

ORIGINAL ARTICLE

Gaze tracker parameters have little association with visual field metrics of intrasession frontloaded SITA-Faster 24–2 visual field results

Jack Phu^{1,2}  | Michael Kalloniatis^{1,2} 

¹Centre for Eye Health, University of New South Wales, Kensington, New South Wales, Australia

²School of Optometry and Vision Science, University of New South Wales, Kensington, New South Wales, Australia

Correspondence

Jack Phu, School of Optometry and Vision Science, University of New South Wales, Kensington, New South Wales, Australia.
Email: jack.phu@unsw.edu.au

Funding information

National Health and Medical Research Council, Grant/Award Number: 1186915; University of New South Wales Faculty of Science Early Career Academics Network

Abstract

Purpose: To determine the usefulness of Humphrey Field Analyser (HFA) SITA-Faster 24–2 gaze tracker outputs on interpreting intra-visit visual field (VF) result pairs.

Methods: Analysis of 1380 right–left eye pairs and 1432 pairs of test 1–test 2 intrasession VF results of patients seen within a university-based glaucoma service was undertaken to understand gaze deviation distributions. Output gaze tracker results were aggregated into total ticks, sum of amplitudes and average amplitudes. Correlations between visual field indices (mean deviation [MD], “events” and overall hill of vision) and independent variables (age and test order) were performed using one eye from each subject.

Results: There was no association of test order (right–left, test 1–test 2) with eye movements. There was a significant, but weak correlation between eye movements and age ($r = 0.16$). Correlations of eye movements with MD were driven by more severe MD values. There were no significant correlations between intrasession difference in eye movements and the change in MD, number of “events” and hill of vision, or in the root mean square of sensitivity and total deviation values. There was also no significant correlation between gaze tracker outputs and another commonly used “reliability” metric, false positive rate.

Conclusions: Eye movement parameters as currently reported by the HFA do not appear to be correlated with key sensitivity parameters when considering the repeatability of intrasession SITA-Faster 24–2 VF results. Thus, current gaze tracker outputs do not appear to provide clinically meaningful information for interpretation of intra-visit visual field results that cannot already be garnered using other strategies.

KEYWORDS

eye movements, false positive, Humphrey field Analyser, SITA

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Ophthalmic and Physiological Optics* published by John Wiley & Sons Ltd on behalf of College of Optometrists.

INTRODUCTION

In clinical practice, the usefulness of information provided by static automated perimetry in the assessment of diseases of the visual pathway is often tempered by sources of test variability.¹ Factors contributing to output variability need to be accounted for to ensure accurate interpretation of the patient's visual field status. As such, clinicians need to be able to recognise when an output result is clinically useful or "reliable."

Historically, several output parameters for quantifying result reliability have been recommended for clinical use. These have included fixation losses,² false positive and negative catch trials³ and the identification of seeding point errors.⁴ These indices are aimed at identifying causes of artificial alterations to the visual field. Over time, an increased understanding of the relationship between these indices and visual field sensitivity metrics have led to the reappraisal of the usefulness of these metrics, with suggestions that "traditional" indices of "reliability" contribute little to reproducibility of sensitivity measurements,⁵ especially in the context of visual field defects.⁶

Recently, the development of SITA-Faster and its clinical implementation has been accompanied by the abandonment of both fixation loss (using the Heijl-Krakau method) and false negative catch trials in its default setting.⁷ The false positive metric and gaze tracker output remain as automated indices reported for interpretation of reliability. Despite the automatic reporting, current recommendations for interpreting visual field test results de-emphasise reliance on automated indices in general.⁸

A recent study by Heijl et al.⁹ challenged the historical—and current manufacturer reported—cut-off values for "elevated" false positives, with such false positive rates having poor relationships with output test results. This work was specifically topical due to reports of higher false positive rates found on SITA-Faster compared to SITA-Standard.^{10,11} The reason for this is thought to be in part due to the nature of the catch trial, which relies on measurements of response and stimulus timing used in the thresholding algorithm of the Humphrey Field Analyser, in which a false positive result is identified if a response is provided within a certain time window before or immediately after stimulus onset.¹² It has been proposed that it may be due to the adoption of a more lenient response criterion due to the use of more near-threshold, rather than supra-threshold, stimulus intensities in SITA-Faster.⁹ Given the questions raised regarding false positive rates and true erroneous perimetric outputs, it has been recommended that careful analysis of other signs of trigger-happy results should be conducted, instead of relying on historical cut-off values for binarised pass/fail in reliability.¹¹

The second of the main "reliability" metrics reported in SITA-Faster, the gaze tracker, is aimed at obtaining an impression of fixation stability. Previous studies examining its relationship with sensitivity outputs have primarily used

Key points

- There was no association between the test order (right versus left, first versus second test) of intrasession (frontloaded) visual field test results and gaze tracker metrics
- Mean deviation and age showed weak correlations, and other key perimetry metrics—sensitivity, "events" and the hill of vision—showed no correlation with gaze tracker metrics
- Gaze tracker metrics as currently reported using vertical scalar "ticks" do not appear to provide clinically meaningful information for interpreting intrasession visual field test results

SITA-Standard,^{13–15} which is a longer test and thus potentially more likely to return more lapses in concentration compared to SITA-Faster. These studies suggested that some gaze tracker metrics may affect output global metrics, such as mean deviation. In a more recent study Camp and colleagues¹⁶ examined four nominal categories of gaze tracker metrics in SITA-Faster, and did not find a clinically significant association with other metrics of "reliability." The other related finding was that a greater number of large eye movements (>6 degrees) was associated with visual field severity.¹⁶

Historically, there have been concerns regarding the resolution (or accuracy) of defining fixation loss or instability using catch trials, and their effects on the perimetric outcome. For example, Vingrys and Demirel¹⁷ demonstrated that although catch trials for monitoring test reliability were generally accurate in estimating false responses, there were wide confidence intervals and thus estimations were not particularly precise. Although their work examined catch trials to infer reliability of perimetric outputs,¹⁷ alongside the work of Newkirk et al.,¹⁸ there are important concerns regarding the amount of sampling required to estimate fixation stability and reliable responses properly.

Demirel and Vingrys¹⁹ also showed that fixation stability in a cohort of normal subjects within 3 degrees of the fixation target did not occur throughout the entirety of a perimetric test (albeit in a longer, older generation thresholding algorithm). Injection of intentional eye movements led to increased short-term fluctuations of threshold measurements, but when extrapolated across the entirety of the visual field test grid, were unlikely to cause significant sensitivity differences. Similarly, a study using retinal stabilised perimetric testing to measure frequency of seeing curves found increased threshold variability near the blind spot, but little difference in threshold sensitivity.²⁰ A similar finding was reported by Kimura et al.²¹ using a head-mounted, eye-tracking perimeter. Taken in combination, fixation instability, in general, appears to have a minor impact on thresholds aside from at the edge of a scotoma. Contextually,

despite some reservations in resolution and debate regarding the importance of monitoring gaze during the perimetric test, SITA-Faster only returns qualitative gaze tracker data for interpreting reliability. Recommendations for interpreting gaze tracker outputs are primarily qualitative, with no quantitative measure for simpler clinical interpretation. Quantitative parameters may be more readily interpretable, rather than relying on subjective analysis of qualitative data. Therefore, it would be clinically informative to understand the relationship between objective measurements of gaze tracker outputs from static automated perimetry and the resultant sensitivity measurements.

The purpose of the present study was to describe and examine the gaze tracker outputs in SITA-Faster visual field results obtained from a cohort of patients seen within a glaucoma service. We were specifically interested in the correlations between visual field results performed within the same clinical visit ("frontloaded": a method used to obtain multiple perimetric data points for clinical interpretation^{22,23}). The central hypothesis was that gaze tracker metrics provide useful information in visual field interpretation in the form of correlations with other gaze tracker metrics and output sensitivity. We performed two main analyses to test our central hypothesis. First, we examined the correlations between the gaze tracker output and false positive rates and patient-specific factors, as these might influence and confound analysis of eye movements. Understanding these factors would be important in identifying potential confounders in developing clinically relevant parameters. Second, we examined the association of gaze tracker metrics with output sensitivity measurements in SITA-Faster. Thus, the combination of understanding gaze tracker metrics, their correlations and the output sensitivity would potentially enable the development of guidelines for clinical interpretation of perceived reliability of the result.

METHODS

Ethics statement

This was a cross-sectional study using prospectively acquired data from the files of patients seen within the Centre for Eye Health, University of New South Wales. Ethics approval was provided by the Human Research Ethics Committee of the University of New South Wales (HC210563). The study adhered to the tenets of the Declaration of Helsinki. All subjects provided their written informed consent for use of their de-identified clinical data for research purposes.

Subjects

Subject data was acquired from consecutive patients seen within the general and glaucoma service of the

Centre for Eye Health, University of New South Wales between 1 September 2020 and 31 March 2021. The clinic is a referral-only, optometry-ophthalmology service, providing assessment and management of patients with (or are referred for suspicion of) diseases of the visual pathways, including glaucoma. The subjects were part of the Frontloading Fields Study (FFS), an ongoing study at the Centre for Eye Health examining the deployment of frontloaded SITA-Faster visual fields in clinical decision making and patient management.²² For the present study, we included all subjects whose visual field result had a gaze tracker output.

We categorised the ocular diagnoses of eyes from patients within the present cohort into one of four categories, based on the review of their medical record. The diagnostic categories for the test eye were no evidence of diseases of the visual pathway including no evidence or suspicion of glaucoma (healthy, $n = 379$ eyes); glaucoma suspect ($n = 849$ eyes); manifest glaucoma ($n = 343$ eyes) or non-glaucomatous optic atrophy ($n = 23$ eyes). The method for diagnosis has been described in detail in our previous studies.^{10,22} The diagnosis of glaucoma was made as per current clinical guidelines,^{24,25} which included glaucomatous structural defects (enlarged or asymmetric cup-to-disc ratio, diffuse or focal rim thinning and adjacent retinal nerve fibre layer defects that were not explained by other retinal or neurological pathologies) with or without retinotopic visual field loss (i.e., patients with pre-perimetric glaucoma were not excluded). Elevated intraocular pressure was not required for diagnosis. The diagnosis of an eye as "glaucoma suspect" was made if the structural or functional findings were suspicious but not conclusive for a diagnosis of glaucoma, or if one or more risk factors for glaucoma were present. An eye with signs of optic atrophy that were attributable to causes other than glaucoma was defined as "non-glaucomatous optic atrophy." A healthy eye had normal structural and functional findings that did not meet any of the above criteria. The diagnoses were extracted from the patient's medical record. As per the clinical protocols of the Centre for Eye Health,²⁶ a diagnosis was made by an examining clinician, with remote review by a senior clinician working within the clinic. A third expert further examined the record for inclusion in the present study.

Visual field data extraction

As part of the FFS²² and the current clinical protocols at Centre for Eye Health, all patients underwent visual field testing twice for each eye within the same test session. The order of testing was at the discretion of the administering technician, with rest breaks between each test as requested by the patient. All testing was performed using the Humphrey Field Analyser 3 instrument, using the 24-2 test grid and SITA-Faster algorithm (Carl Zeiss Meditec, zeiss.com).

Visual field data of interest were the right and left eye (or only eye in cases where the patient was monocular) results collected within the same clinical visit. A custom written MATLAB program (MathWorks, mathworks.com) was used to extract the following parameters of interest from each visual field printout: pointwise visual field sensitivity, pointwise total deviation numerical values, pointwise pattern deviation probability scores, mean deviation, pattern standard deviation and false positive rate. In addition, an image analysis component of the custom program was used to extract out gaze tracker ticks (see more below). Other demographic information was extracted from the subject's medical record (VIP, Best Practice Software, bposoftware.net).

Gaze tracker metrics and parameterisation

Interpretation of the gaze tracker output has been provided by both manufacturer guidelines and by previous studies.^{8,13} In brief, the horizontal line indicates the instances where eye tracking was performed. A tick above the horizontal line indicates eye movement away from fixation, with a taller tick indicating greater amplitude (incremented in 1 degree steps). An example of a pair of intrasession repeat visual field test results with unstable fixation is shown in [Figure 1a](#). A short tick below the line indicates a tracking error and a long tick indicates a blink artefact. The retest

variability of the gaze tracker has been reported to be 2 degrees,²⁷ with previous studies commonly reporting gaze deviations in increments of 2–3 degrees.^{14,28} The tallest tick represents 10 degrees, with the smallest tick indicating 2 degrees and no upward tick indicating 0–1 degrees, in line with the presumed resolution of the device. Since no upward tick is provided, the minimum eye movement tick that was incorporated into subsequent analyses was 2 degrees of movement. We also note that the ticks represent the gaze deviation at the point of measurement, but may not necessarily reflect active eye movements. Nonetheless, the methods described herein represent interpretation of current clinically-available outputs.

In addition to counting the proportion of each tick identified in each gaze tracker result, we used three aggregate measurements to describe the eye movements shown by each subject ([Figure 1b](#)). The first aggregate measurement was the overall number of ticks above the line, i.e., any deviation from fixation. The second aggregate measurement was the total amplitude of movement. This measurement was defined as the sum of the upward ticks and their associated amplitude (i.e., the “sum of amplitudes”) expressed as total degrees per test, i.e., the number of 2 degree ticks multiplied by 2 degrees, number of 3 degree ticks multiplied by 3 degrees, and so on. The third aggregate measurement was the average eye movement over the total test duration. This was calculated dividing the second aggregate measurement (sum of amplitudes of

(a) Case examples of visual field tests with frequent large eye movements



(b) Aggregate eye tracker parameters examined in the present study

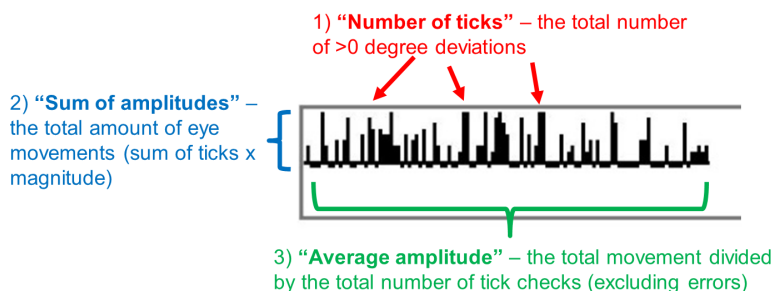


FIGURE 1 (a) Case examples of visual field test results with frequent large eye movements. The sensitivity and pattern deviation maps are shown at the top, false positive rate and mean deviation in red and the gaze tracker output at the bottom. (b) Aggregate eye tracker parameters used in the present study. Red: The number of individual ticks above the horizontal line (deviations greater than 0 degrees). Blue: The sum of amplitudes (sum of the ticks multiplied by their magnitude). Green: Average amplitudes (sum of amplitudes divided by the total number of tick checks excluding errors).

movement) by the number of ticks counted during the test. For all calculations, all tracking errors were excluded from the analysis. We noted that the maximum pixel increment for the ticks above the horizontal line was nine steps, thus implying that deviations greater than 10 degrees would be included in this group. This potentially means that subjects with more instances of “10 degree ticks” may have underestimated eye movements.

Exploratory analysis of correlations between gaze tracker outputs

As part of devising the above aggregate gaze tracker parameters, we performed an exploratory analysis to examine for correlations between the different outputs from the gaze tracker (i.e., the specific amplitudes) and false positive rates. For example, we examined whether there were internal correlations between individual gaze tracker outcomes (for example, 2 degree movements, 3 degree movements and so on) and false positive rates. The presence of internal correlations means that the use of aggregate gaze tracker or “reliability” metrics may be suitable as they would reflect an ascending amount of eye movements. The potential correlations between gaze tracker outputs and false positive rates (but not including other visual field metrics) were analysed using principal components analysis, as per our previous reported methods.²⁹ In brief, the results of this analysis revealed positive correlations between the specific upward ticks that indicated amplitudes of 4 degrees or greater, which were in turn negatively correlated with the specific ticks indicating amplitudes of 0–3 degrees. Simply put, this meant that subjects who had some gaze deviations of at least 4 degrees were likely to have more of such “higher” amplitude deviations, and fewer small or negligible deviations (0–3 degrees). If the converse were true, i.e., no correlations between groups of relatively small or relatively large amplitudes of movements, then the use of aggregate gaze metrics would be less logical. False positive rates were not correlated with any eye movement variable. Further details are provided in the Appendix S1.

Given these results, it appeared that eyes could potentially be distinguished by the proportion of low or high amplitude eye movements (i.e., a propensity to have more or fewer deviations as per the above correlations). An aggregate gaze deviation metric would thus be expected to identify eyes in which greater amounts of movement were present, providing contrast along a spectrum of potential gaze deviations.

Whilst the aggregate measures themselves may be correlated (thus potentially increasing the type I error when performing multiple separate correlations), the goal was to identify potentially useful, summary metrics for clinical interpretation. Thus, we continued to report them separately in our initial exploratory analysis, which could then undergo further analyses to confirm significance.

Analysis 1: The influence of patient-specific factors on gaze tracker metrics

We next examined patient-specific factors that might influence and confound analysis of gaze tracker metrics. Factors that contribute significantly to the output of gaze tracker metrics might therefore be confounders for analysis of sensitivity or reliability measurements.

The independent factors analysed were age, eye laterality (right or left), test order (first or second) and visual field mean deviation. These were analysed against the aggregate gaze tracker metrics. For the continuous variables age and mean deviation, a correlation analysis was used to obtain the significance of the slope and the correlation coefficient (r). For the categorical variables, we analysed the difference between tests (i.e., the difference between right and left eye results and the difference between first and second test results). The results were then grouped together and a one-sample t -test was used to determine if there were significant differences from zero (no difference in gaze tracker metric).

From this analysis, our anticipated outcome is the identification of independent factors that might influence aggregate gaze tracker metrics that might need to be accounted for in clinical interpretation.

Analysis 2: The relationship between gaze tracker metrics and sensitivity measurements

After identifying potential confounders and internal correlations in the two approaches noted above, we then examined the relationship between gaze tracker metrics and sensitivity measurements.

We performed analyses of correlation of the gaze tracker metrics against several parameters related to visual field sensitivity: mean deviation, number of “events” (points identified as significant at the $p < 0.05$ or lower level on the pattern deviation map) and the average hill of vision. We were interested in mean deviation and number of “events” as these parameters are typically used in the staging of glaucoma. These values were extracted directly from the Humphrey Field Analyser printout.

The hill of vision has been variously defined in the literature.^{30–33} In static automated perimetry, an individual's hill of vision is typically used to scale their overall visual field sensitivity and deviation results to create the pattern deviation map. Accounting for the hill of vision aims to facilitate identification of subtle clusters of defects that might otherwise be missed if the patient had a high (for example, abnormally high sensitivity) or low (for example, due to generalised media opacities like cataracts) hill of vision. One method for estimating the hill of vision is to take the 7th most positive value on the total deviation map (approximately corresponding to the 85th percentile).^{34,35} To capture the extent of the height of the hill of vision, we took the average of the seven most positive values for each subject to represent the average hill of vision. An average of the seven highest sensitivity

values, instead of selecting only the 7th ranked sensitivity value, may provide an impression of whether more points with “higher sensitivity” were present that may otherwise be missed (i.e., a higher “ceiling”).

For the above variables, we determined the correlations between the difference in gaze tracker and sensitivity results between tests 1 and 2. This was to account for inter-individual differences in gaze patterns.

We also examined the relationship between gaze tracker metrics and the root mean squared error³⁶ of the correlations of intrasession pairs of sensitivity and total deviation values, which provides an impression of the degree of result repeatability. In contrast to the differences described above, the root mean squared error is positively signed, and thus was correlated against the absolute difference in gaze tracker metrics.

From this analysis, our anticipated outcome is the ability to account for differences in clinically relevant visual field sensitivity outputs through the interpretation of aggregate gaze tracker metrics.

Quartile analysis

Since there may be temporal variations in gaze behaviour (such as at the start or at the end of the test), we also performed a sub-analysis where we analysed separately the gaze tracker outputs into first and fourth (last) quartiles. The output sensitivity results from the Humphrey Field Analyser do not provide accompanying information regarding the time that it was measured. However, the test locations are assessed in pseudo-random order. Therefore, for the first “quartile”, we extracted the first four primary seeding points (at locations 9 degrees vertical and 9 degrees horizontal from the midline in each quadrant) and for the fourth “quartile”, the 14 peripheral-most points were used.

For all analyses, if a subject contributed multiple results (for example, in situations where both eyes or both first and second visual field results were valid), only one visual field result was selected at random for analysis. This was to reduce the contributions of intra-individual correlations and the association with test order. As only one result from each subject was used, we analysed the data using simple Spearman rho correlations, rather than linear mixed models.

RESULTS

We analysed 2947 visual field results (mean age 59.6 years, SD 13.4; 409 males, 371 females) with valid gaze tracker outputs to describe the frequency of each gaze tracker metric tick. For the correlation analyses, we examined the results of a subset of 1380 right-left eye pairs (mean age 59.5 years, SD 13.5; 379 males, 343 females) and 1432 pairs of test 1-test 2 (mean age 59.6 years, SD 13.5; 400 males, 360 females) intrasession visual field results. The mean deviation characteristics of the visual field test results analysed

in the present study were as follows: median, -0.95 dB, interquartile range -2.34 to 0.03 dB and full range from -29.73 to 8.83 dB. We noted that the upper limit of the full range of mean deviation was high. This was because the study criteria intended to include a diverse range of perimetric sensitivities encompassing those that are typically deemed to be unreliable or “supra-sensitive.”

The proportions of gaze tracker ticks for each subject were divided into their respective outcomes (for example, tracking error, 0–1 degrees, 2 degrees and so on) and their relative frequency across all subjects is shown in [Figure 2](#). There were several findings evident from these distributions. A large proportion of subjects had low occurrences of tracking error and large eye movements of 6 degrees or higher. Most of the eye movements were 0 to 3 degrees in magnitude. Interestingly, there were very few apparent movements with a magnitude of 10 degrees and no blinking artefacts noted in the present cohort. Overall, most patients had a propensity for few or small magnitude eye movements, and this is reflected in the ensuing results.

Analysis 1: The influence of patient-specific factors on gaze tracker metrics

Comparison of the distribution of aggregate eye movement metrics based on the eye and test order variables is shown in [Figure 3](#). The distributions showed no significant difference from zero for nearly all conditions, except between the first and second tests within the same eye and the sum of amplitudes, which was statistically ($p = 0.04$) but not clinically significant (there was an average difference of 9.3 degrees across the entirety of the test in test 1 compared to test 2).

Analysis of the Spearman correlations between gaze tracker metrics and age and mean deviation are shown in [Figure 4](#) (note that the correlation line is not shown for mean deviation, b; see further below). For all conditions, the correlation was significant at the $p < 0.0001$ level. With increasing age ([Figure 4a](#)), there was an increase in the number of eye movements during the test. However, in all cases the correlations were weak. Notably, the correlation between the mean deviation and the number of ticks showed a very wide distribution of number of ticks at near normal levels of mean deviation (-2 dB or better). Nominal bracketed correlations (-2 dB or better, -2 dB to -6 dB or -6 dB to -12 dB) showed no correlations between the number of ticks and mean deviation (all $p > 0.05$). Despite a tendency towards more ticks with worsening mean deviation, the correlation across the entire cohort appeared to be driven primarily by the outlier points at more severe mean deviation levels, and thus the correlation line is not shown in [Figure 4b](#). There remained substantial spread of data across all bins of mean deviation level. Using the other aggregate indices (sum of amplitudes and average amplitude), as defined in [Figure 1](#), showed similarly weak correlations. Therefore, although these reflected

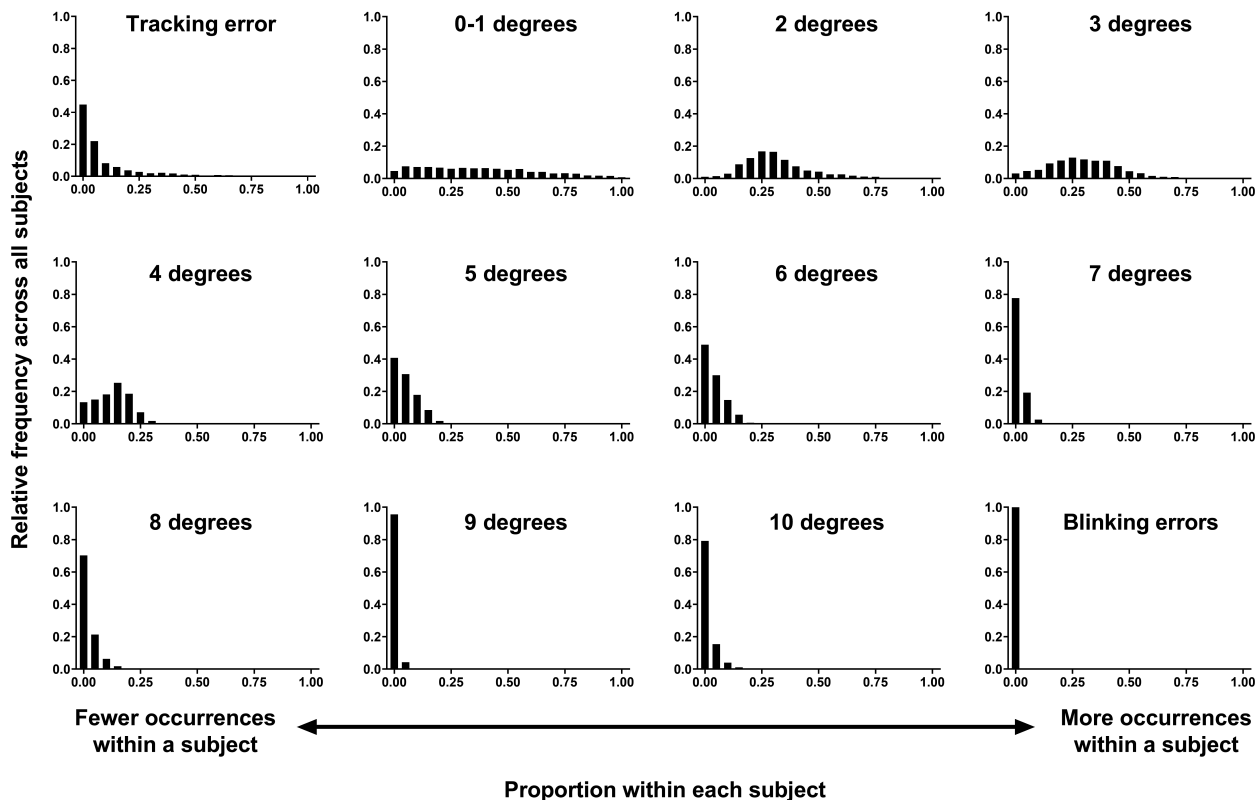


FIGURE 2 Frequency distributions (proportion across all subjects, y-axis) of difference eye tracker outcomes as a function of different proportions of occurrences of those outcomes within each subject (x-axis). A higher y-axis value indicates more instances across all subjects, and a higher x-axis value indicates more instances within a subject.

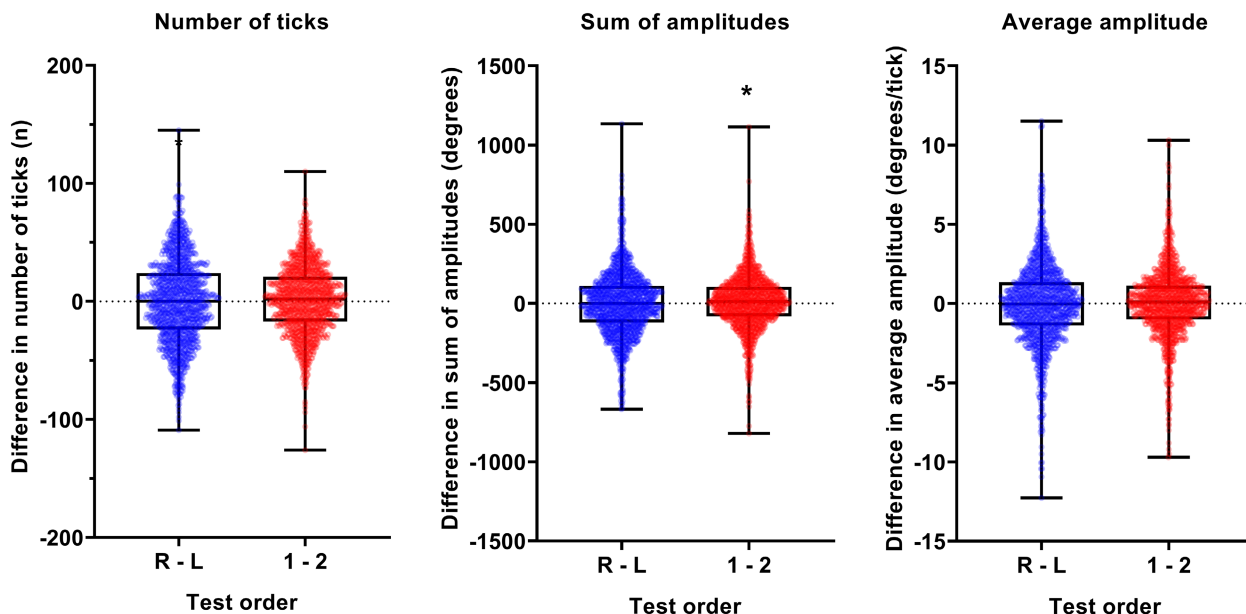


FIGURE 3 Difference in number of ticks (left), sum of amplitudes (middle) and average amplitude (right) when examining test order (right—Left, blue; test 1—Test 2, red). The black dotted line indicates no difference. The datum points indicate the result from one subject, and the box and whiskers indicate the median, interquartile range and full range of values.

amplitude of movements during the test, they are unlikely to contribute to clinical interpretation and thus are not shown for clarity.

Overall, the assessed independent variables either did not show or demonstrated only very weak relationships with the aggregate gaze tracker metrics. As such, we did

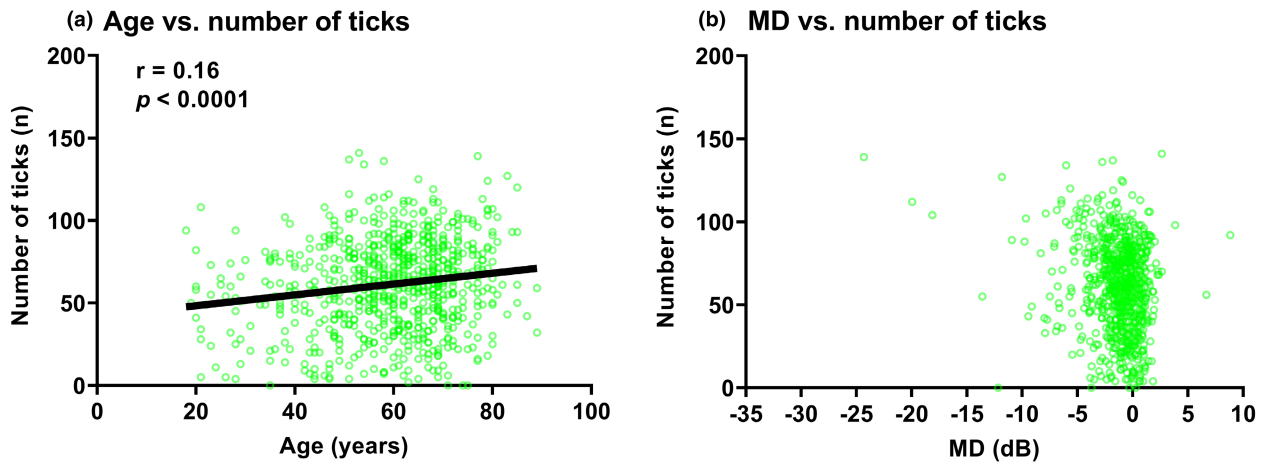


FIGURE 4 Correlations between number of ticks and age (years, (a)) and mean deviation (MD) (dB, (b)). The datum points indicate the result from one of eye of one subject, and the black solid line indicates the result of the correlation analysis. The inset values are the correlation coefficient (r) and p -value. As mentioned in the text, due to the wide distribution of data points in (b), the correlation results are not shown.

not incorporate these variables into models examining the relationship between gaze tracker metrics and output sensitivity results.

Analysis 2: The relationship between gaze tracker metrics and sensitivity measurements

The first correlations performed were between the difference in mean deviation, “events” and the overall average hill of vision as a function of difference in aggregate eye movements (Figure 5). There were no statistically significant correlations between eye movement parameters and key visual field outputs.

The second Spearman correlation analysis performed was between the root mean squared error (sensitivity and total deviation) and the absolute difference in aggregate eye movements (Figure 6). Similar to the results with key visual field output metrics, there were no significant correlations between eye movement parameters and root mean squared error on intra-session visual field results. Again, there was no difference in the output correlations between the number of ticks and the other aggregate metrics, and thus, sum of amplitude and average amplitude results are not shown for clarity.

Quartile analysis

We performed the correlation analysis after dividing the visual field sensitivity, total deviation and gaze tracker results into the approximate first and fourth quartiles of the test. Correlations were performed on the number of “events,” root mean squared error of sensitivity and total deviation values (mean deviation and average hill of vision calculations would not be meaningful on this subset of points).

Descriptive statistics showed that there tended to be more eye movements (mean 0.5, $p < 0.0001$) and greater

sum of amplitudes (mean 4.6 degrees, $p < 0.0001$) made in the fourth quarter of the test compared to the first quarter, but the magnitude of difference indicated little clinical significance. There was no difference in the average eye movement between first and last quartiles ($p = 0.32$). These results are summarised in Figure S1.

The correlations performed on “event” analysis for quartiles 1 and 4 showed similar results to those reported above when the entirety of the gaze tracker output was analysed, with no apparent correlations between eye movements and “events” (Figure S2). The lack of correlations with root mean squared error for sensitivity and total deviation values for quartiles 1 (Figure S3) and 4 (Figure S4) were also similar to that when the entirety of the gaze tracker output was analysed.

DISCUSSION

In the present study, we aimed to identify correlations between SITA-Faster 24–2 gaze tracker results and patient-related independent variables and resultant sensitivity outputs. The specific interest in gaze tracker outputs, despite the contention surrounding their interpretation,¹⁹ arose due to the recent recommendations for interpreting reliability in SITA-Faster.⁸ None of the independent factors nor sensitivity measures appeared to have clinically meaningful correlations with gaze tracker aggregate metrics. Inspection of the data from the present study showed wide variation in gaze tracker output data. Even large differences in aggregate eye movement measurements were associated with only very small changes in visual field outputs such as events detected or the average height of the hill of vision (when defined as the average of the top seven sensitivity values across the visual field).

Overall, there was little apparent contribution of gaze tracker metrics to the output sensitivity results. The present study focussed on the associations between gaze

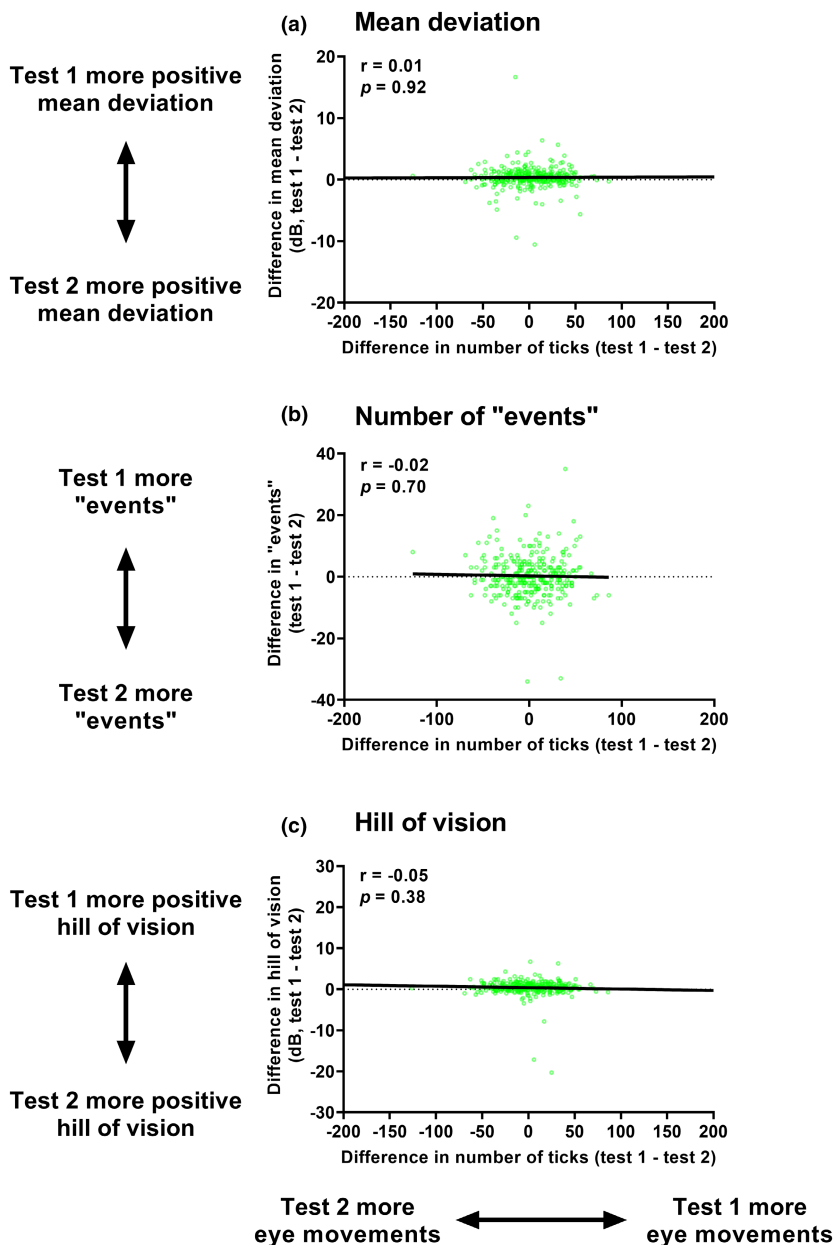


FIGURE 5 Correlations between the difference in mean deviation (dB, top row (a)), number of “events” (n , middle row (b)), and overall average hill of vision (dB, bottom row (c)) and number of ticks. All differences are test 1—Test 2. The datum points indicate the result from one of eye of one subject, and the black solid line indicates the result of the correlation analysis. The inset values are the correlation coefficient (r) and p -value.

tracker results and visual field metrics in frontloaded visual field results. However, one might infer that, due to generally strong correlations between frontloaded visual field tests, the overall variation in gaze tracker deviations means little in the clinical interpretation of perimetric results.

Factors affecting gaze tracker outputs—Test order

Test results after the first are typically predicted to return more reliable data, overcoming issues related to

procedural learning. Due to the relatively short test length of SITA-Faster, fatigue is less likely to play a role in reliability in comparison to a longer test like SITA-Standard. Thus, a prediction was that, under normal perimetric conditions using a short algorithm like SITA-Faster, test 2, left eye, and quartile 4 results would return fewer overall eye movements. Our present results showed, at times, statistically significant differences in line with this prediction, but the overall association of test order with differences in gaze tracker outputs was small. Thus, test order did not seem to have a clinically meaningful association with resultant eye movement data.

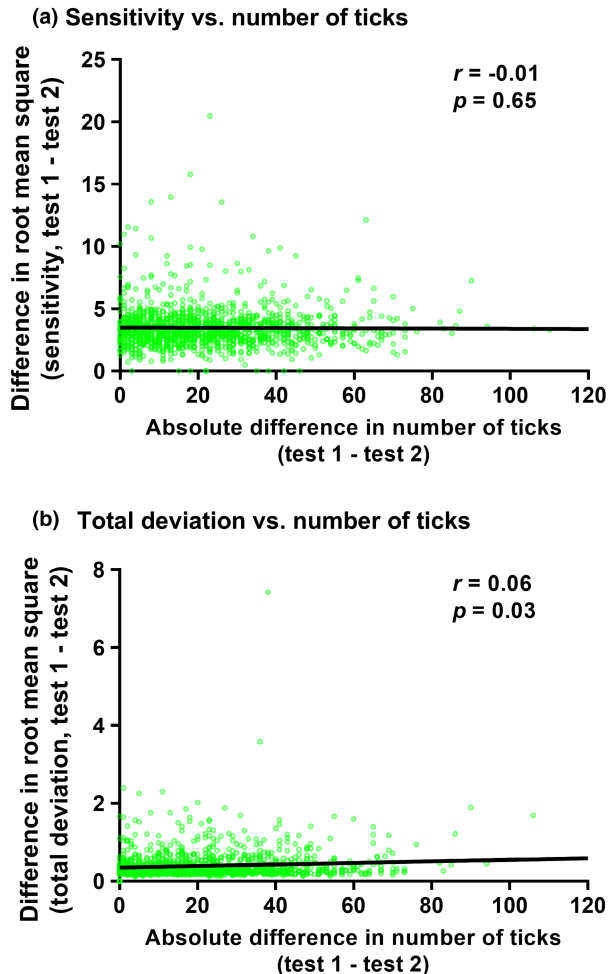


FIGURE 6 Correlations between the difference in root mean squared error in sensitivity (top row (a)) and total deviation (bottom row (b)) with number of ticks. All differences are test 1—Test 2. The datum points indicate the result from one of eye of one subject, and the black solid line indicates the result of the correlation analysis. The inset values are the correlation coefficient (r) and p -value. Note that unlike in Figure 5, the differences have been converted to absolute values.

Factors associated with gaze tracker outputs—Age, false positive rate and visual field mean deviation

We found only weak correlations between age and visual field mean deviation—but not false positive rate—and the aggregate gaze tracker metrics. The correlation found with age may be multifactorial, with contributions from psychomotor, age-related neural and sensitivity decline, and cognitive factors such as attention having been described in the literature in reference to fixation stability and inhibition of eye movements.^{37–39} The correlation found with mean deviation was expected, as previous studies have shown fixation instability to be more prevalent in patients with glaucoma, even in its early stages, and worsening with greater vision loss.^{40,41} However, the weak correlations suggest little clinical relevance using this description of gaze data.

The association of gaze tracker deviations with intrasession visual field sensitivity outputs

The direction of eye movements deviating from fixation may be important for eliciting its anticipated association with visual field parameters. Foveation towards the stimulus presentation location is expected to increase sensitivity (and thus, a more positive mean deviation, fewer “events” and a higher average hill of vision), whilst movements away from the stimulus expectedly lead to the opposite associations. Our results suggested a trend with more eye movements leading to the following: a less positive mean deviation, more “events” and a more positive average hill of vision when comparing intrasession tests. The present results therefore did not appear to fit any prediction and may be indicative of the randomness of eye movements. This remained a product of the scalar output, and techniques that employ vector quantities may still be beneficial for reducing result variability and increasing fidelity.

Implications for clinical practice

Our results differed slightly from previous reports^{13–15} that have described the potential usefulness of incorporating gaze tracker metrics into visual field interpretation. There have been concerns regarding eye movements impairing accurate progression analysis, as sensitivity values across tests may therefore not be correlated.¹⁵ Interestingly, our results found no correlation between intrasession variability and gaze tracker metrics, contrary to these previous reports.¹⁹ Our study design was different as we compared tests within the same session (potentially overcoming issues such as learning). However, a fundamental difference is the test algorithm and therefore test length. As SITA-Faster is a shorter test compared to SITA-Standard, patients may be less likely to return excessive eye movements due to fatigue and inattention. Further longitudinal studies using SITA-Faster would be required to determine the usefulness of the gaze tracker output for progression analysis.

Although most gaze deviations in our patients were small, with many returning deviations fewer than 3 degrees unlike the results of Demirel and Vingrys,¹⁹ the relative associations with key output metrics were generally aligned with their report. Specifically, our results demonstrated generally weak correlations between aggregate gaze deviations and important perimetric outcomes, such as mean deviation, “events” and the overall average hill of vision (as defined by the average of the top seven sensitivity results).

A recent study by Camp et al.¹⁶ suggested that gaze tracker and false positive rates assess different aspects of result reliability due to their poor correlations. Our findings provide support for the poor correlations between these “reliability” metrics. Additionally, neither metric, both in the work of Camp et al.¹⁶ and within previous reports,⁹ has

been shown to have strong correlations with resultant sensitivity measurements when using SITA-Faster. Therefore, the poor correlations may not simply be reflective of different aspects of reliability, but rather that neither are fully reflective of the usability of the visual field result.

The association of different magnitudes of eye movements can be estimated by understanding the shape of the hill of vision across the 24–2 test grid. Using a Goldmann size III target, the difference in sensitivity between the fovea and the 3 degree eccentricity ring is approximately 3 dB.³¹ Thus, the difference between 3, 9 and 15 degree rings are relatively small (<1 dB) before a larger difference occurs at 21 degrees or beyond (represented by similar sensitivity isocontours).³¹ Therefore, the change in sensitivity would be greatest with very large eye movements of 21 degrees or greater. Such changes in the shape of the hill of vision should be obvious upon inspection of the sensitivity map.

Practical constraints of current gaze tracker outputs

As described above, it is impossible to predict the potential effect of gaze deviations due to the current manner in which gaze tracker deviations are reported. The emergence of fundus-tracking perimetry has allowed for measurement of visual field sensitivity whilst accounting for eye movements. Reports in the literature on the variability of sensitivity measurements from fundus-tracking and static automated perimetry have been mixed,^{42,43} with little clinically meaningful improvement in repeatability by using fundus tracking. The notable advantage of accounting for eye movements is related to potentially improving structure–function relationships, but again, reports have been mixed.^{44–46} Given small differences between fixation-compensating or tracked perimetric techniques and non-tracked methods, the clinical value of gaze data remains a complementary method for assessing visual field reliability and integrity.

An alternative strategy that targets lapses in attention and other sources of test noncompliance uses automated feedback if significant eye movement is detected. Currently, the Humphrey Field Analyser provides an audio cue indicating significant eye or head tracking errors. However, specific feedback on how to reduce gaze deviations is reliant upon the technician. Additional biometric measurements aside from head and eye deviations have been proposed by Jones and colleagues⁴⁷ for identifying sources of measurement error. Even so, Jones and colleagues⁴⁷ acknowledge that patient-derived biomarkers and metrics of “reliability” account for little of the measurement error associated with perimetry.

Limitations

Although we used a large sample of consecutively examined patients, most of the gaze deviations were small, and

there was an under-representation of large magnitude errors. Thus, we did not perform correlations with individual ticks and instead chose to analyse aggregate measures. Similarly, prior studies have examined nominal groups of deviations (for example, less than 2 degrees, 3–5 degrees, 6+ degrees and others), but given the results of the factor analysis (see Appendix S1) and the distribution of gaze deviation magnitude, we only reported aggregate metrics. Whilst this skew in the data could have affected the resultant correlation analyses, this distribution of gaze data is representative of the clinical population examined in the present study. Furthermore, although there was a risk of introducing a type I error with multiple separate correlation analyses, all resultant correlations were weak, and thus further analyses confirming significance were not required.

A fundamental assumption made in the present study was regarding the accuracy and precision of the output gaze tracker data for representing a patient's gaze during the test. The present study was not designed to assess the precision of eye traces using the Humphrey Field Analyser; however, the output metrics remain representative of a clinician-facing, supposedly interpretable metric.

There was also an under-representation of subjects with more advanced visual field loss. This was a product of the clinic from which the subjects were sampled. Whilst the distribution of mean deviation values is diverse, targeted examination of subjects with more advanced loss may strengthen the analyses involving mean deviation as an independent variable.

The present study focussed on the 24–2 test grid, and thus the gaze tracker outputs and correlations would only apply under conditions where the test locations are spaced 6 degrees apart. In the 10–2 grid where the locations are spaced 2 degrees apart, it is possible that an equivalent magnitude of eye movements may be more significantly associated with sensitivity measurements.

CONCLUSIONS

Aggregate quantitative metrics do not appear to provide clinically meaningful information in the interpretation of intrasession sensitivity metrics in perimetry conducted using the SITA-Faster algorithm and the 24–2 test grid. Instead, an approach when considering other facets of test repeatability—the comparison of sensitivity results in combination with classical “reliability” metrics—is recommended.

ACKNOWLEDGEMENTS

The work was supported, in part, by an NHMRC Ideas Grant to MK and JP (1186915), and a University of New South Wales Science Early Career Academic Network Seeding Grant to JP. Guide Dogs NSW/ACT provided funding for the clinical services enabling data collection for this study. JP and MK receive salary support from Guide Dogs NSW/ACT. Guide

Dogs NSW/ACT also provides salary support for JP and MK and support for clinical service delivery at the Centre for Eye Health, from which the clinical data was derived. The funding body had no role in the conception or design of the study. Open access publishing facilitated by University of New South Wales, as part of the Wiley - University of New South Wales agreement via the Council of Australian University Librarians. [Correction added on 04-July-2022, after first online publication: CAUL funding statement has been added.]

AUTHOR CONTRIBUTIONS

Jack Phu: Conceptualization (lead); data curation (lead); formal analysis (lead); funding acquisition (supporting); investigation (lead); methodology (lead); visualization (equal); writing – original draft (lead); writing – review and editing (equal). **Michael Kalloniatis:** Formal analysis (supporting); funding acquisition (lead); methodology (supporting); resources (lead); visualization (equal); writing – review and editing (equal).

CONFLICT OF INTEREST

No conflicting relationship exists for any author.

ORCID

Jack Phu  <https://orcid.org/0000-0002-9933-6780>

Michael Kalloniatis  <https://orcid.org/0000-0002-5264-4639>

REFERENCES

- Stewart WC, Hunt HH. Threshold variation in automated perimetry. *Surv Ophthalmol.* 1993;37:353–61.
- Asman P, Fingeret M, Robin A, Wild J, Pacey I, Greenfield D, et al. Kinetic and static fixation methods in automated threshold perimetry. *J Glaucoma.* 1999;8:290–6.
- Frankhauser F, Spahr J, Bebie H. Some aspects of the automation of perimetry. *Surv Ophthalmol.* 1977;22:131–41.
- Phu J, Kalloniatis M. A strategy for seeding point error assessment for retesting (SPEAR) in perimetry applied to normal subjects, glaucoma suspects, and patients with glaucoma. *Am J Ophthalmol.* 2021;221:115–30.
- 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;78:519–22.
- Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Invest Ophthalmol Vis Sci.* 2000;41:2201–4.
- Heijl A, Patella VM, Chong LX, Iwase A, Leung CK, Tuulonen A, et al. A new SITA perimetric threshold testing algorithm: construction and a multicenter clinical study. *Am J Ophthalmol.* 2019;198:154–65.
- Heijl A, Patella VM, Bengtsson B. 5. Statpac analysis of single fields. Excellent perimetry – the field analyzer primer. Dublin, CA: Carl Zeiss Meditec; 2021. p. 79–96.
- Heijl A, Patella VM, Flanagan JG, Iwase A, Leung CK, Tuulonen A, et al. False positive responses in standard automated perimetry. *Am J Ophthalmol.* 2021;233:180–8.
- 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;208:251–64.
- Phu J, Kalloniatis M. The frontloading fields study: the impact of false positives and seeding point errors on visual field reliability when using SITA-faster. *Transl Vis Sci Technol.* 2022;11:20. <https://doi.org/10.1167/tvst.11.2.20>
- 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;75:181–3.
- Ishiyama Y, Murata H, Hirasawa H, Asaoka R. Estimating the usefulness of Humphrey perimetry gaze tracking for evaluating structure-function relationship in glaucoma. *Invest Ophthalmol Vis Sci.* 2015;56:7801–5.
- Ishiyama Y, Murata H, Mayama C, Asaoka R. An objective evaluation of gaze tracking in Humphrey perimetry and the relation with the reproducibility of visual fields: a pilot study in glaucoma. *Invest Ophthalmol Vis Sci.* 2014;55:8149–52.
- Asaoka R, Fujino Y, Aoki S, Matsuura M, Murata H. Estimating the reliability of glaucomatous visual field for the accurate assessment of progression using the gaze-tracking and reliability indices. *Ophthalmol Glaucoma.* 2019;2:111–9.
- Camp AS, Long CP, Patella VM, Proudfoot JA, Weinreb RN. Standard reliability and gaze tracking metrics in glaucoma and glaucoma suspects. *Am J Ophthalmol.* 2021;234:91–8.
- Vingrys AJ, Demirel S. False-response monitoring during automated perimetry. *Optom Vis Sci.* 1998;75:513–7.
- Newkirk MR, Gardiner SK, Demirel S, Johnson CA. Assessment of false positives with the Humphrey field analyzer II perimeter with the SITA algorithm. *Invest Ophthalmol Vis Sci.* 2006;47:4632–7.
- Demirel S, Vingrys AJ. Eye movements during perimetry and the effect that fixational instability has on perimetric outcomes. *J Glaucoma.* 1994;3:28–35.
- Demirel S, Johnson LN, Fendrich R, Vingrys AJ. The slope of frequency-of-seeing curves in normal, amblyopic and pathologic vision. JOSA Technical Digest Series; 1997. p. 244–7.
- Kimura T, Matsumoto C, Nomoto H. Comparison of head-mounted perimeter (Imo([R])) and Humphrey field analyzer. *Clin Ophthalmol.* 2019;13:501–13.
- Phu J, Kalloniatis M. Viability of performing multiple 24-2 visual field examinations at the same clinical visit: the frontloading fields study (FFS). *Am J Ophthalmol.* 2021;230:48–59.
- Phu J, Kalloniatis M. The frontloading fields study (FFS): detecting changes in mean deviation in glaucoma using multiple visual field tests per clinical visit. *Transl Vis Sci Technol.* 2021;10:21. <https://doi.org/10.1167/tvst.10.13.21>
- Prum BE Jr, Rosenberg LF, Gedde SJ, Mansberger SL, Stein JD, Moroi SE, et al. Primary open-angle glaucoma preferred practice pattern([R]) guidelines. *Ophthalmology.* 2016;123:P41–111.
- National Health and Medical Research Council. Guidelines for the screening, prognosis, diagnosis, management and prevention of glaucoma. Internet: Commonwealth of Australia; 2010.
- Wang H, Kalloniatis M. Clinical outcomes of the Centre for eye Health: an intra-professional optometry-led collaborative eye care clinic in Australia. *Clin Exp Optom.* 2021;104:795–804.
- Heijl A, Patella VM, Bengtsson B. 2. Review of basic principles. Excellent perimetry – the field analyzer primer. Dublin, CA: Carl Zeiss Meditec; 2021. p. 13–42.
- Kunimatsu S, Suzuki Y, Shirato S, Araie M. Usefulness of gaze tracking during perimetry in glaucomatous eyes. *Jpn J Ophthalmol.* 2000;44:190–1.
- Phu J, Khuu SK, Agar A, Domadiou I, Ng A, Kalloniatis M. Visualizing the consistency of clinical characteristics that distinguish healthy persons, glaucoma suspect patients, and manifest glaucoma patients. *Ophthalmol Glaucoma.* 2020;3:274–87.
- Josan AS, Buckley TMW, Wood LJ, Jolly JK, Cehajic-Kapetanovic J, MacLaren RE. Microperimetry hill of vision and volumetric measures of retinal sensitivity. *Transl Vis Sci Technol.* 2021;10:12. <https://doi.org/10.1167/tvst.10.7.12>

31. Phu J, Khuu SK, Nivison-Smith L, Zangerl B, Choi AYJ, Jones BW, et al. Pattern recognition analysis reveals unique contrast sensitivity isocontours using static perimetry thresholds across the visual field. *Invest Ophthalmol Vis Sci.* 2017;58:4863–76.
32. Traquair HM. An introduction to clinical perimetry. St Louis: The C V Mosby Company; 1949.
33. Weleber RG, Smith TB, Peters D, Chegarnov EN, Gillespie SP, Francis PJ, et al. VFMA: topographic analysis of sensitivity data from full-field static perimetry. *Transl Vis Sci Technol.* 2015;4:14. <https://doi.org/10.1167/tvst.4.2.14>
34. Heijl A, Lindgren G, Olsson J. A package for the statistical analysis of visual fields. *Doc Ophthalmol Proc Ser.* 1987;49:153–68.
35. Heijl A, Lindgren G, Olsson J, Asman P. Visual field interpretation with empiric probability maps. *Arch Ophthalmol.* 1989;107:204–8.
36. 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;43:2654–9.
37. Fragiotta S, Carnevale C, Cutini A, Rigoni E, Grenga PL, Vingolo EM. Factors influencing fixation stability area: a comparison of two methods of recording. *Optom Vis Sci.* 2018;95:384–90.
38. Morales MU, Saker S, Wilde C, Pellizzari C, Pallikaris A, Notaroberto N, et al. Reference clinical database for fixation stability metrics in normal subjects measured with the MAIA Microperimeter. *Transl Vis Sci Technol.* 2016;5:6. <https://doi.org/10.1167/tvst.5.6.6>
39. Plomecka MB, Baranczuk-Turska Z, Pfeiffer C, Langer N. Aging effects and test-retest reliability of inhibitory control for saccadic eye movements. *eNeuro.* 2020;7:ENEURO.0459-19.2020. <https://doi.org/10.1523/ENEURO.0459-19.2020>
40. Shi Y, Liu M, Wang X, Zhang C, Huang P. Fixation behavior in primary open angle glaucoma at early and moderate stage assessed by the MicroPerimeter MP-1. *J Glaucoma.* 2013;22:169–73.
41. Kameda T, Tanabe T, Hangai M, Ojima T, Aikawa H, Yoshimura N. Fixation behavior in advanced stage glaucoma assessed by the MicroPerimeter MP-1. *Jpn J Ophthalmol.* 2009;53:580–7.
42. Matsuura M, Murata H, Fujino Y, Hirasawa K, Yanagisawa M, Asaoka R. Evaluating the usefulness of MP-3 microperimetry in glaucoma patients. *Am J Ophthalmol.* 2018;187:1–9.
43. Montesano G, Bryan SR, Crabb DP, Fogagnolo P, Oddone F, McKendrick AM, et al. A comparison between the compass fundus perimeter and the Humphrey field analyzer. *Ophthalmology.* 2019;126:242–51.
44. Shin JW, Song MK, Won HJ, Jo Y, Kook MS. Comparison of the structure-function relationship between compass microperimetry and Humphrey field analyzer in myopic open-angle glaucoma eyes. *Br J Ophthalmol.* 2020;106:485–90.
45. Rao HL, Januwada M, Hussain RS, Pillutla LN, Begum VU, Chaitanya A, et al. Comparing the structure-function relationship at the macula with standard automated perimetry and microperimetry. *Invest Ophthalmol Vis Sci.* 2015;56:8063–8.
46. Tepelus TC, Song S, Nittala MG, Nassisi M, Sadda SVR, Chopra V. Comparison and correlation of retinal sensitivity between microperimetry and standard automated perimetry in low-tension glaucoma. *J Glaucoma.* 2020;29:975–80.
47. Jones PR, Demaria G, Tigchelaar I, Asfaw DS, Edgar DF, Campbell P, et al. The human touch: using a webcam to autonomously monitor compliance during visual field assessments. *Transl Vis Sci Technol.* 2020;9:31. <https://doi.org/10.1167/tvst.9.8.31>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Phu J, Kalloniatis M. Gaze tracker parameters have little association with visual field metrics of intrasession frontloaded SITA-Faster 24–2 visual field results. *Ophthalmic Physiol Opt.* 2022;42:973–985. <https://doi.org/10.1111/opo.13006>