Measurement Precision in a Series of Visual Fields Acquired by the Standard and Fast Versions of the Swedish Interactive Thresholding Algorithm Analysis of Large-Scale Data From Clinics

Luke J. Saunders, MSc; Richard A. Russell, PhD; David P. Crabb, PhD

**IMPORTANCE** Swedish Interactive Thresholding Algorithm (SITA) testing strategies for the Humphrey Field Analyzer have become a clinical standard. Measurements from SITA Fast are thought to be more variable than SITA Standard, yet some clinics routinely use SITA Fast because it is quicker.

**OBJECTIVE** To examine the measurement precision of the 2 SITA strategies across a range of sensitivities using a large number of visual field (VF) series from 4 glaucoma clinics in England.

**DESIGN, SETTING, AND PARTICIPANTS** Retrospective cohort study at Moorfields Eye Hospital in London, England; Gloucestershire Eye Unit at Cheltenham General Hospital; Queen Alexandra Hospital in Portsmouth, England; and the Calderdale and Huddersfield National Health Service Foundation Trust that included 66,974 Humphrey 24-2 SITA Standard VFs (10,124 eyes) and 19,819 Humphrey 24-2 SITA Fast VFs (3,654 eyes) recorded between May 20, 1997, and September 20, 2012. Pointwise ordinary least squares linear regression of measured sensitivity over time was conducted using VF series of 1 random eye from each patient. Residuals from the regression were pooled according to fitted sensitivities. For each sensitivity (decibel) level, the standard deviation of the residuals was used to estimate measurement precision and were compared for SITA Standard and SITA Fast. Simulations of progression from different VF baselines were used to evaluate how different levels of precision would affect time to detect VF progression.

**MAIN OUTCOME AND MEASURE** Median years required to detect progression.

**RESULTS** Median (interquartile range) patient age, follow-up, and series lengths for SITA Standard were 64 (53-72) years, 6.0 (4.0-8.5) years, and 6 (4-8) VFs, respectively; for SITA Fast, medians (interquartile range) were 70 (61-78) years, 5.1 (3.2-7.3) years, and 5 (4-6) VFs. Measurement precision worsened as sensitivity decreased for both test strategies. In the 20 to 5 dB range, SITA Fast was less precise than SITA Standard; this difference was largest between 15 to 10 dB, where variability in both methods peaked. Translated to median time to detection, differences in measurement precision were negligible, suggesting minimal effects on time to detect progression.

**CONCLUSIONS AND RELEVANCE** Although SITA Standard is a more precise testing algorithm than SITA Fast at lower VF sensitivities, it is unlikely to make a sizeable difference to improving the time to detect VF progression.

Author Affiliations: Department of Optometry and Visual Science, School of Health Sciences, City University London, London, England. Corresponding Author: David P. Crabb, PhD, Department of Optometry and Visual Science, School of Health Sciences, City University London, Northampton Square, London EC1V OHB, England (d.crabb@city.ac.uk).
ogression of glaucoma needs to be monitored carefully to ensure appropriate treatment can be given to prevent sight loss. Visual field (VF) testing is the only direct means of measuring functional impairment in glaucoma but yields highly variable measurements. Moreover, measurement variability increases with declining VF sensitivity. As a result, detecting VF progression is far from straightforward, with significant potential for detecting false change or failing to detect definite change.

Visual field testing algorithms implemented in standard automated perimetry should generate measurements with precision and accuracy sufficient for monitoring progression effectively. Accuracy and precision are terms that are often confused; an accurate test is one that produces results with as little bias from true measurements as possible, whereas a precise test is one with high repeatability. The precision of a VF test can be optimized by repeated or extended testing but this is offset by the demand for quick examination in a clinical environment and a need to reduce the effect of fatigue on test performance. The Swedish Interactive Thresholding Algorithm (SITA) Standard was developed for the Humphrey Field Analyzer (HFA; Carl Zeiss, Meditec) in the 1990s and allows VF testing to be performed in about half the time taken by older full-threshold strategies, with no significant decrease in performance. Thus, SITA Standard has become a clinical standard for acquiring VF measurements. In contrast to full-threshold testing, SITA Standard uses prior information based on the sensitivities of the surrounding points; stimulus sequences are interrupted when the measurement error of the test point is small compared with a predetermined level of accuracy known as the error-related factor. A variant of SITA Standard, also available on the HFA, is called SITA Fast and provides an even quicker test time, often taking 5 minutes or less on average to administer.

In comparison with SITA Standard, SITA Fast achieves faster test times by presenting starting stimuli closer to expected thresholds and because stimulus staircases are interrupted at an earlier stage by increasing the error-related factor cutoff (thereby accepting lower accuracy of test results). The shorter duration of SITA Fast means that it is an appealing method from a practical and clinical perspective. Yet, there remains uncertainty about the precision of measurements from SITA Fast relative to SITA Standard and how this impacts time to detect VF progression. Therefore, the choice of testing algorithm presents a dilemma to physicians. The main aim of this article is to provide evidence for deciding which test should be used in clinical practice. We do this by clarifying the difference in precision between SITA Standard and SITA Fast across the full range of VF sensitivities in large volumes of patient data from ordinary clinics. We then estimate the clinical impact of choice of VF test algorithm by considering the average time to detect progression using computer simulations.

**Methods**

We analyzed 473,252 anonymized VFs from 88,954 patients from databases at Moorfields Eye Hospital glaucoma clinic in London, England (320,334 VFs); Gloucestershire Eye Unit at Cheltenham General Hospital (50,144 VFs); Queen Alexandra Hospital in Portsmouth, England (31,879 VFs); and the Calderdale and Huddersfield National Health Service Foundation Trust (70,955 VFs). The study adhered to the Declaration of Helsinki and was approved by a research governance committee of City University London, and all anonymized data were transferred to a secure database. Patients did not provide written informed consent for participation owing to the retrospective nature of this study. Only VFs from the HFA using Goldmann size III stimuli with the 24-2 test pattern and either SITA Standard or Fast were included in the study. Eyes were excluded if they had fewer than 5 VF examinations; if both eyes fulfilled these criteria for an individual patient, 1 eye was selected at random. In addition, patients younger than 35 years were excluded from the study. The first VF was removed to attempt to account for perimetric learning effects, leaving a total of 86,793 VFs from 13,778 patients for analysis, 66,974 SITA Standard VFs from 10,124 patients and 19,819 SITA Fast VFs from 3,654 patients.

Precision of SITA Standard and SITA Fast was assessed by fitting ordinary least squares regression to each VF test location’s sensitivity series (pointwise linear regression [PLR]). Fitted values were obtained from the PLR model for each test location (excluding the blind spot test locations) and raw residuals were then calculated by subtracting the calculated fitted values from the observed sensitivities at each point in the individual series. The raw residuals were pooled and binned according to fitted sensitivity (rounded to the nearest decibel). This approach described in detail by Russell et al was carried out separately for eyes tested using SITA Standard and SITA Fast so that the precision of the 2 methods could be compared.

To illustrate the clinical relevance of any additional variability associated with either of the 2 algorithms, computer simulations determined how soon VF progression could be diagnosed. The methods used to simulate VF progression are reported in more detail elsewhere. Ten thousand pointwise VF series were simulated to deteriorate at 3 different rates (or speeds) of VF loss over time (−0.5, −1, and −2 dB per year) from 3 separate starting sensitivities (30 dB, 20 dB, and 10 dB). Variability was derived from the distributions of PLR residuals extracted for each algorithm and input into the computer simulation. The times taken to detect deterioration at the P < .01 significance level with 2 tests per year were recorded and median detection times were compared between SITA Standard and SITA Fast. All statistical analyses were carried out in the open-source programming language, R.
Measurement Precision in the Swedish Interactive Thresholding Algorithm

Table 1. Characteristics of Study Sample

<table>
<thead>
<tr>
<th>Measurement</th>
<th>SITA Standard (10 124 Eyes)</th>
<th>SITA Fast (3654 Eyes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VF tests, No.</td>
<td>6 (4 to 8)</td>
<td>5 (4 to 6)</td>
</tr>
<tr>
<td>Baseline VF sensitivity, dB</td>
<td>27 (23 to 30)</td>
<td>28 (24 to 30)</td>
</tr>
<tr>
<td>Baseline MD, dB</td>
<td>−3.23 (−7.37 to −1.03)</td>
<td>−2.20 (−5.42 to −0.38)</td>
</tr>
<tr>
<td>Pointwise progression rate, dB/y</td>
<td>−0.13 (−0.55 to 0.17)</td>
<td>−0.16 (−0.62 to 0.17)</td>
</tr>
<tr>
<td>MD progression rate, dB/y</td>
<td>−0.11 (−0.43 to 0.12)</td>
<td>−0.14 (−0.49 to 0.09)</td>
</tr>
<tr>
<td>Baseline age, y</td>
<td>64 (53 to 72)</td>
<td>70 (61 to 78)</td>
</tr>
<tr>
<td>Follow-up period, y</td>
<td>6.0 (4.0 to 8.5)</td>
<td>5.1 (3.2 to 7.3)</td>
</tr>
</tbody>
</table>

Abbreviations: MD, mean deviation; SITA, Swedish Interactive Thresholding Algorithm; VF, visual field.

Discussion

Consistent with previous studies, the variability of measurements generated using both the SITA Standard and SITA Fast algorithms increased with decreasing sensitivity for both tests before declining approaching perimetric blindness. One reason why variability did not continually increase with VF loss is owing to no measurements possible below 0 dB; this floor effect seemed to have some impact from below 20 dB (Figure 1).

When compared with SITA Standard, SITA Fast had slightly worse precision overall, although this difference was negligible at test locations, with little or no sensitivity loss; in fact, SITA Fast seemed to be marginally more precise at this stage. Differences in precision between methods were more apparent when thresholds dropped closer to 12 dB and were largest between 15 dB to 10 dB, which matched the findings of Artes et al. Because variability differences between the 2 tests were low at near-normal thresholds, the shorter test times associated with SITA Fast may be preferable in glaucoma suspects or patients with early VF loss. Furthermore, simulation results indicated that any modest sacrifice in precision using SITA Fast makes little difference to the time to detect pointwise progression (Table 2). Research suggests that a 30% to 60% drop in variability is required in standard automated perimetry to make a significant difference to the speed at which deterioration is diagnosed; this level of difference was certainly not apparent in our study. Conversely, the reduced test time of SITA Fast relative to SITA Standard (roughly 2 to 3 minutes, according to the literature) may not be sufficient to make using this test worthwhile in the clinic.

Analysis of average times to detect different rates of VF deterioration using a published model for simulating VF progression was very informative (Table 2). When the level of pointwise VF damage dropped below 20 dB, the precision of the 2 algorithms was perhaps too poor for detecting significant pointwise VF progression in a reasonable period. For example, at 20 dB, the model suggests it takes more than 8 years to detect a rate of −1 dB per year with an examination every 6 months, whether using SITA Standard or SITA Fast. This limitation of standard automated perimetry has been well reported and recently published data indicate that recording at these sensitivities, regardless of testing algorithm, may simply be unreliable; an apparent change in sensitivity within this range may not be informative of disease progression. If this is the case, differences in precision of this part of the measurement scale would be meaningless; SITA Fast could be as effective as SITA Standard for follow-up.

Our study did not consider the accuracy of the 2 testing algorithms. Evidence from computer simulation and real patient data suggests that both SITA strategies (although SITA Fast more so) tend to produce threshold values systematically higher than the original HFA full-threshold algorithm. However, this does not necessarily mean that SITA Standard and SITA Fast are less accurate than full-
Figure 1. Distributions of Residuals From Linear Regression at Different Fitted Sensitivities

Back-to-back histograms showing the frequency density of raw residuals generated through linear modeling of each test location at rounded fitted values of 0 dB (A), 10 dB (B), 15 dB (C), 20 dB (D), 25 dB (E), and 30 dB (F). SITA indicates Swedish Interactive Thresholding Algorithm.

threshold testing because this assumes that full threshold is a gold standard. There is no gold standard good enough to perfectly evaluate the accuracy of these test algorithms; repeated testing of points to generate frequency of seeing curves could be considered 1 potential approach to calculate thresholds more accurately; this could still be subject to nonconstant variability across sensitivities, interpatient variability, and fatigue. Most importantly, VF total deviation measurements are age corrected and centered so they should not be affected by bias. Given the similar scale of the rival methods, the relative accuracy of the tests should not affect utility in monitoring progression. In fact, smaller variability in the SITA normative database has been shown to result in these algorithms actually being more sensitive to change than full-threshold testing. Because both methods use surrounding thresholds to help derive sensitivity values and because SITA Fast uses a lower threshold of measurement certainty to reduce the number of stimulus...
presentations,\textsuperscript{10} it is sometimes thought that VF damage has a higher likelihood of being missed or wrongly recorded in SITA Fast compared with SITA Standard. This phenomenon was not tested in our study and can only be investigated in a prospective study. Calculating signal-noise ratios in such a study as described by Artes et al could be one useful technique for comparing the utility of these algorithms.\textsuperscript{23}

One strength of our study was that the precision estimates were based on millions of VF points from real clinical data rather than from a small number of people taking part in a more controlled experiment. The sheer size of this data set means that while there were bound to be anomalies and outliers, the results were overall most likely a good reflection of the precision of perimetric testing found in day-to-day clinical practice. One drawback to our approach was the lack of definitive clinical knowledge about the individual patients included in the study. For instance, patients may have had comorbidities that affected the VF (like cataract) that could have affected estimates. Even the type of diagnosis of glaucoma was not available; it was only known that patients were in glaucoma clinics. No exclusion criteria based on reliability criteria (false positives, false negatives, and fixation losses) were applied to the data sets because much of these data were unavailable for our study. This is a limitation from the perspective that VFs with values for these indices greater than an arbitrary determined criterion are often excluded in clinical practice. However, we believe that the reliability indices are possibly the most unreliable measures on the printout and make the assumption that unreliable VFs would have existed in equal measure in SITA Fast and SITA Standard VFs and that this would not impact our findings. Our estimates of precision only represent an average rather than a measure that can be applied to different individuals; patients perform differently at perimetry and some produce measurements with higher variability than others.\textsuperscript{15,24} Furthermore, the distribution of cases is likely dominated by those with smaller amounts of VF loss and tests that are within normal limits, as the higher density of points in the bottom right corner of Figure 2 suggests. In addition, our results estimated precision based on following up patients longitudinally rather than obtaining test-retest data; therefore, they are only relevant to the follow-up rather than the diagnosis of a single VF.

There are other aspects of our study that demand greater scrutiny. For instance, as a group, patients who tested with SITA Fast were on average older than those tested using SITA Standard. We speculate this may reflect the preconception that older people are thought to be worse at perimetry,\textsuperscript{25-27} which could lead to these patients being offered the quicker test in the belief that it is more suited to them. Also, the baseline MDs for SITA Fast tended to be higher than those for SITA Standard. This may be explained by selection bias because the clinic databases would include people being followed up without VF defects (patients at risk of developing glaucoma and ocular hypertensives) who may have been more likely to be assigned to the quicker test. One assumption made in the study was that the model-fitted VF sensitivities were akin to patients’ true sensitivities, which is reliant on assuming a linear relationship between change in VF thresholds and time. Alternative trend analyses modeling sensitivity changes over time have been proposed,\textsuperscript{28-31} yet PLR is commonly used and in practice is likely to be as effective (if not more) at predicting future loss as these other methods.\textsuperscript{32,33} For this study, the regression is used to extract an estimate of precision via the residuals. Pointwise VF variability was analyzed without regard for location in the VF, although it is already known that variability tends to differ according to test location.\textsuperscript{24}

However, this is confounded by the fact that sensitivity, which is not independent of variability, also varies according to test location.\textsuperscript{13,23} As of yet, there is no clear evidence to suggest that there is a clinically significant difference in variability in test locations independent of sensitivity.

Finally, it is most important to reflect on patient preferences for different types of perimetry; it is imperative...
to account for this when choosing a testing algorithm. For example, some patients may feel fatigue more acutely than others and benefit from the shorter testing time of SITA Fast while for other patients, it may be valuable to make use of the additional precision that comes with SITA Standard, particularly if the disease has progressed beyond an early stage. In our experience and that of clinical colleagues, some patients complained of having to delineate a high number of barely perceptible stimuli with SITA Fast because the algorithm presented stimuli closer to thresholds to reduce the test time. Therefore, it is wrong to assume that SITA Fast is an easier test. In practice, it seems important to select the algorithm most suited to the patient and use this algorithm consistently throughout their follow-up, as clinical recommendations already stipulate.34

Conclusions

Overall, this study suggests there is some reduction in measurement precision associated with using SITA Fast instead of SITA Standard. However, this difference is unlikely to have a significant clinical impact on the time to detect VF progression. This study did not allow for any conclusion to be drawn about comparative accuracy of the measurements from SITA Standard and SITA Fast. Nevertheless, in the absence of patient preferences for a particular testing strategy, we conclude that clinicians can use either strategy to follow up patients over time, with the knowledge that precision in detecting VF progression will be practically equivalent.


