Quantifying Gaze Tracking for Visual Field Reliability Estimation. A Systematic Review

Peña-Pulgar Luisa Fernanda[1,3], Tello Alejandro[2,3,4,5], Acuña María Fernanda[2,3,5], Galvis Virgilio[2,3,5]

[1]Universidad del Rosario, Bogotá, Colombia
[2]Centro Oftalmológico Virgilio Gálvis, Floridablanca, Colombia
[3]Universidad Autónoma de Bucaramanga (UNAB), Bucaramanga, Colombia
[4]Universidad Industrial de Santander (UIS), Bucaramanga, Colombia
[5]Fundación Oftalmológica de Santander FOSCAL, Floridablanca, Colombia

Submitted to the editorial board: April 25, 2025
Accepted for publication: July 2, 2025
Available on-line: September 4, 2025

The authors have not received funding for this research from any public, commercial, or non-profit organizations. Article has not been simultaneously submitted to another journal, and is not printed elsewhere.

First author:
Luisa Fernanda Peña-Pulgar, MD

Corresponding author:
Alejandro Tello, MD
Office 301, Tower C, 3rd Floor, Cl. 157 #20-94
Floridablanca, Santander
Colombia
E-mail: alejandrotello@gmail.com

SUMMARY

Aims: This systematic review aimed to investigate whether quantitative metrics derived from gaze tracking (GT) outputs during visual field (VF) testing with an automated perimeter could enhance the evaluation of test reliability.

Materials and Methods: A systematic search of PubMed, Cochrane, LILACS, and IBECS databases, from inception to August 31, 2024, was conducted.

Results: Eight studies – four cross-sectional and four cohort – met the inclusion criteria, comprising 8,181 visual field tests from 3,687 patients. The studies were categorized based on testing strategy: SITA Standard, Fast, and Faster. In the SITA Standard group, GT parameters were associated with visual field result reproducibility and the structure-function relationship in glaucoma, but were influenced by ocular surface variables. In the SITA Fast and Faster group, results were mixed: some studies suggested GT metrics could complement conventional reliability parameters, while others concluded that GT quantitative metrics did not offer clinically meaningful insights beyond existing methods.

Conclusion: GT trace quantification shows promise as an objective reliability parameter for VF testing, particularly within the SITA Standard framework. Advanced image analysis techniques, including artificial intelligence, could facilitate automated GT parameter quantification, streamlining processes and supporting further studies to evaluate their impact on VF data reliability and clinical decision-making.

Key words: glaucoma, visual field, automated perimetry, gaze tracking

Čes. a slov. Oftal., 82, 2026, No. 1, p. 3–11

INTRODUCTION

Chronic glaucoma is the leading cause of irreversible blindness worldwide, affecting over 76 million people [1]. This chronically progressive optic neuropathy primarily affects retinal ganglion cells, whose axons form the optic nerve. It typically begins with asymptomatic peripheral loss in the visual field (VF), which often delays diagnosis, and may eventually impair central vision, significantly affecting quality of life. Beyond its personal impact, glaucoma represents a considerable economic and social challenge for healthcare systems globally. Despite advancements in diagnosis and treatment, early detection remains challenging, with an estimated 50-80% of cases in developed countries going undiagnosed [2,3].

Proper evaluation, classification, and management of glaucoma rely on detecting structural changes in the optic nerve, typically accompanied by corresponding functional impairments. Testing is essential for assessing functional damages. It provides key information on the extent of vision loss and its progression, the results of which are crucial for determining disease severity and treatment. Among the available devices, the Humphrey Field Analyzer (HFA; Carl Zeiss Meditec AG, Jena, Germany) is one of the most widely used due to its reliability and comprehensive set of features [4]. The HFA offers several reliability indices, such as false positives (FP), false negatives (FN), and fixation losses (FL). These are determined using the Heijl-Krakau method, by periodically presenting stimuli within the previously mapped blind spot [4,5]. Such indices help clinicians assess the consistency and accuracy of VF results.

Currently, perimetry tests in the HFA mostly use the Swedish Interactive Threshold Algorithm (SITA), a family of strategies used to significantly reduce testing time while maintaining or improving accuracy, compared to older methods. SITA includes various test patterns: 30-2, with 76 points spaced 6° apart covering 30° of the central field; 24-2, with 54 points also spaced 6° apart covering 24°; and 10-2 with a denser grid of points every 2° focused on the central 10°. The threshold strategies – SITA Standard, SITA Fast, and SITA Faster – offer progressively shorter test times, with SITA Faster enabling reliable results in 2-4 minutes [6].

For nearly 25 years, the HFA has included a gaze tracking (GT) functionality, which records deviations in the patient's gaze during testing – each time a stimulus appears – and displays them on the gaze tracking plot at the bottom of the VF printout. This chart consists of an irregular horizontal line with consecutive – often adjacent – vertical bars. Upward bars indicate instances where the gaze shifted outside the fixation target, with the largest bars reflecting deviations of ≥ 10°. While the size of the upward bar correlates with the amount of gaze deviation, it does not indicate the deviation's orientation relative to the fixation target, which can be in any direction. Downward bars represent cases where the gaze direction could not be determined (small bars), or interference caused by the eyelid, such as when the patient squints or blinks (large bars). The gaze fixation chart provides an overview of fixation stability throughout the examination, where larger or more frequent deviations suggest poorer fixation reliability. However, the GT functionality does not automatically generate numerical metrics. Consequently, interpreting GT data has traditionally relied on the subjective assessment of the examiner [6]. (Figure 1).

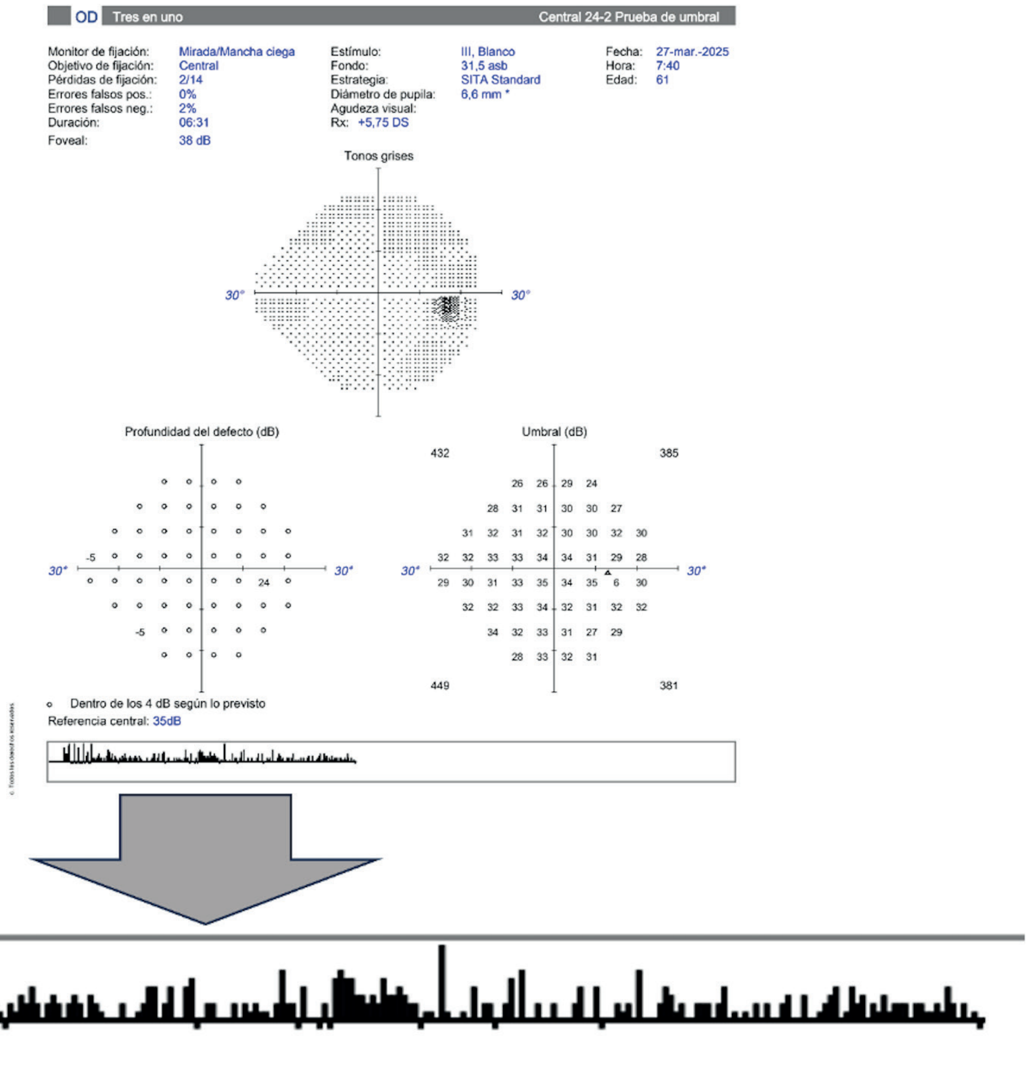


Figure 1. Gaze Tracking Trace from a Humphrey Visual Field Test (HFA 24-2).
This figure illustrates the gaze tracking (GT) record from a standard automated perimetry test performed with the Humphrey Field Analyzer (HFA). The GT graph, located in the lower section of the visual field printout (top panel), has been magnified in the bottom panel for better visualization. Each upward vertical bar represents a deviation of the patient's gaze from the fixation target at the moment a visual stimulus was presented, with longer bars indicating greater deviation (≥ 10°). The horizontal axis reflects the sequence of stimuli. Downward bars correspond to undetectable gaze positions (short bars) or interference caused by eyelid movements, such as blinking (tall bars). While this trace provides valuable insight into fixation stability during the test, it does not offer numerical outputs and must be interpreted subjectively by the examiner

While the gaze-tracking feature has shown potential to enhance the interpretation of VF results by offering additional insights into test reliability, its lack of quantitative metrics limits its full utility in clinical research and practice. Some research groups have been working on addressing this issue for over two decades [4,5,7-10]. Developing a standardized framework for utilizing quantitative metrics from GT data could potentially further improve the diagnostic accuracy and management of glaucoma. For this reason, the present study aims to review the current published knowledge on using values derived from GT graphical records in the HFA as a reliability index for VF testing.

MATERIAL AND METHODS

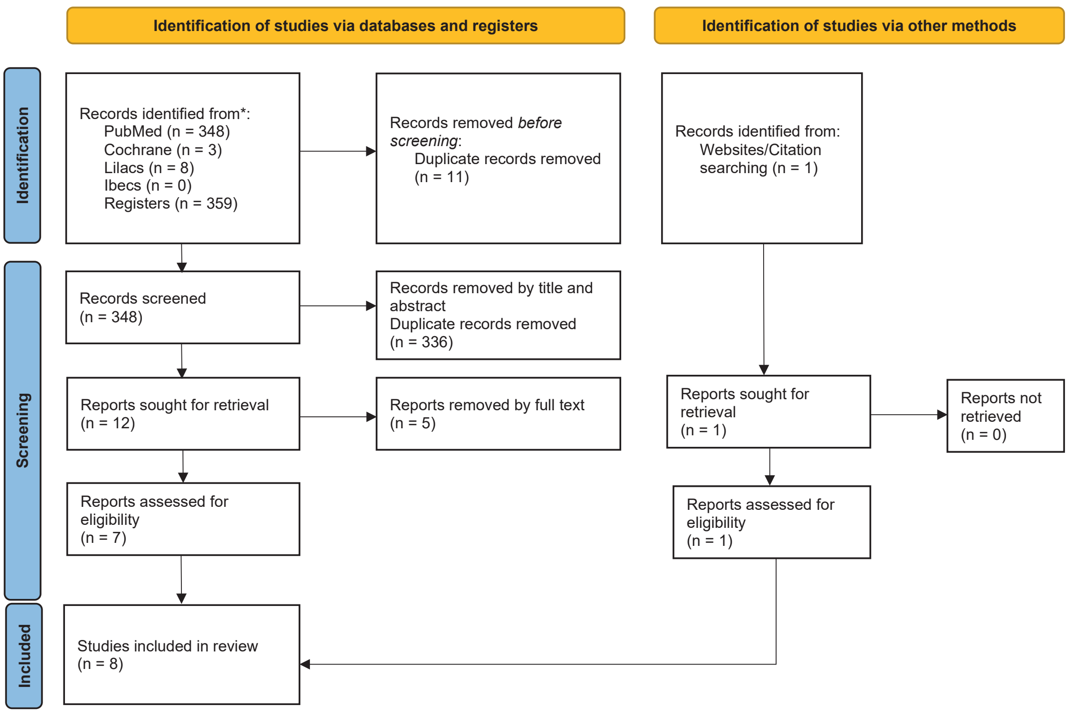
For the purposes of this article, a systematic search was conducted in PubMed, Cochrane, Lilacs, and Ibecs, using the following terms in both English and Spanish: ("Gaze tracking" / "Seguimiento de la mirada" [All fields] OR "Eyetracking" / "Seguimiento ocular" [All fields]) AND ("Visual Field" / "Campo visual" [All fields] OR "Campimetry" / "Campimetría" [All fields] OR "Perimetry" / "Perimetría" [All fields]) AND ("Glaucoma" [All fields]). Studies were included based on the following criteria: primary studies published in English or Spanish involving patients diagnosed with glaucoma, studies that employed the HFA, and those with patients of all ethnicities, ages, and sexes. Exclusion criteria included secondary and tertiary studies, studies involving patients with ocular conditions other than glaucoma or using visual field analyzers other than the HFA, animal studies, and studies that did not meet the inclusion criteria.

The search was restricted to studies published until August 31, 2024. Results were initially filtered by title and abstract; eligible studies then underwent full-text review. The selection process was conducted by independent reviewers, with disagreements resolved through consensus. Finally, a narrative synthesis of the identified studies was conducted. Considering that secondary data were used and no interventions or modifications were made to the individuals' biological, physiological, psychological, or social variables, this research is classified as low risk and did not require approval from an ethics committee or informed consent from patients. In addition, the intellectual property rights of the authors of the primary studies were ensured.

RESULTS

Search Results

The initial search retrieved 359 articles. After removing duplicates and filtering the articles by title, abstract, and full text, 7 articles were selected that met the inclusion and exclusion criteria. Finally, an additional article was added through snowball sampling, bringing the total to 8 studies included [11]. (Figure 2).



From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

Figure 2. PRISMA 2020 flow diagram of study selection.
Flowchart illustrating the identification, screening, eligibility, and inclusion of studies in the systematic review. A total of 359 records were identified from databases and registers, and 1 from other sources. After removing duplicates and screening titles and abstracts, 12 full-text reports were assessed, and 8 studies were ultimately included in the review

Demographics and Clinical Characteristics

Table 1 shows the demographic and clinical characteristics of the included studies, and Table 2 summarizes their main findings

The 8 studies were conducted between 2014 and 2022, 5 of them in Japan, 2 in the United States, and 1 in Australia. Four were cross-sectional studies and 4 cohort studies. Three included patients with glaucoma and suspected glaucoma, and 5 included only patients with open-angle glaucoma. Among the included studies, 1 used both the 24-2 and 10-2 SITA Standard protocols, 2 employed both 24-2 and 30-2, one used only 30-2, and 4 studies applied exclusively the 24-2 protocol. In addition, some studies utilized SITA Fast and SITA Faster 24-2 strategies [6].

Studies on GT Metrics with SITA Standard Testing Strategies

In 2014, in their pilot study, Ishiyama et al. proposed a novel method to manually quantify parameters from GT traces from 24-2 or 10-2 SITA standard protocols [5]. They included 42 eyes from 42 glaucoma patients. Their analysis focused on measuring the frequency of upward and downward bars of varying lengths in the GT records. They also analyzed the average frequency of tracking failures per stimulus (TFF) and the average frequency of eye movements per stimulus, categorized by angular deviation: movements between 1° and 2°, 3° and 5°, and 6° or greater [5]. The researchers examined the relationship between GT parameters – such as the average frequency of TFF and eye movements across degrees – and the test-retest reproducibility of VF measurements using Humphrey VF tests (24-2 and 10-2 SITA Standard) conducted within a 3-month period. The study found that GT parameters, particularly TFF and the frequency of eye movements deviating 3° to 5° from the fixation point, were closely related to VF result reproducibility. They concluded that the best model for predicting test-retest variability in the SITA Standard 24-2 strategy was the TFF, used as a quantitative variable derived from GT. The most predictive factor for the SITA Standard 10-2 strategy was the average frequency of eye movements recorded by GT (between 3° and 5°) [5].

A follow-up study by the same group in 2015 included 244 eyes from 155 glaucoma patients, aiming to validate previous findings by integrating GT metrics into models assessing the structure-function relationship in glaucoma [7]. Although focused on correlating VF data with OCT parameters, the study reinforced the clinical value of GT metrics, such as tracking failures and eye movement frequency, by demonstrating their influence on the accuracy of VF interpretation alongside classic reliability indices.

These same authors published another study in 2015, using the SITA Standard 30-2 strategy [4]. In this study, they focused on evaluating GT parameters as indicators of VF reliability, specifically analyzing their role in predicting over- or underestimation of VF results. Unlike the previous studies from the same group, this study used a larger sample of 631 eyes from 400 patients, analyzing 10 VF tests per patient. They examined relationships between mean deviation (MD) and GT parameters, such as eye movement frequency, TFF, and blink frequency (BF), the latter derived from the downward bars in the GT trace indicating eyelid closures. The findings revealed that high FL and FP elevated MD values, while high values of eye movements, TFF, BF, and pattern standard deviation decreased MD values, suggesting GT parameters as new reliability indices for VF testing. In particular, they found that high rates of FL and FP tended to overestimate the MD values. Conversely, higher values of eye movement (≥ 3° in GT), TFF, and BF showed a tendency to underestimate the MD values. The authors suggested that GT parameters could serve as new reliability indices for VF assessments, aiding in the prediction of overestimation or underestimation of MD results [4].

Researchers from the same institution (Arai et al.) published another study in 2018, examining the relationship between ocular surface condition and GT metrics in patients with preperimetric glaucoma [8]. The authors measured ocular surface parameters (tear breakup time, Schirmer test, tear meniscus volume, and the presence of superficial punctate keratopathy), as well as GT parameters. Using statistical models, they found a significant association between superficial punctate keratopathy and a higher frequency of larger fixation deviations, and between high tear meniscus volume and a higher frequency of GT failures. They concluded that these ocular conditions can affect the interpretation of GT results in glaucoma evaluation, necessitating careful consideration when interpreting GT parameters in patients with superficial punctate keratopathy or high tear meniscus volume [8].

Asaoka et al., also from the same institution in Japan, employed a method for quantifying GT traces. Using the SITA Standard 24-2 and only the central 52 test locations overlapping with the 24-2 test pattern of the SITA Standard 30-2 strategies, they analyzed the significance of traditional reliability indices and GT parameters in assessing the progression of sensitivity loss in the visual field [9]. They found no improvement in diagnostic accuracy when stricter criteria were applied for FL and FP. However, significant improvements were observed when stricter thresholds were used for FN and GT parameters. The authors proposed that GT parameters should be represented as quantitative indices in clinical evaluations [9].

Studies on GT Metrics with SITA Fast and SITA Faster Testing Strategies

The SITA Fast and SITA Faster strategies offer the advantage of shorter test durations compared to the SITA Standard, which helps reduce visual fatigue during the examination [12]. In 2021, Camp et al. from the United States applied a quantitative method for GT bars using the SITA Fast strategy in a clinical study with a larger sample (988 eyes). The metrics were calculated as the percentage of stimuli with gaze deviations between 1°–2° (M1), 3°–5° (M3), and ≥ 6° (M6), as well as the percentage

Table 1. Characteristics of the included studies. Summary of the main features of the studies included in this review

Author/ Year	Institution*	Title	Type of Study	Sample Size	Type of Subjects	Visual Field Strategy	Follow-up time
Ishiyama et al/ 2014	The University of Tokyo, Tokyo, Japan	An objective evaluation of GT in Humphrey perimetry and the relation with the reproducibility of visual fields: a pilot study in glaucoma	Prospective cohort	42 eyes, 42 patients (20 males / 22 females)	Open angle glaucoma	SITA Standard 24-2 and 10-2	3 Months
Ishiyama et al/ 2015	The University of Tokyo, Tokyo, Japan	Estimating the Usefulness of Humphrey Perimetry GT for Evaluating Structure-Function Relationship in Glaucoma	Cross sectional study	244 eyes, 155 patients	Open angle glaucoma	SITA-Standard 24-2 or 30-2 program. In the latter only the central 52 test locations overlapping with the 24-2 test pattern were used to calculate the mean total deviation (mTD) value	n/a
Ishiyama et al/ 2015	The University of Tokyo, Tokyo, Japan	The Usefulness of GT as an Index of Visual Field Reliability in Glaucoma Patients	Prospective cohort	631 eyes, 400 patients	Open angle glaucoma	SITA Standard 30-2	n/a
Arai et al/ 2018	The University of Tokyo, Tokyo, Japan	The association between ocular surface measurements with visual field reliability indices and GT results in preperimetric glaucoma	Prospective cohort	34 eyes, 30 patients	Open angle glaucoma	SITA Standard 24-2	6 months or more
Asaoka et al/ 2019	The University of Tokyo, Tokyo, Japan	Estimating the Reliability of Glaucomatous Visual Field for the Accurate Assessment of Progression Using the GT and Reliability Indices	Prospective cohort	483 eyes, 304 patients	Open angle glaucoma	SITA Standard 24-2 or 30-2 program. In the latter only the central 52 test locations overlapping with the 24-2 test pattern were used to calculate the mean total deviation (mTD) value	6 months or more

Camp et al/ 2021	University of California San Diego, San Diego, California, USA	Standard Reliability and GT Metrics in Glaucoma and Glaucoma Suspects	Cross sectional study	988 eyes, 494 patients**	Glaucoma and glaucoma suspects	SITA Fast 24-2	n/a
Camp et al/ 2022	University of California San Diego, San Diego, California, USA	Dry Eye Symptom Severity and Visual Field Reliability Metrics	Cross sectional study	988 eyes, 494 patients**	Glaucoma, glaucoma suspects ang ocular hypertension	SITA Fast 24-2	n/a
Phu et al/ 2022	University of New South Wales, Kensington, New South Wales, Australia	GT parameters have little association with visual field metrics of intrasession frontloaded SITA-Faster 24–2 visual field results	Cross sectional study	2947 visual field results (mean age 59.6 years, SD 13.4; 409 males, 371 females). 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) intrassession visual field results	Healthy, glaucoma suspect, manifest glaucoma and non-glaucomatous optic atrophy	SITA-Faster 24-2	n/a

*Institution listed corresponds to the first author's primary affiliation
**Based on identical methodologies, demographic data, and gaze tracking metrics, it can be reasonably inferred that the two studies by Camp et al. were conducted on the same patient cohort
GT – Gaze Tracking, VF – Visual Field, n/a – Not applicable

Table 2. Summary of findings from the included studies. Overview of key results and performance metrics derived from gaze tracking data, along with the main conclusions reported by each study included in this review

Author/ Year	Title	Definition of Quantitative Metrics Derived from GT Data	Results of Quantitative Metrics Derived from GT Data	Conclusion
Ishiyama et al/ 2014	An objective evaluation of GT in Humphrey perimetry and the relation with the reproducibility of visual fields: a pilot study in glaucoma	Average TFF per stimulus, defined by the number of downward bars. Average frequency of eye movements per stimulus within 1–2°; within 3–5°; and ≥ 6°.	**SITA Standard 24-2:** The average eccentricity of eye movements was 2.3 ± 1.8° [0.92–9.6]. Per stimulus, the average frequencies were: – 1–2°: 0.62 ± 0.18 [0.07–0.93]* – 3–5°: 0.12 ± 0.08 [0.004–0.41]* – ≥ 6°: 0.10 ± 0.19 [0–0.75]* –TFF: 0.07 ± 0.13 [0.001–0.61]* **SITA Standard 10-2:** The average eccentricity of eye movements was 2.1 ± 1.4° [0.88–7.4]. Per stimulus, the average frequencies were: – 1–2°: 0.64 ± 0.15 [0.21–0.92]* – 3–5°: 0.11 ± 0.08 [0.004–0.33]* – ≥ 6°: 0.079 ± 0.12 [0–0.58]* – TFF: 0.10 ± 0.15 [0.003–0.73]*	GT metrics, particularly TFF, were significantly associated with VF test-retest variability in both 24-2 and 10-2 strategies. For 10-2, eye movements between 3° and 5° were among the strongest predictors of variability, suggesting their value as reliability indicators.

Ishiyama et al/ 2015	Estimating the Usefulness of Humphrey Perimetry GT for Evaluating Structure-Function Relationship in Glaucoma	Average frequency of eye movement between 3°–5°, and ≥ 6°; average TFF and average BF.	TFF: 0.035 ± 0.074 [0.00–0.91]* per stimulus BF: 0.020 ± 0.044 [0.00–0.49]* per stimulus Move 3–5°: 0.10 ± 0.11 [0.00–0.55]* per stimulus Move ≥ 6°: 0.51 ± 0.12 [0.00–0.94]* per stimulus	In a model predicting VF mTD, GT metrics (TFF, BF, eye movements), along with cpRNFL, GCC, axial length, FL, and FP, were significant predictors. These findings support incorporating GT parameters when interpreting VF results in glaucoma.
Ishiyama et al/ 2015	The Usefulness of GT as an Index of Visual Field Reliability in Glaucoma Patients	Average frequency of eye movement within 1–2°, within 3–5°, and ≥ 6°; average TFF and average BF.	Average eccentricity: 1.9 ± 1.5° [0–12]* Move 1–2°: 0.65 ± 0.17 [0.00–0.97]* per stimulus Move 3–5°: 0.10 ± 0.12 [0.00–0.97]* per stimulus Move ≥ 6°: 0.07 ± 0.15 [0–1.00]* per stimulus TFF: 0.037 ± 0.95 [0.00–0.90]* per stimulus BF: 0.045 ± 0.086 [0.00–0.99]* per stimulus	Higher values of GT metrics (move 3–5°, ≥6°, TFF, BF) were significantly associated with lower MD values in 30-2 SITA Standard tests. These metrics may serve as indicators of VF reliability and sensitivity underestimation.
Arai et al/ 2018	The association between ocular surface measurements with visual field reliability indices and GT results in preperimetric glaucoma	Average frequency of eye movement within 3–5°, and ≥ 6°, average TFF and average BF.	TFF: 0.02 ± 0.03 [0.00–0.15]* per stimulus BF: 0.03 ± 0.10 [0.00–0.59]* per stimulus Move 1–2°: 0.70 ± 0.14 [0.15–0.92]* per stimulus Move 3–5°: 0.12 ± 0.14 [0.0044–0.66]* per stimulus Move ≥ 6°: 0.02 ± 0.03 [0.00–0.11]* per stimulus	In eyes with superficial SPK or high tear TMV, GT metrics (move 3–5°, ≥.6°, TFF) were significantly altered. These findings suggest that ocular surface conditions may influence GT data, warranting cautious interpretation in such cases.
Asaoka et al/ 2019	Estimating the Reliability of Glaucomatous Visual Field for the Accurate Assessment of Progression Using the GT and Reliability Indices	Average TFF, average BF, and total amount of MPS, defined as the frequency of eye MPS multiplied by the magnitude of eye movement in degrees.	TFF: 0.15 ± 0.17 [0.0089–0.94]* per stimulus BF: 0.17 ± 0.22 [0.00–0.96]* per stimulus Eye MPS: 0.011 ± 0.063 [0.0030–0.041]* per stimulus	Stricter thresholds for FN, TFF, BF, and MPS improved the diagnostic accuracy of VF progression (based on mTD rates), as shown by higher AUC values. GT indices may enhance progression assessment when stricter reliability criteria are applied.
Camp et al/ 2021	Standard Reliability and GT Metrics in Glaucoma and Glaucoma Suspects	GT metrics were calculated as the percentage of stimuli with gaze deviations between 1–2°, 3–5°, and ≥ 6°; and percentage of TFF. Tmove and Tmag were also calculated.	GT metrics (upper 95th percentile)**: – M1: 28.6% (20.7%), 66.7% – M3: 24.6% (21.8%), 67.5% – M6: 10.5% (17.3%), 49.5% – TFF: 17.0% (27.9%), 79.8% – Tmove: 62.4% (26.4%), 95.2% – Tmag: 504° (337°), 1173°	Although GT and standard reliability metrics were significantly correlated in SITA Fast 24-2 tests, low correlation strength and AUROC values suggest limited clinical applicability. GT may still serve as a complementary measure in some cases.

Camp et al/ 2022	Dry Eye Symptom Severity and Visual Field Reliability Metrics	GT metrics were calculated as the percentage of stimuli with gaze deviations between 1–2°, within 3–5°, and ≥ 6°, and percentage of TFF.	GT metrics (upper 95th percentile)**: – M1: 28.6% (20.7%, 66.7%) – M3: 24.6% (21.8%, 67.5%) – M6: 10.5% (17.3%, 49.5%) – TFF: 17.0% (27.9%, 79.8%)	TFF was the only GT metric associated with dry eye symptom severity and testing order (higher in the left eye). Other GT and standard metrics showed no relationship with dry eye status.
Phu et al/ 2022	GT parameters have little association with visual field metrics of intrasession frontloaded SITA-Faster 24–2 visual field results	A custom MATLAB program was used to extract GT ticks, with three aggregate measurements describing eye movements: the total number of ticks above the line, the total amplitude of movement (sum of tick amplitudes in degrees), and the average movement per test (amplitude sum divided by tick count). Movements over 10° were grouped as "10° ticks," potentially underestimating eye movements in these cases.	Gaze tracker data from 2947 VF tests showed that most eye movements ranged between 0° and 3°, with few cases ≥ 6°, and very few reaching 10°. No blinking artifacts were observed. Most participants showed low tracking error rates and minimal large eye deviations. Results reflect a predominance of small or infrequent gaze shifts.	In SITA-Faster testing, GT metrics showed no meaningful correlation with VF sensitivity or variability. These findings suggest limited utility of GT in this setting and support the need for longitudinal data to assess its value in progression analysis.

AUC – Area Under the Receiver Operating Characteristic Curve, BF – blinking frequency, cpRNFL – Circumpapillary Retinal Nerve Fiber Layer, FL – fixation loss, FN – false-negatives, FP – false-positives, GCC – Ganglion Cell Complex, GT – Gaze Tracking, MPS – movement per stimulus, mTD – Mean Total Deviation, SD – Standard Deviation, SPK – Superficial punctate keratopathy, TFF – Tracking Failure Frequency, Tmag – Total number of degrees of eye movements, Tmove – Percentage of stimuli with any gaze deviation, TMV – tear meniscus volume, VF – Visual Field
**Data are presented as mean ± standard deviation [minimum–maximum].*
*** Data are presented as mean (standard deviation), upper 95th percentile.*

of stimuli with TFF. In addition, they calculated the percentage of stimuli showing any gaze deviation (Tmove), representing how often the gaze shifted from the fixation target during the test; and the total magnitude of eye movements (Tmag), which sums the degrees of these deviations across all the stimuli. The researchers found a statistically significant correlation between standard reliability metrics and GT. However, due to the low strength of these correlations, they questioned the clinical relevance of the associations [14]. The researchers suggested that their findings indicated that these parameters might complement conventional metrics [13].

In a second study, the same authors investigated, in this patient group, the relationship between the severity of dry eye symptoms (according to the 5-item Dry Eye Questionnaire) and the reliability of VF tests [10]. Standard VF reliability metrics and the previously mentioned GT metrics were used. They found that only the TFF was significantly associated with the severity of dry eye symptoms; greater severity correlated with higher TFF. In addition, it was observed that TFF was higher in the left eye (examined second), suggesting that the order of the examination and the presence of dry eye could affect the interpretation of TFF in VF tests. The authors concluded that TFF should be interpreted with caution in patients with dry eye, considering both symptom severity and the eye examined [10].

On the other hand, Phu et al. from Australia studied intrasession SITA-Faster 24-2 VF outcomes [14]. They used three aggregated measures to describe the eye movements of each subject in their study on GT metrics. These measures, extracted from the VF printout using a custom MATLAB program, were based on the interpretation of the GT output, where marks above a horizontal line represent gaze deviations. The three aggregated measures were as follows: firstly, the total number of ticks above the baseline, which represented the total number of fixation deviations – any mark above the horizontal line, regardless of amplitude. Secondly, the sum of amplitudes, calculated by adding the amplitudes of all deviations: the number of 2° marks was multiplied by 2, the number of 3° marks by 3, and so on, with the result expressed in total degrees per test. Thirdly, the average amplitude, obtained by dividing the sum of amplitudes by the total number of deviation ticks, excluding tracking failures. Their results indicated that the current outputs of the GT did not appear to provide unique or clinically meaningful insights for interpreting VF results beyond what is already achievable through existing methods [14].

DISCUSSION

In 2014, Ishiyama et al. in Japan became the first group to conduct a study using quantitative parameters derived from GT trace data [5]. They proposed that these values could serve as valuable tools for clinicians to estimate the reliability of VF testing and to improve the interpretation of results. This approach represented an innovative step toward integrating GT data into clinical practice, moving beyond a mere subjective evaluation of the GT trace and potentially enabling the use of these metrics to enhance the diagnostic precision of VF in glaucoma management.

The studies identified and reviewed on this topic, in the context of SITA Standard 24-2, 10-2, and 30-2 strategies, consistently highlight the value of GT in assessing VF test reliability. Direct eye position measurements through GT provide a more accurate reflection of test reliability compared to traditional FP metrics. High FP rates are known to inflate MD values, especially in the SITA Standard algorithm, where responses recorded before the minimum response time of approximately 180 milliseconds are classified as FPs [4,5]. This algorithmic limitation leads to the omission of FPs occurring after the minimum response time from the final reliability calculation. In contrast, GT parameters capture real-time eye movements, providing a direct measurement of fixation stability during actual threshold determinations.

Several of the reviewed studies demonstrated significant associations between specific GT parameters and test reliability. For instance, fixation deviations ranging from 3° to 5°, deviations ≥ 6°, the average TFF, and the average BF were linked to lower MD values [4], likely attributable to poor fixation during the examination. These findings suggest that GT metrics may reveal subtle disruptions in fixation that could otherwise go unnoticed with standard reliability indices.

However, not all studies arrived at the same conclusion. Investigations comparing traditional reliability metrics with GT parameters in VF tests, utilizing shorter SITA Fast and SITA Faster 24-2 strategies, found no clinically significant correlations [13,14]. One possible explanation is that the shorter duration of these strategies reduces the likelihood of excessive eye movements, thereby diminishing the impact of GT parameters on reliability assessment in these contexts. This discrepancy highlights the need for further research to identify the specific scenarios where GT data provide the most value. Moreover, one of the revised studies demonstrated that GT parameters not only influenced the interpretation of VF reliability, but also significantly influenced the correlation between VF functional outcomes and OCT structural measures. This suggests that GT data may play a role in evaluating the consistency between functional and structural assessments in glaucoma, potentially enhancing diagnostic confidence and monitoring accuracy [7].

On the other hand, the relationship between GT parameters and ocular surface conditions highlights the complexity of interpreting GT data in VF testing. The findings by Camp et al. highlight the significant impact of dry eye symptoms on GT metrics. The severity of these symptoms may affect the acceptable range or threshold, particularly for TFF, when determining VF reliability. Moreover, the acceptable range or threshold for TFF may differ between the right and left eyes, depending on the order in which the VF tests are conducted [10]. Similarly, Arai et al.'s observation of increased eye movement values, as registered in the GT line in eyes with superficial punctate keratopathy, and a substantially higher TFF in eyes with larger tear meniscus volume further underscores the need for cautious interpretation of GT parameters in such clinical contexts [8]. The influence of these ocular surface conditions on proposed GT metrics indicates that external factors unrelated to glaucomatous damage can significantly affect VF reliability indices [8-10]. The potential integration of artificial intelligence (AI)-based methods in advanced image analysis could further improve the precision and automation of GT parameter quantification. Future research should explore these AI techniques to assess their feasibility and impact on the reliability of visual field testing in clinical settings.

Nevertheless, several limitations should be considered when interpreting the results of this review. Firstly, the number of available studies exploring GT in VF testing remains limited, which restricts the strength of the conclusions and underscores the need for further research. In addition, a significant proportion of the studies included were conducted in Japan, which may introduce geographic bias and limit the generalizability of the findings to broader clinical settings. Another relevant limitation is that while some studies were prospective, the majority focused on cross-sectional data, and longitudinal analyses evaluating the role of GT in tracking glaucoma progression over time remain scarce. Further large-scale and standardized studies across diverse populations are needed to clarify the specific contexts in which GT data offer the most clinical value.

CONCLUSION

Despite mixed findings, GT trace quantification seems to hold significant promise as a reliability parameter for VF testing, particularly within the SITA Standard framework. The development and application of advanced image analysis techniques could enable precise and automated quantification of GT parameters. These innovations would streamline processes and support further clinical studies to determine whether GT metrics indeed enhance the evaluation of VF data reliability in clinical practice, ultimately assisting clinicians to improve patient diagnosis and monitoring. Continued research and validation are essential to fully integrate these approaches into routine ophthalmic assessments.

REFERENCES

1. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global Prevalence of Glaucoma and Projections of Glaucoma Burden through 2040. Ophthalmology. 2014;121(11):2081-2090.
2. Lee SSY, Mackey DA. Glaucoma – risk factors and current challenges in the diagnosis of a leading cause of visual impairment. Maturitas. 2022;163:15-22.
3. Saifi AI, Nagrale P, Ansari KK, Saifi I, Chaurasia S. Advancement in Understanding Glaucoma: A Comprehensive Review. Cureus. 2023;15(9):e46254.
4. Ishiyama Y, Murata H, Asaoka R. The Usefulness of Gaze Tracking as an Index of Visual Field Reliability in Glaucoma Patients. Invest Ophthalmol Vis Sci. 2015;56(11):6233-6236.
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(12):8149-8152.
6. Heijl A, Patella VM, Bengtsson B. The Field Analyzer Primer: Excellent Perimetry. 5th ed. Jena: Carl Zeiss Meditec; 2021. p. 22–27, 43-45, 90–94.
7. 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(13):7801-7805.
8. Arai T, Murata H, Matsuura M, Usui T, Asaoka R. The association between ocular surface measurements with visual field reliability indices and gaze tracking results in preperimetric glaucoma. Br J Ophthalmol. 2018;102(4):525-530.
9. 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. Ophthalmology Glaucoma. 2019;2(2):111-119.
10. Camp AS, Long CP, Galor A, Yamane M, Proudfoot JA, Weinreb RN. Dry Eye Symptom Severity and Visual Field Reliability Metrics. Journal of Glaucoma. 2022;31(5):305-309.
11. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021; 372:n71.
12. Bengtsson B, Heijl A. SITA Fast, a new rapid perimetric threshold test. Description of methods and evaluation in patients with manifest and suspect glaucoma. Acta Ophthalmol Scand. 1998;76(4):431-437.
13. Camp AS, Long CP, Patella VM, Proudfoot JA, Weinreb RN. Standard Reliability and Gaze Tracking Metrics in Glaucoma and Glaucoma Suspects. Am J Ophthalmol. 2022;234:91-98.
14. 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 Physiologic Optic. 2022;42(5):973-985.