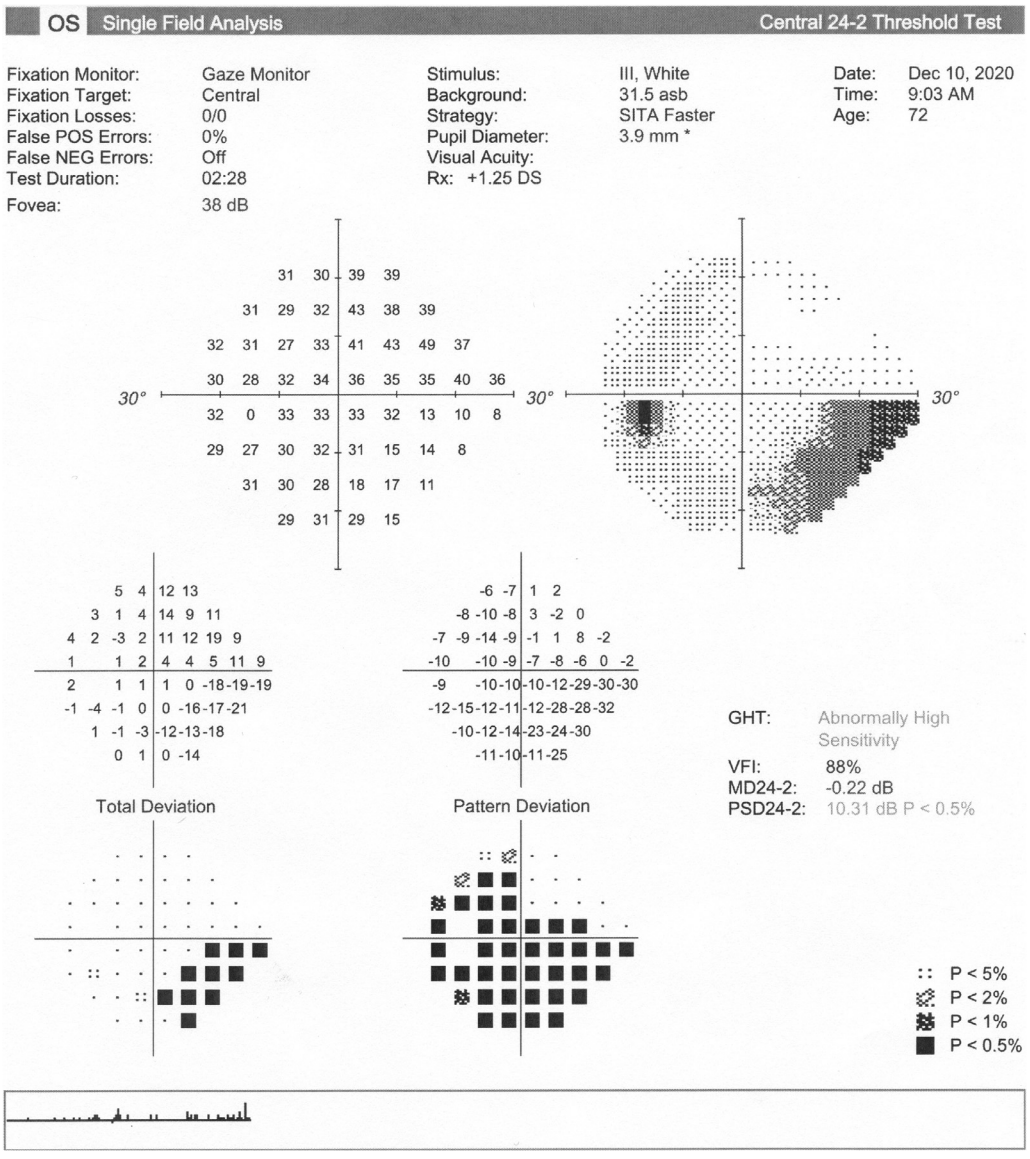

FIGURE 1. Continued

as an exclusion criterion for the visual field tests used to define normative significance limits for the first Humphrey Statpac interpretation package.[12] Later, we suggested that FP rates >15% might indicate unreliable test results, a recommendation that was based upon the distribution of FP levels seen in a sample of field test results. Thus, FP rates >15% were flagged because they were uncommon, not because tests with higher FP rates were unreliable.[13]

While in perimetry FP responses have traditionally been regarded as errors, signal detection theory provides a different perspective.[14] In signal detection theory, FP responses are merely a reflection of the subject's response criterion.

Recently, while developing the SITA Faster (SFR) test strategy, we noticed that the percentage of FP answers was higher with the new program than with SITA Fast (SF),[4] and this has been subsequently reported by other investigators.[15] It has been known for >20 years that FP rate estimates are typically slightly higher with SF than with SITA Standard (SS). Despite the higher FP rates with SFR, the results of a multicenter clinical trial showed almost identical SFR and SF threshold test results.[4] We realized that further analysis of our multicenter SFR study data might provide an opportunity to study the relationship between FP answers and perimetric test results in greater detail. The distinctive advantage of using this recent study material was that all patients had been tested twice within such a short period of time, <2 weeks, that it was reasonable to postulate that no significant visual field progression would have occurred between the 2 tests. A second advantage was that we could simultaneously evaluate FP effects in all 3 SITA

testing algorithms. Therefore, we hoped to determine the extent to which differences in FP measurements between the first and second tests were associated with observed differences in measured threshold sensitivity and associated metrics.

The aim of the current investigation was to analyze our recent multicenter data set, focusing on the relationship between FP rates and perimetric test results in each of 3 different testing strategies.

METHODS

• **SETTING:** This prospective multicenter study was conducted at 5 centers located in 5 different countries in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Gifu Prefecture Medical Association, the Ethics Committee of Tampere University Hospital, the Committee for Protection of Human Subjects of the University of California Berkeley, and the Hong Kong Hospital Authority Kowloon Central Research Ethics Committee. The study was also submitted to the Regional Ethics Review Board in Lund, Sweden. The Lund Board concluded that the study did not need their approval but that they saw no ethical issues.

• **STUDY POPULATION:** The acquisition of study data has been previously described.[4] The study included 126 patients with manifest or suspect glaucoma. No stages of glaucomatous visual field loss were excluded.

• **OBSERVATION PROCEDURE:** All participants underwent Humphrey 24-2 visual field testing in a single study eye using 3 different threshold testing strategies (SFR, SF, and SS) in randomized order. All perimetric testing was repeated at a second visit, between 1 day and 2 weeks later, with testing order reversed. At each study site, participants underwent all testing on the same Humphrey 860 perimeter (Carl Zeiss Meditec, Dublin, California, USA).

If, during testing, the perimetrist observed patient gaze instability or results consistent with false responses, patient misunderstanding, or inattentiveness, the perimetrist was allowed to stop the test, reinstruct the patient, and restart the test from the beginning, thus discarding the interrupted test. However, once a test had been completed, it could not be deleted, and it was included in all statistical analyses.

• **MAIN OUTCOME MEASURES:** From each visual field test we tabulated the percentage of FP responses, visual field index (VFI), and MD values, and the number of significantly depressed test points at the 1% and 0.5% significance levels in the total deviation (TD) and pattern deviation (PD) probability maps.

• **STATISTICAL ANALYSIS:** First, we registered FP rates and MD and VFI values for the 3 algorithms. For each tested eye and each test strategy, we then calculated differences between visit 1 and visit 2 FP rates, as well as intertest differences in VFI, MD, and the number of significantly depressed test points. We then performed linear regression analyses with intrasubject FP differences as the explanatory variable and intrasubject differences in VFI, MD, and number of significantly depressed test points as the dependent variables. We also calculated intertest differences in general height (GH).[16,17] GH is the difference between the numerical TD values and the PD values in the Statpac program of the Humphrey perimeter. We then performed the same regression analyses with GH differences, instead of FP differences, as the explanatory variable.

We also performed regression analyses with FP differences as the explanatory variable and differences in MD, VFI, and in numbers of significantly depressed points with study eyes divided into 3 groups with early, moderate, or severe visual field loss using the MD values of the staging systems of Hoddap and associates[18] and Mills and associates.[19] The MD stage for each eye was defined as the average of the visit 1 and visit 2 MD values, for each test algorithm. Assumptions for linear regression were tested by residual analysis between differences in FP vs differences in MD and VFI. Histograms of residuals were produced, as were scatterplots of standardized residuals over standardized predicted values.

RESULTS

We analyzed test results from 125 patients, including 64 women (51%) and 61 men (49%). The mean age was 67 years (range 26-82 years). Results from 1 subject were excluded because testing of this patient had been interrupted because of the observation of large eye movements. The patient was reinstructed and a new test was started, but fixation stability was still considered unacceptable.

The 3 test strategies showed significantly different FP rates, while MD and VFI values were very similar[4] (Table 1). Intervisit differences shown in Table 1 were all distributed normally.

For each of the 3 strategies, intervisit differences in FP explained only a small part of the intervisit differences in MD and VFI, despite reaching statistical significance in half of the analyses (Table 2). Statistical significance may have been reached simply because of the relatively large number of observations. The coefficients of determination—r^2, the variability in the dependent variable that is explained by the explanatory variable—were small for all strategies for FP vs MD and even smaller for FP vs VFI. Higher FP rates were associated with greater increases in MD values, as expected, but the effects were small (0.4-0.5 dB), depending upon testing strategy, for an increase of 10 per-

TABLE 1. Descriptive Analysis of Mean, Median, Minimum, and Maximum Values for Parameters With Skewed Distributions, and Means and Standard Deviations for Variables That Were Normally Distributed

	Visit 1 Mean; Median (Minimum, Maximum), Skewed Distribution	Visit 2 Mean; Median (Minimum, Maximum), Skewed Distribution	Intrapatient Difference (Visit 1 – Visit 2), Mean (SD), All Gaussian
FP (%) SS	2.8; 2 (0, 28)	2.8; 2 (0, 13)	0.0 (4.1)
FP (%) SF	3.3; 2 (0, 41)	3.65; 2 (0, 32)	−0.4 (5.5)
FP (%) SFR	4.9; 0 (0, 39)	5.0; 3 (0, 43)	−0.1 (9.5)
MD (dB) SS	−8.5; −6.0 (−28.3, 0.56)	−8.5; −6,4 (−28.7, 0.58)	−0.1 (1.3)
MD (dB) SF	−8.6; −6.2 (−28.7, 1.33)	−8.4; −6.1 (−28.9, 0.8)	−0.2 (1.6)
MD (dB) SFR	−8.4; −5.8 (−28.5, 1.9)	−8.5; −6.4 (−28.2, 2.9)	0.1 (1.5)
VFI (%) SS	75.9; 83 (8, 100)	75.9; 83 (6, 100)	−0.0 (3.7)
VFI (%) SF	76.6; 82 (9, 100)	77.1; 84 (11, 100)	−0.5 (4.6)
VFI (%) SFR	77.6; 85 (11, 100)	77.1; 85 (11, 100)	0.4 (4.6)

MD = mean deviation; SD = standard deviation; SF = SITA Fast; SFR = SITA Faster; SITA = Swedish Interactive Thresholding Algorithm; SS = SITA Standard; VFI = visual field index.

TABLE 2. Relationships Between Differences in False Positive Response Rate Percentages and Mean Deviation and Visual Field Index Values, and Number of Significant Test Points in Total and Pattern Deviation Probability Maps Differences Between Visits 1 and 2

	r^2	Slope (Change per Percentage Point Increase in FP Rate) and 95% CI	Effect per 10 Percentage Point–Increase in FP
Diff MD/diff FP SS	0.01	0.04 (−0.02 to 0.08)	0.36 dB
Diff MD/diff FP SF	0.04	0.06 (0.01-0.11)[a]	0.60 dB
Diff MD/diff FP SFR	0.11	0.05 (0.03-0.08)[a]	0.51 dB
Diff VFI/diff FP SS	0.00	0.07 (−0.10 to 0.22)	0.56%
Diff VFI/diff FP SF	0.03	0.14 (−0.01 to 0.28)	1.37%
Diff VFI/diff FP SFR	0.04	0.10 (0.01-0.18)[a]	0.95%
Diff TD 1%/diff FP SS	0.00	−0.04 (−0.25 to 0.10)	−0.4 points
Diff TD 1%/diff FP SF	0.03	−0.15 (−0.30 to 0.01)	−1.5 points
Diff TD 1%/diff FP SFR	0.04	−0.09 (−0.17 to −0.01)[a]	−0.9 points
Diff PD 1%/diff FP SS	0.00	−0.02 (−0.17 to 0.13)	−0.2 points
Diff PD 1%/diff FP SF	0.00	0.00 (−0.11 to 0.10)	0.0 points
Diff PD 1%/diff FP SFR	0.00	−0.02 (−0.09 to 0.05)	−0.2 points

CI = confidence interval; diff = difference; FP = false positive; MD = mean deviation; PD = pattern deviation; SF = SITA Fast; SFR = SITA Faster; SITA = Swedish Interactive Thresholding Algorithm; SS = SITA Standard; TD = total deviation; VFI = visual field index.
[a]Statistically significant slope.

centage points in FP rates (for example, an increase in FP rate from 5% to 15%). Effects for VFI were even smaller 0.6%-1.4% (approximately corresponding to 0.2-0.4 dB), for an increase of 10 percentage points in FP rates. Similarly, the associations between FP intervisit differences and differences in numbers of significantly depressed test points were weak for all 3 test strategies, with many r^2 values close to 0. Most of those relationships were not statistically significant.

The relationships between intervisit differences in GH and MD were markedly stronger, with r^2 values ranging from 0.22 to 0.46 for the 3 strategies. The relationships between GH and VFI intervisit changes were weak but still much stronger than for FP vs VFI (Tables 2 and 3). The relationships between GH and number of significant TD test points were fairly strong across testing strategies but were much weaker for points in PD maps. This latter observation is not surprising, because GH was designed to correct for generalized changes in visual field sensitivity, such as those associated with cataract development.

Analysis of linear regression residual values revealed that assumptions implicit in linear regression were supported. Histograms of standardized residuals were normally distributed around zero. Most residual points in the scatter plots were within the ±2 intervals of standardized residuals on the y axes and randomly dispersed around standard-

TABLE 3. Relationships Between Differences in General Height and Mean Deviation and Visual Field Index Values, and Number of Significant Test Points in Total and Pattern Deviation Probability Maps Differences Between Visits 1 and 2

	r^2	Slope (Change per dB of GH) and 95% CI	Effect per 10-dB Change in GH
Diff MD/diff GH SS	0.35	0.55 (−0.42 to 0.69)	5.53 dB
Diff MD/diff GH SF	0.46	0.93 (0.75-1.1)[a]	9.25 dB
Diff MD/diff GH SFR	0.22	0.56 (0.37-0.74)[a]	5.55 dB
Diff VFI/diff GH SS	0.10	0.84 (0.39-1.28)[a]	8.35%
Diff VFI/diff GH SF	0.22	1.83 (1.22-2.43)[a]	18.25%
Diff VFI/diff GH SFR	0.03	0.64 (0.00-1.28)[a]	6.38%
Diff TD 1%/diff GH SS	0.35	−1.95 (−2.42 to −1.47)[a]	−19.5 points
Diff TD 1%/diff GH SF	0.39	−2.56 (−3.17 to −2.00)[a]	−25.6 points
Diff TD 1%/diff GH SFR	0.20	−1.51 (−2.05 to −0.90)[a]	−15.1 points
Diff PD 1%/diff GH SS	0.00	0.06 (−0.37 to 0.50)	0.6 points
Diff PD 1%/diff GH SF	0.00	−0.18 (−0.68 to 0.32)	−1.8 points
Diff PD 1%/diff GH SFR	0.11	0.92 (0.44-1.40)[a]	9.2 points

CI = confidence interval; diff = difference; GH = general height; MD = mean deviation; PD = pattern deviation; SF = SITA Fast; SFR = SITA Faster; SITA = Swedish Interactive Thresholding Algorithm; SS = SITA Standard; TD = total deviation; VFI = visual field index.
[a]Statistically significant slope.

TABLE 4. Relationships Between False Positive Response Rate Percentages and Mean Deviation Values at Different Stages of Glaucoma

Eyes (n)	Strategy	Stage	r^2	Slope (Decibel Change per Percentage Point Change in FP Rate)	95% CI	Effect per 10 Percentage Point–Increase in FP Rate (dB)
61	SS[c]	Early	0.01	0.03	−0.04 to 0.09	0.3[a]
25	SS	Moderate	0.00	0.02	−0.14 to 0.18	0.2
39	SS	Severe	0.05	0.11	−0.05 to 0.27	1.1
58	SF[d]	Early	0.06	0.04	−0.004 to 0.09	0.4
30	SF	Moderate	0.01	0.04	−0.13 to 0.22	0.4
37	SF	Severe	0.09	0.14	−0.01 to 0.28	1.4
63	SFR[e]	Early	0.06	0.03	0.001-0.05[b]	0.3
23	SFR	Moderate	0.14	0.05	−0.01 to 0.11	0.5
39	SFR	Severe	0.33	0.20	0.11-0.30[b]	2.0
37[a]	SFR	Severe	0.12	0.14	0.01-0.26[b]	1.4

CI = confidence interval; FP = false positive; SF = SITA Fast; SFR = SITA Faster; SS = SITA Standard.
[a]Two outliers excluded.
[b]Statistically significant slope.

ized predicted MD and VFI values on the horizontal axes (Supplemental Figure 1).

The influence of FP rates on MD and VFI when dividing the field tests into severity stages is presented in Tables 4 and Table 5. For MD, the relationships did not reach statistical significance at any disease stage with SS and SF. With SFR, they were statistically significant in severe glaucoma and borderline significant for SS and SF. Influences on VFI were statistically significant only for SF and SFR in severe glaucoma. Eliminating 2 outliers among the SFR results in the group of eyes with severe glaucoma (Supplemental Figure 2) reduced the slopes considerably. Corresponding results at different stages of glaucoma on numbers of significantly depressed TD and PD test points are shown in Tables 6 and 7. None of the relationships were statistically significant.

DISCUSSION

Our results indicate that across 3 different perimetric thresholding strategies, FP rate measurements generally showed only weak associations with visual field threshold

TABLE 5. Relationships Between False Positive Response Rate Percentages and Visual Field Index Values at Different Stages of Glaucoma

Eyes (n)	Strategy	Stage	r^2	Slope VFI Percentage Change per Percentage Change in FP	95% CI	Effect per 10 Percentage Point–Increase in FP (%)
61	SS	Early	0.00	0.01	−0.17 to 0.18	0.08
25	SS	Moderate	0.01	0.10	−0.36 to 0.55	0.95
39	SS	Severe	0.04	0.28	−0.20 to 0.76	2.81
58	SF	Early	0.01	0.05	−0.07 to 0.16	0.47
30	SF	Moderate	0.01	0,12	−0.38 to 0.61	1.17
37	SF	Severe	0.11	0.44	0.01-0.88[a]	4.43
63	SFR	Early	0.01	0.02	−0.04 to 0.08	0.23
23	SFR	Moderate	0.04	0.08	−0.11 to 0.26	0.76
39	SFR	Severe	0.25	0.62	0.26-0.97[a]	6.15
37[b]	SFR	Severe	0.09	0.43	−0.04 to 0.90	4.29

CI = confidence interval; FP = false positive; SF = SITA Fast; SFR = SITA Faster; SS = SITA Standard.
[a] Statistically significant slope.
[b] Two outliers excluded.

TABLE 6. Relationships Between False Positive Response Rate Percentages and Numbers of Significant Points at the $P < .01$ Level in Total Deviation Probability Maps

Eyes (n)	Strategy	Stage	r^2	Slope (Change in Number of 1% Points per Percentage Point Change in FP Rate)	*P* Value	Effect per 10 Percentage Point–Increase in FP
61	SS	Early	0.01	−0.05	.68	−0.5 points
25	SS	Moderate	0.00	−0.008	.98	−0.1 points
39	SS	Severe	0.00	−0.002	1.00	0.0 points
58	SF	Early	0.02	−0.08	.36	−0.8 points
30	SF	Moderate	0.02	−0.16	.52	−1.6 points
37	SF	Severe	0.07	−0.29	.11	−2.9 points
63	SFR	Early	0.01	−0.04	.42	−0.4 points
23	SFR	Moderate	0.07	−0.12	.21	−1.2 points
39	SFR	Severe	0.09	−0.25	.06	−2.5 points

FP = false positive; SF = SITA Fast; SFR = SITA Faster; SS = SITA Standard.

TABLE 7. Relationships Between False Positive Response Rate Percentages and Numbers of Significant Points at the $P < .01$ Level in Pattern Deviation Probability Maps

Eyes (n)	Strategy	Stage	r^2	Slope (Change in Number of 1% Points per Percentage Point Change in FP Rate)	*P* Value	Effect per 10 Percentage Point–Increase in FP Rate
61	SS	Early	0.00	0.01	.89	0.1 points
25	SS	Moderate	0.01	−0.11	.57	−1.1 points
39	SS	Severe	0.00	−0.05	.79	−0.5 points
58	SF	Early	0.01	0.03	.56	0.3 points
30	SF	Moderate	0.02	0.13	.52	1.3 points
37	SF	Severe	0.08	−0.21	.08	−2.1 points
63	SFR	Early	0.00	−0.02	.69	−0.2 points
23	SFR	Moderate	0.00	0.15	.82	1.5 points
39	SFR	Severe	0.06	−0.16	.14	−1.6 points

FP = false positive; SF = SITA Fast; SFR = SITA Faster; SS = SITA Standard.

sensitivity and associated analysis metrics. This finding is somewhat unexpected, but we see a parallel in the evolution of our thinking regarding the role of FN response rates >20 years ago.

The traditional definition of reliability in research is reproducibility. If that is what we want FP metrics to assess, then our findings suggest that the current FP index may be of little use. This is not at all a new finding, however, as similar results were reported 20 years ago; the results of reliability testing had almost negligible correlation with test reliability as expressed as threshold reproducibility.[11] If we instead define reliability as indicating the "usefulness" of test results, our findings show that FP measurement changes were associated with changes in test results in the same direction as in other published studies.[9,10,20] Therefore, in the current study, increasing rates of FP responses were associated with increases in MD values, but the effects on MD were small, except in severe disease, and even smaller for VFI. PD probability maps were not influenced at all, which is interesting because a higher number of significantly depressed PD test points than TD points is one of the classical hallmarks of a trigger-happy field. That observation alone shows the results reported herein: that the relationship of higher FP rates to signs of trigger-happy fields is weak to poor.

The effects of FP rates on MD were larger with SFR than with SS and SF, but in early glaucoma the slopes were small with all 3 algorithms, while in severe disease the slope with SFR was considerably larger than those of SS and SF. The SFR results in severe glaucoma were partially explained by 2 outliers (Supplemental Figure 2). In line with earlier reports, we thus found that FP rates seemed to be more important in eyes with severe field loss. PD probability maps were not influenced.

Tan and associates[9] and Yohannan and associates[10] both reported that FP influenced MD to a greater extent with higher frequencies of FP. We applied the analyses of Tan and associates[9] on our own data but could not confirm their findings. Therefore, in our material, the effect of FP on MD did not differ between eyes with high vs low FP values. Each percentage point of higher FP rate was associated with an increase in MD of 0.06 dB in eyes with FP ≤15% and 0.04 dB in eyes with FP >15%.

The main strength of the current study was that 2 visual fields were obtained with each of 3 perimetric thresholding algorithms within a very short time interval, eliminating the need to compare single test results to a model of an expected field, as has been the case in earlier studies.[10] Other strengths include our study's multicenter design and the fact that we could assess FP performance in subjects who were tested using 3 different threshold testing algorithms, making it possible to determine if observed trends were consistent across testing strategies, and the fact that we also studied the effects on the results expressed in probability maps.

A weakness of this study is the somewhat limited number of enrolled subjects. The material consisted mostly of eyes with manifest glaucoma but also contained glaucoma suspects. It would have been interesting to have had an equally large age-matched cohort of entirely normal subjects, each tested twice with all 3 strategies.

This study is not the first attempt to address the relationship between FP response rate measurements and MD values. Our results referring to MD values go in the same direction but are of a smaller magnitude than those reported in a large similar population of patients with suspect and manifest glaucoma[10] and in another large study of normal subjects.[9] These earlier studies have not reported the relationship of FP rates to the VFI, where we found even smaller effects, and, therefore, we cannot make any comparisons.

We have found no previous publications reporting the relationship of FP rates to the number of significantly depressed TD and PD points, which are central to the clinical interpretation of perimetric results. We found no significant influence of FP rates on PD probability maps.

One may speculate as to why SFR tests generate a larger number of FP responses than SF, and why SF generates more such responses than SS. According to signal detection theory, a more lenient response criterion leads to a higher rate of FP responses.[14] This has also been shown to happen in computerized perimetric testing, where instructions encouraging test subjects to use more lenient response criteria resulted in higher FP rates.[21] In the beginning of a visual field test, patients must set their own subjective response criteria. In SS and SF, the test starts with stimuli that are quite a bit more intense than normal threshold sensitivity, which usually are easily perceived. It seems likely that patients taking a SF test may then require stronger stimuli before responding than in SFR tests, which start out at the normal age-corrected threshold. The lack of clearly visible (supraliminal) stimuli in SFR tests makes it reasonable to assume that patients might then tend to adopt a more lenient response criterion and respond more often when not being sure of having seen a stimulus. This might explain the higher FP rates with SFR. During most of the test, SFR presents stimuli at the patient's predicted 50% threshold level, while SF presents stimuli that are approximately 1 dB brighter and SS 3 dB brighter, possibly explaining the smaller FP difference between SF and SS. The timing algorithms in SS and SF are identical, and the one used in SFR differs little from that of SS and SF. We do not believe that the differences of FP rates among the 3 algorithms are explained by the method used in SITA to assess FP rates, ie, to register as FP responses any button presses that occur during the first 180 milliseconds after stimulus initiation or during a period from the end of a response window until the onset the next stimulus exposure.[8]

The intended aim of the FP index was to flag perimetry test results from "trigger-happy" patients where those results cannot be trusted, and on that basis current methods of es-

timating FP rates are far from optimal. Therefore, it seems likely that test results should never be discarded solely based on FP response rates. It is encouraging to note that the correlation of GH with other test metrics was much greater than that of FP rates and was usually highly significant. It might perhaps be possible to construct a metric that is based in part on GH to better identify trigger-happy fields where test results should not be trusted.

ALL AUTHORS HAVE COMPLETED AND SUBMITTED THE ICMJE FORM FOR DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST and none were reported.

Funding/Support: The clinical study was supported by the Herman Järnhardt Foundation, the Foundation for Visually Impaired in Former Malmöhus County, Sweden. These funding organizations had no role in the design or conduct of this research. Carl Zeiss Meditec Inc., Dublin, California, USA was directly involved in the development of SITA Faster and loaned perimeters to 4 of the participating clinical sites. Carl Zeiss Meditec Inc. provided research funding for this study to J.G.F. at the University of California, Berkeley. Financial Disclosures: A.H. and B.B. are consultants of and are entitled to royalties from Carl Zeiss Meditec. A.H. is a consultant for Allergan plc and has received speaker honoraria from Allergan and Zeiss. V.M.P. is a consultant for Carl Zeiss Meditec and was a Carl Zeiss Meditec employee during the development and evaluation of SITA Faster. J.G.F. is a consultant for and received research support from Carl Zeiss Meditec, Inc. A.I. is a consultant for Santen and has received speaker honoraria from Pfizer, Santen, Kowa, Alcon, Heidelberg Engineering through Japan focus Company, and Carl Zeiss Meditec, Tokyo. A.I. also holds a patent licensed to Topcon without any royalties. C.K.L. has received speaker honoraria from Carl Zeiss Meditec, Topcon, Tomey, Allergan, Novartis, Santen, Glaukos, and Global Vision; research support in the form of instruments from Carl Zeiss Meditec, Heidelberg Engineering, Topcon, Tomey, and Optovue; research grants from Carl Zeiss Meditec, Topcon, Novartis, Glaukos, Alcon, and Optovue; consultant fees from Allergan and Novartis; and has patents with Carl Zeiss Meditec. A.T. has no financial disclosures, but Carl Zeiss Meditec provided the perimeter used in the clinical evaluation. G.C.L. and T.C. are employees of Carl Zeiss Meditec. All authors attest that they meet the current ICMJE criteria for authorship.

REFERENCES

1. Heijl A, Krakau CE. An automatic static perimeter, design and pilot study. *Acta Ophthalmol (Copenh)*. 1975;53(3):293–310.
2. Fankhauser F, Spahr J, Bebié H. Some aspects of the automation of perimetry. *Surv Ophthalmol*. 1977;22(2):131–141.
3. Anderson DR, Patella VM, eds. Automated Static Perimetry. Mosby; 1999.
4. Heijl A, Patella VM, Chong LX, et al. A new SITA perimetric threshold testing algorithm: construction and a multicenter clinical study. *Am J Ophthalmol*. 2019;198:154–165.
5. Katz J, Sommer A. Reliability indexes of automated perimetric tests. *Arch Ophthalmol*. 1988;106(9):1252–1254.
6. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Invest Ophthalmol Vis Sci*. 2000;41(8):2201–2204.
7. Haley M. *The Field Analyzer Primer*. 2nd ed. Humphrey Instruments; 1987.
8. 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(2):181–183.
9. Tan NYQ, Tham YC, Koh V, et al. The effect of testing reliability on visual field sensitivity in normal eyes: The Singapore Chinese Eye Study. *Ophthalmology*. 2018;125(1):15–21.
10. Yohannan J, Wang J, Brown J, et al. Evidence-based criteria for assessment of visual field reliability. *Ophthalmology*. 2017;124(11):1612–1620.
11. Bengtsson B. Reliability of computerized perimetric threshold tests as assessed by reliability indices and threshold reproducibility in patients with suspect and manifest glaucoma. *Acta Ophthalmol Scand*. 2000;78(5):519–522.
12. Heijl A, Lindgren G, Olsson J. A package for the statistical analysis of visual fields. In: Greve EL, Heijl A, eds. Seventh International Visual Field Symposium, Amsterdam, September 1986. Documenta Ophthalmologica Proceedings Series, vol 49. Springer, Dordrecht. doi:10.1007/978-94-009-3325-5_23.
13. Bengtsson B, Heijl A. Acceptable frequencies of false positive answers in computerized perimetry. *Invest Ophthalmol Vis Sci*. 2000;41(4):478.
14. Green DM, Swets JA. *Signal Detection Theory and Psychophysics*. Wiley; 1966.
15. Phu J, Khuu SK, Agar A, Kalloniatis M. Clinical evaluation of Swedish Interactive Thresholding Algorithm-Faster compared with Swedish Interactive Thresholding Algorithm-Standard in normal subjects, glaucoma suspects, and patients with glaucoma. *Am J Ophthalmol*. 2019;208:251–264.
16. Olsson J. *Statistic in Perimetry*. Lund, Sweden: Lund University, Department of Mathematical Statistics; 1991.
17. Asman P, Heijl A, Olsson J, Rootzen H. Spatial analyses of glaucomatous visual fields; a comparison with traditional visual field indices. *Acta Ophthalmol (Copenh)*. 1992;70(5):679–686.
18. Hodapp E, Parrish 2nd RK, Anderson DR. *Clinical Decisions in Glaucoma*. Mosby; 1993.
19. 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(1):24–30.
20. Junoy Montolio FG, Wesselink C, Gordijn M, Jansonius NM. Factors that influence standard automated perimetry test results in glaucoma: test reliability, technician experience, time of day, and season. *Invest Ophthalmol Vis Sci*. 2012;53(11):7010–7017.
21. Kutzko KE, Brito CF, Wall M. Effect of instructions on conventional automated perimetry. *Invest Ophthalmol Vis Sci*. 2000;41(7):2006–2013.