AJO International 3 (2026) 100231

Contents lists available at ScienceDirect

AJO International

journal homepage: www.sciencedirect.com/journal/ajo-international

Advancing glaucoma progression detection: Simulated evaluation of 24-2 and 10-2 visual field testing strategies using real-world data

Bethany E Higgins [a,b], Giovanni Montesano [a,b,c], Masaki Tanito [d], Shiro Mizoue [e], Kazuhiko Mori [f], Katsuyoshi Suzuki [g], Takehiro Yamashita [h], Kazunori Hirasawa [i], Nobuyuki Shoji [i], David Crabb [a], Ryo Asakoa [j,k,l,m,*]

[a] Optometry and Visual Sciences, City St. George's, Northampton Square, London, United Kingdom
[b] Institute of Ophthalmology, University College of London, London, United Kingdom
[c] Moorfields Eye Hospital NHS Foundation Trust, London, United Kingdom
[d] Department of Ophthalmology, Shimane University, Shimane, Japan
[e] Department of Ophthalmology, Ehime University Graduate School of Medicine, Ehime, Japan
[f] Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan
[g] Department of Ophthalmology, Yamaguchi University Graduate School of Medicine, Yamaguchi, Japan
[h] Department of Ophthalmology, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan
[i] Department of Ophthalmology, Kitasato University School of Medicine, Sagamihara, Kanagawa, Japan
[j] Department of Ophthalmology, Seirei Hamamatsu General Hospital, Hamamatsu, Japan
[k] Seirei Christopher University, Shizuoka, Japan
[l] Organization for Innovation and Social Collaboration, Shizuoka University, Shizuoka Japan
[m] The Graduate School for the Creation of New Photonics Industries, Shizuoka Japan

ARTICLE INFO

Keywords:
Glaucoma
Glaucoma progression
Visual field testing
10–2
24–2
Simulation
Hierarchical modeling

ABSTRACT

Objective/Purpose: To evaluate whether alternating 24–2 and 10–2 visual field (VF) testing improves overall glaucoma progression detection compared with 24–2 or 10–2 alone, and to determine whether this enhances detection within the macular region, using simulated series derived from real-world data.
Design: Retrospective, observational simulation study.
Subjects, Participants, and/or Controls: Retrospective VF data from 426 eyes were used for simulations.
Methods, Intervention, or Testing: Pointwise slopes were estimated from 24–2 and 10–2 VFs using hierarchical modelling with location-level random intercepts/slopes. Eye-specific residuals formed noise templates. These were combined with estimated slopes to simulate 426 dense VF series (every 3-months), representing stable (slope=0) and progressing eyes. Four strategies were created and tested over 24-months: (A) simultaneous 24–2 and 10–2 at every visit, (B) 24–2 alone, (C) 10–2 alone and (D) alternating 24–2 and 10–2. Progression was determined using hierarchical modeling and strategy detection (hit) rate was assessed via cluster-adjusted ROC analyses, reporting sensitivity at 95% specificity and partial AUCs.
Results: At 12-months, sensitivity at 95% specificity was 79%, 67%, 67% and 63% (A–D); corresponding partial AUCs were 0.033 (A) and 0.025 (B–D). A had a significantly higher AUC than B, C, and D at 12-months ($p < 0.001$ each). B exceeded C at 12 months ($p = 0.015$), while D and B did not differ ($p = 0.570$) and D exceeded C ($p = 0.016$). By 24-months, sensitivities converged for A and B (87% each), with 79% (C) and 80% (D). Similar results were seen in 244 eyes defined as early-stage glaucoma at baseline. For 284 eyes defined as macular progressors at 12-months, B detected 80% and D 81% of eyes identified by 10–2.
Conclusions: Alternating 24–2 and 10–2 testing provided no clear benefit over 24–2 alone, which remained robust for progression detection. Simultaneous testing is optimal but not clinically practical.

* Corresponding author at: Department of Ophthalmology, Seirei Hamamatsu General Hospital, 2 Chome-12-12 Sumiyoshi, Chuo Ward, Hamamatsu, Shizuoka 430-8558, Japan.
E-mail addresses: ryoasa0120@googlemail.com, rasaoka-tky@umin.ac.jp (R. Asakoa).

https://doi.org/10.1016/j.ajoint.2026.100231
Received 12 October 2025; Received in revised form 19 January 2026; Accepted 26 January 2026
Available online 28 January 2026

1. Introduction

Glaucoma, a progressive optic neuropathy, is marked by gradual damage to the optic nerve and ganglion cells, which in turn gives rise to corresponding visual field (VF) defects.[1] The VF test serves as a crucial measure in both the diagnosis and ongoing monitoring of glaucoma. The 24–2 VF test pattern is most commonly used in clinical practice and provides coverage of both central vision and the mid-periphery. However, the central VF is only sparsely sampled with this pattern (6° interval), and this can lead to early central defects being missed.[2,3] This is problematic, because even early central VF defects have been shown to affect the quality of life in patients with glaucoma.[4] Moreover, central VF defects are more common in patients with normal-pressure glaucoma,[5] which constitutes up to 92 % of open-angle glaucoma cases in the Japanese population.[6] The 10–2 pattern, instead, offers a more spatially detailed assessment of the central VF and has been suggested as an alternative to detect such early defects.[2]

Common practice involves the repetition of VF tests during successive glaucoma clinic visits to evaluate the progression of VF damage, encompassing assessments of both the entire VF through global metrics and specific locations via pointwise analysis. Testing patients with both a 24–2 and 10–2 pattern would allow comprehensive monitoring of both the central and peripheral VF. One of the challenges faced in the clinical setting is that performing a single VF test with sufficient frequency is a significant burden[7,8] undergoing an HFA 24–2 test in addition to the HFA 10–2 test adds to this burden. Alternative approaches could leverage the use of interleaved 10–2 and 24–2 tests over time, each performed every other appointment. Such an approach could more effectively sample the VF over time without additional burden to the patient and the clinics.

To analyse progression, a simple ordinary least squares linear regression on global VF metrics, such as mean deviation and VF index, is typically used. However, this approach can only provide a progression summary of individual test patterns, because such indices are not comparable between different grids. Hierarchical modelling would allow an efficient combination of locations from different tests. This would efficiently combine progression information from multiple testing patterns and be a viable approach to combine 10–2 and 24–2 interleaved over multiple visits.

The primary aim of this study was to determine whether interleaved VF testing (alternating 24–2 and 10–2 tests at successive visits) provides greater efficiency in detecting glaucoma progression compared to strategies based on a single test type (24–2 alone or 10–2 alone). A secondary aim was to evaluate whether alternating 24–2 and 10–2 testing improves detection of progression within the central 10°, compared with repeated 24–2 testing alone. As prospective data with sufficiently dense follow-up with both 10–2 and 24–2 test types was not available, a simulation framework[9] using hierarchical modelling based on real-world VF data was utilised. This approach allowed a direct, controlled comparison of alternative VF testing schedules, while preserving the spatial and temporal features of glaucoma progression observed in clinical practice.

2. Methods

2.1. Retrospective dataset

This study retrospectively assessed longitudinal clinical data from 24–2 and 10–2 VF test patterns from Humphrey Field Analyzer (HFA; Carl Zeiss Meditec) during successive clinic visits collected from patients with glaucoma. Patients with glaucoma were recruited from the JAMDIG dataset[10] or sourced from medical records from Seirei Hamamatsu General Hospital or Shimane University Hospital. All patient data were anonymised during data extraction and were transferred to a single secure database held at Seirei Hamamatsu General hospital. This study was approved by the institutional review board of Seirei Hamamatsu General Hospital and Seirei Center for Health Promotion and Preventive Medicine (Institutional Review Board No 3306) and conducted according to the tenets of the Declaration of Helsinki. Patients provided written consent for their information to be stored in the hospital database and used for research.

All VFs were recorded on the HFA using a Goldmann size III stimulus with a 30–2, 24–2 or 10–2 test pattern and the Swedish Interactive Testing Algorithms (SITA Standard). 55,718 30–2 VFs were transformed into 24–2 for the purpose of this analysis. The database contained 136,771 VFs from 29,953 eyes recorded between November 1996 and June 2024. Data series for eyes were truncated after any ocular surgery other than cataract extraction during the follow-up period. Eyes that had ocular comorbidities such as macular degeneration (based on participant medical records), or had fewer than four 10–2 and four 24–2 VFs over a minimum follow-up of one year were excluded. No exclusions were made based on glaucoma subtype or whether glaucoma was present in one or both eyes, and eyes were included irrespective of whether VF loss was observed on the 10–2 or 24–2 test pattern. Individual VFs were excluded if they had ≥15 % false-positive errors or ≥20 % fixation losses. No exclusions were made based on false-negative errors.[11]

The final dataset included 7468 VFs from 426 eyes (309 participants). The mean (±standard deviation; SD) Mean Deviation (MD) was −9.80 (±7.07) dB and the average age of participants was 72 (±13) years at the start of their follow-up. See Table 1 for full details.

2.2. Data simulation

To generate densely sampled VF series for analysis, the previously published simulation framework by Wu and Medeiros[9] was utilised. This approach was used to create simulated 10–2 and 24–2 VF tests at regular 3-month intervals over a 2-year period, based on real-world longitudinal VF data. See Fig. 1 for schematic diagram of simulation pipeline.

2.2.1. Observed rates of progression from retrospective data

For each eye, pointwise rates of progression were estimated using a hierarchical linear mixed model (LMM) jointly fitted to the retrospectively collected 24–2 and 10–2 VF data. This model incorporated random slopes and intercepts at the test location level, to preserve the individual patterns of progression and correlation between 10–2 and 24–2 locations over time within the same eye. Specifically, all 10–2 and 24–2 locations were included in a single model, with data from overlapping locations treated as distinct data points for the same location and each assigned its individual random intercept and slope. After fitting, residuals were calculated for each location. The random effects for each location were estimated as Best Linear Unbiased Predictions (BLUPs), which provide the most likely deviation of each location from

Table 1
Characteristics and demographics of dataset used for simulation.

Data Characteristics	N of eyes
Date Source:	
JAMDIG	113
Shimane University Hospital	313
Test Type:	
10–2	3401
24–2	4067
Demographic and Clinical Characteristics	**Mean (±SD)**
Age (years; at start of follow-up)	72 (13)
Mean Deviation (dB)	−9.80 (7.07)
Mean Deviation (dB) for 24–2 tests only	−8.56 (7.28)
Mean Deviation (dB) for 10–2 tests only	−10.90 (8.70)
Pattern Standard Deviation (dB)	8.87 (4.56)
Pattern Standard Deviation (dB) for 24–2 tests only	8.76 (5.24)
Pattern Standard Deviation (dB) for 10–2 tests only	8.55 (5.09)
Rate of Mean Deviation Change:	
24–2 (dB per year)	−0.39 (0.69)
10–2 (dB per year)	−0.55 (0.97)

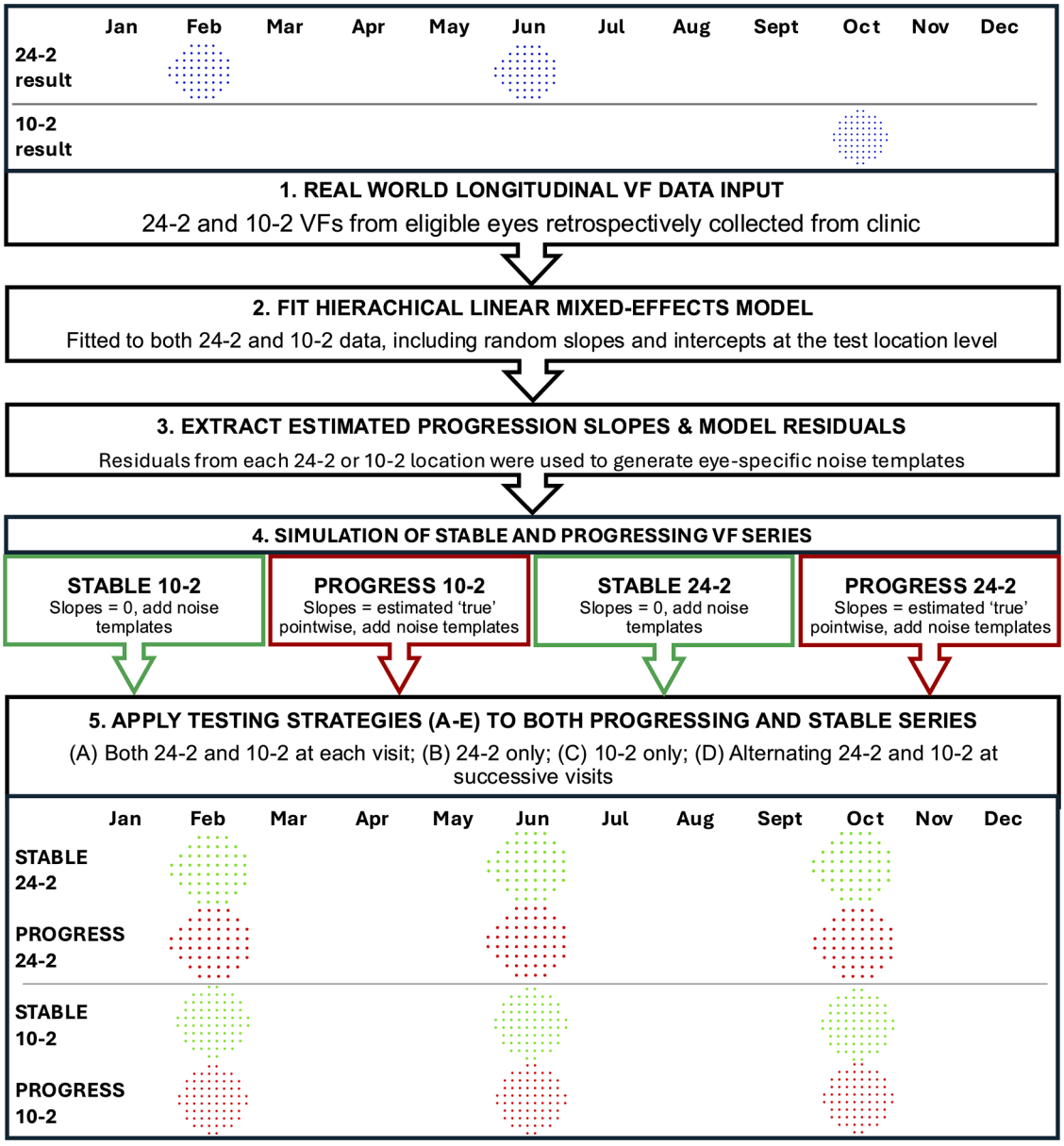

Fig. 1. A schematic diagram of data simulation pipeline. Note: VF tests sourced from clinics can be inconsistent and sparse for a patient, as illustrated by the figure. This simulation process shows how the authors created a more robust, consistent dataset to analyse.

the global trend. Eye-specific noise templates were created by extracting these residuals for each test location using the 24–2 or 10–2 pattern. These templates captured the variability patterns unique to each eye and test type[12] and global fluctuations within the same test (global visit effect). Location-specific variability was not fixed, but scaled according to sensitivity level using established variability models,[13] such that variability increased as sensitivity declined. The estimated pointwise slopes from this model were extracted to represent the underlying, or "true," rate of VF change for the following steps.

2.2.2. *Simulated progressing and stable series*

Simulated VF series were then created under two conditions. To replicate the rates observed in our dataset in our simulated series, the 'true' pointwise slopes estimated in the previous step were applied to reflect realistic patterns of disease progression. Noise templates were then added, following Wu and Medeiros,[9] to introduce variability, which would vary with the level of VF sensitivity and to retain the global visit effect. To simulate stable series, all slopes were set to zero while maintaining the same noise templates as the progressing series. This design ensured that test variability was matched between progressing and stable series for each eye, enabling controlled comparisons between the two conditions.

These simulated tests were scheduled at baseline and every three months thereafter over a 24-month period. Four testing strategies were evaluated: (A) Simultaneous testing with both 24–2 and 10–2 at each visit (our benchmark); (B) 24–2 alone at every visit; (C) 10–2 alone at every visit; (D) alternating 24–2 and 10–2 (see Table 2). Tests were scheduled at baseline (visit 1) and every 3-months thereafter.

Table 2
The Four Strategies Simulated for Comparison.

Strategy		Description
A	Simultaneous	1 × 24–2 test and 1 × 10–2 test at each visit
B	24–2 Alone	1 × 24–2 test at each visit
C	10–2 Alone	1 × 10–2 test at each visit
D	Alternating	1 × 24–2 test at one visit, then 1 × 10–2 test at next visit

Evaluations were conducted at multiple timepoints (6, 12, 18, and 24 months), based on global progression metrics derived from the combined pointwise sensitivities across all locations, rather than using individual pointwise sensitivities alone. Global progression was quantified by the fixed effect parameter describing the average change over time. Random effects modelled the trajectories for the individual locations. This approach used real-world data to inform both the progression and fluctuation patterns, allowing the evaluation of different test scheduling strategies under realistic conditions, with greater statistical power than would have been possible using the original clinical data alone.

2.3. Data analysis

All series simulated using the observed slopes were treated as potentially progressed for the subsequent analyses, while series with slopes set to zero were considered stable. This aligns with other published analyses, for which true progression cannot be established.[14–16]

The performance of different testing schemes was evaluated by comparing the proportion of eyes detected as progressed (hit-rate) at the same specificity. Here, progression was assessed statistically at the individual eye level using a LMM, and a "hit" was recorded only if the slope declined significantly (one-tailed p-value below the chosen threshold) in eyes simulated from the observed slopes. Specificity was calculated using the simulated stable series (i.e. series with a slope set to zero), which were used as a surrogate of stable test-retest series.[17,18] While hit-rate is not equivalent to sensitivity, a higher hit-rate at the same specificity implies a higher sensitivity and can be used as a surrogate measure of test performance.[17,18]

The LMM used assessed the relationship between sensitivity (*Sens*) and time (*Time*), including a random slope and intercept for *Time* within clusters (*Clust*), nested within locations (*LocN*). A random intercept for *TestID* accounted for the global visit effect.[12] Four clusters were defined: central superior, central inferior, peripheral inferior and peripheral superior. This spatial grouping enabled the model to capture region-specific variations in sensitivity progression by accounting for correlations between locations within the same cluster, while progression itself was still defined based on the LMM slope p-values.

$$Sens_i = \beta_0 + \beta_1 Time_i + u_{Clust(LocN)i} + v_{TestIDi} + \epsilon_i$$

To evaluate progression at the individual eye level, the model was fitted separately for each eye at each time point (e.g., 6 months, 12 months). For each eye, a subset of the data was created, and the LMM was re-fitted using the same structure as above. From the fitted model, the fixed effect estimates for the slope of sensitivity change over time, along with its standard error (SE), was extracted. A one-tailed p-value for the slope was computed using the t-statistic derived from the estimate and SE, as the primary interest was in detecting a significant decline in sensitivity over time, rather than any change in either direction. These values were stored for each eye to facilitate comparison of the p-values across testing strategies. This procedure was repeated for both the progressing and stable series for each of the five strategies. In total, 8 datasets were produced (four for progressing and four for stable eyes) using strategies A-D, compared at 6-month, 12-month, 18-month, and 24-month time points.

To ensure that the Simultaneous strategy (A), which includes both the 10–2 and 24–2 tests, did not perform differently simply because it had more data compared to strategy B (24–2 Alone) and strategy C (10–2 Alone), we conducted an additional comparison. For this, strategies B and C were re-run to generate a second test at the same time point using a new random seed. These independently simulated tests ensured that all three strategies were modeled with the same number of tests per visit, allowing observed differences to reflect the testing strategy rather than differences in data quantity.

To evaluate and compare the performance of each test strategy, cluster-adjusted receiver operating characteristic (ROC) curves and area under the curve (AUC) values with 95 % confidence intervals (CI) were computed. ROC curves were generated for each strategy, comparing fixed-effect one-sided p-values between progressing and stable eyes, while accounting for the clustering of eyes within individuals. These curves allow the proportion of true positives (hit-rate/sensitivity) to be plotted against 1-specificity at different thresholds. As aforementioned, strategy performance was evaluated at multiple follow-up durations (6, 12, 18, and 24 months). For each strategy and timepoint, two key metrics were extracted from the ROC analysis: sensitivity at 95 % specificity and the partial AUC (pAUC) between 95 % and 100 % specificity. These metrics were chosen to assess the ability of each test strategy to detect progression while maintaining high specificity.

As a secondary measure to compare the discriminatory power of each strategy, non-parametric bootstrapping was performed to assess whether the differences in AUC between strategies were statistically significant. Resampling was performed at the subject level (not eye level) to preserve the within-subject correlation structure. As only one simulated series per eye was used, the effective sample size and statistical power were determined solely by the number of eyes included. Note, bootstrapping does not inherently increase statistical significance; rather, the precision of the estimated p-value improves with a higher number of bootstrap samples.

In secondary analyses, we repeated the ROC analysis in a subset of 244 eyes with baseline MD >-10 on 24–2 VFs (i.e. eyes with *earlier* glaucoma). We also isolated eyes identified as progressors on the 10–2 test (strategy C), based on individual progression slopes. These same eyes were then evaluated across the other testing strategies to determine whether they were also detected as progressing, allowing assessment of strategy sensitivity for the same individuals with prominent macular field loss.

While the primary and secondary analyses focused on detection of global progression across all locations on the VF, we ran supplementary analyses to (1) examine rate estimates and (2) assess progression in central locations only. To assess whether testing strategy influenced estimated rates of progression, we compared fixed-effect slopes obtained using simultaneous 24–2 and 10–2 testing (strategy A) with slopes derived from 24 to 2-only (strategy B) and 10–2-only strategies (strategy C). To assess central locations only, we re-ran the LMM but treated the central cluster as a fixed effect and compared simultaneous 24–2 and 10–2 testing (strategy A) with 24–2-only (strategy B) and alternating 24–2 and 10–2 (strategy D). See supplemental materials for full details.

All data processing and analysis were performed in R (v4.0.5; RStudio, Boston, MA, USA), with the following packages: pROC, lme4, lmerTest, dplyr, ggplot2, tidyr, stringr, and gridExtra.

3. Results

The data demonstrated an improved ability to detect progression when both 10–2 and 24–2 tests were performed at the same visit (Strategy A; simultaneous), particularly at later visits. This finding was consistent across the entire cohort as well as in subgroups, including eyes with early disease and those classified as macular progressors (see Tables 3–5 and Fig. 2). Strategy A was therefore used as a benchmark, representing the upper bound of detectable progression, rather than a clinically practical approach.

At 12-months, strategy A achieved a sensitivity (95 % CI) of 79 % (72, 83 %) at 95 % specificity, compared to 67 % (57, 74 %) for strategy B, 67 % (58, 72 %) for strategy C and 63 % (58, 70 %) for strategy D. In other words, strategy A identified approximately 12 % more progressors compared to strategies B and C and approximately 16 % more progressors compared to strategy D at 12-months (see Table 3). The corresponding partial AUCs (95–100 % specificity) were 0.033 (A) and 0.025 (B-D) (see Table 3 and Fig. 2). This pattern was consistent across time points until 24-months, with strategy A outperforming the others from as early as 6-months (sensitivity [95 % CI] 55 % [43, 63 %] vs. 39 % [31, 47 %] for B and 46 % [38, 52 %] for C), and reaching 87 % (82,

Table 3
Comparison of ROC Outputs (Partial AUC [95 %−100 %] and Sensitivity at 95 % Specificity).

Strategy		Visit 3 (6 m)		Visit 5 (12 m)		Visit 7 (18 m)		Visit 9 (24 m)	
		Sens at 95 % Spec (95 % CI)	Partial AUC*	Sens at 95 % Spec (95 % CI)	Partial AUC*	Sens at 95 % Spec (95 % CI)	Partial AUC*	Sens at 95 % Spec (95 % CI)	Partial AUC*
A	Simultaneous	55 % (43, 63)	0.019	79 % (72, 83)	0.033	86 % (81, 91)	0.040	87 % (82, 90)	0.040
B	24–2 Alone	39 % (31, 47)	0.012	67 % (57, 74)	0.025	78 % (71, 84)	0.033	87 % (80, 90)	0.037
C	10–2 Alone	46 % (38, 52)	0.016	67 % (58, 72)	0.025	76 % (69, 81)	0.033	79 % (80, 90)	0.035
D	Alternating	-	-	63 % (58, 70)	0.025	79 % (72, 85)	0.033	80 % (74, 87)	0.035

Table 4
Comparison of ROC Curves for 6 and 12-Months.

Strategy 1	Strategy 2	Visit 3 (6 m)				Visit 5 (12 m)			
		Strategy 1 AUC	Strategy 2 AUC	AUC Difference 95 % Confidence Intervals	P Value	Strategy 1 AUC	Strategy 2 AUC	AUC Difference 95 % Confidence Intervals	P Value
Simultaneous (A)	10–2 (C)	0.87	0.83	0.028, 0.053	**<0.001**	0.93	0.89	0.029, 0.048	**<0.001**
Simultaneous (A)	24–2 (B)	0.87	0.81	0.045, 0.068	**<0.001**	0.93	0.91	0.010, 0.029	**<0.001**
24–2 (B)	10–2 (C)	0.81	0.83	−0.033, 0.09	0.08	0.91	0.89	0.004, 0.003	**0.015**
Simultaneous (A)	Alternating (D)	-	-	-	-	0.93	0.91	0.014, 0.030	**<0.001**
Alternating (D)	24–2 (B)	-	-	-	-	0.91	0.91	−0.010, 0.006	0.570
Alternating (D)	10–2 (C)	-	-	-	-	0.91	0.89	0.003, 0.030	**0.016**

Table 5
Comparison of Number of Progressing Eyes Identified at 95 % Specificity.

Strategy		Visit 3 (6 m) N progressors identified at 95 % Spec (%)*	Visit 5 (12 m) N progressors identified at 95 % Spec (%)*
A	Simultaneous	232 (54 %)	336 (79 %)
B	24–2 Alone	168 (39 %)	287 (67 %)
C	10–2 Alone	196 (46 %)	284 (66 %)
D	Alternating	-	269 (63 %)

* Total number of progressing eyes in dataset was $n = 426$.

90 %) sensitivity at 24-months, where it matched strategy B (87 % [80,90 %]), but outperformed C (79 % [80, 90 %]) and D (80 % [74, 87 %]).

At both 6 and 12-months, the AUC for strategy A was significantly higher than those of the other three strategies (see Table 4 and Fig. 2). At 6-months, strategy A's AUC differed significantly from both strategy B and strategy C ($p < 0.001$ for both). At 12-months, strategy A's AUC remained significantly greater than those of strategies B, C and D ($p < 0.001$ for each). Additional comparisons at 12-months revealed strategy B's (24–2 only) AUC was significantly greater than strategy C (10–2 only; $p = 0.015$). Also, strategy D's (alternating) AUC was significantly greater than strategy C (10–2 only; $p = 0.016$).

The results were similar when the analysis was repeated after equating the number of tests per visit, i.e. comparing the simultaneous strategy (A) to repeating the same test pattern twice on the same visit (strategy B and C repeated, respectively, see Fig. 3). The main analysis

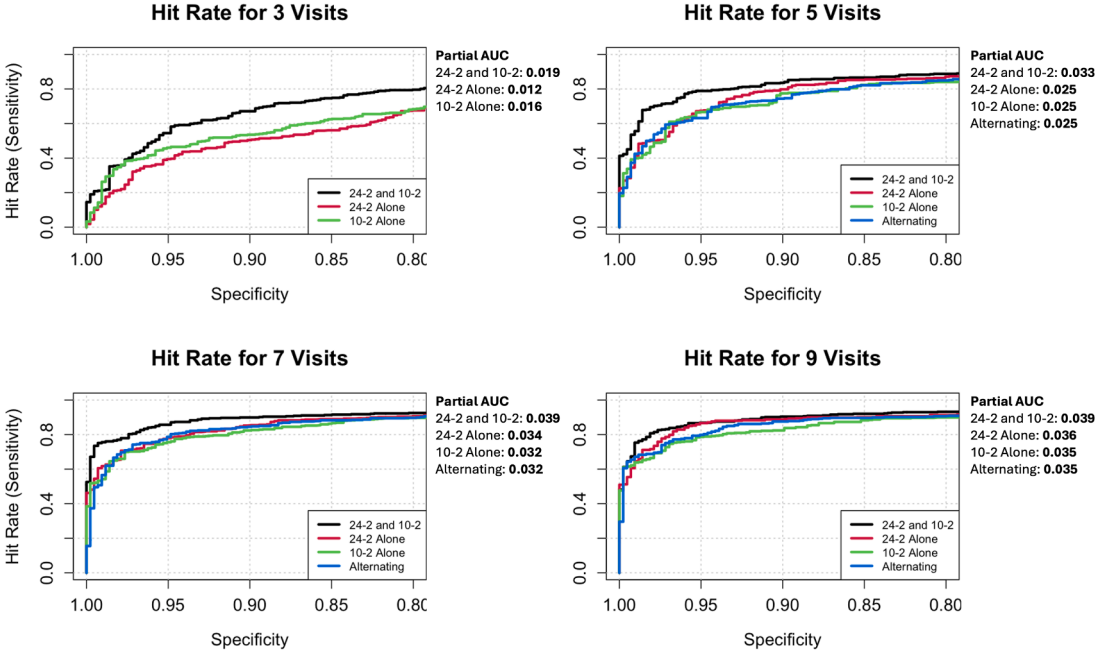

Fig. 2. ROC curves comparing detection (hit) rates to specificity of all strategies tested.

was repeated in 244 eyes with early disease at baseline. While strategy C (10–2 Alone) was comparable in its ability to detect progression in earlier glaucoma disease severity compared to the simultaneous strategy (A) for first 6-months (partial AUC 0.019 and 0.017, respectively), this difference was lost at 12-months (partial AUC 0.026 and 0.034, respectively; see Fig. 4).

A secondary aim of this study was to investigate if alternating 24–2 and 10–2 tests (strategy D) allowed a better detection of progression within the central 10 degrees, as opposed to just 24–2 tests. Strategy C (10–2 Alone) identified 284 (67 %) progressors at 95 % specificity at 12-months (see Table 5) and out of these, strategy B, identified 80 % of these 284 at the same specificity while the interleaved strategy D identified 81 % at the same specificity of these macular progressors (see Table 6 and Fig. 5).

Our supplementary analyses compared progression rate estimates between strategies (A vs B and A vs C) and also progression tested in only central locations (A vs B vs D). These comparisons demonstrated no systematic differences in estimated rates nor difference in progression detection between strategies tested. See supplemental materials.

4. Discussion

This study aimed to compare different VF testing strategies for detecting progression by using hierarchical modelling[19] as a flexible framework. Simulated VF series were generated from retrospectively collected, longitudinal real-world data series of 24–2 and 10–2 VF tests. Preserving the spatial and temporal structure of each eye, hierarchical models were applied to these series to combine information across test locations and visits. The data demonstrated that performing both 10–2 and 24–2 tests during the same visit significantly improved the ability to detect progression and served as a benchmark for maximal information, particularly at 6 and 12-months. While alternating 24–2 and 10–2 tests provided marginal improvements over using 10–2 alone, they did not match the performance of simultaneous testing and performed similarly to 24–2 alone for both global and central (macular) progression. This represents the main finding of the study: interleaved (alternating) test types offers limited benefit over single-test approaches.

When the analysis was repeated with an equal number of tests per visit, comparing the simultaneous strategy to repeated 24–2 or 10–2 tests within a visit, the pattern of findings remained consistent. This suggests the observed benefit of combining test types was not merely a result of having more data points, but rather due to the complementary information captured when both tests are conducted together. The results were consistent across the full cohort, including eyes with early glaucoma, suggesting broad applicability across disease stages. Furthermore, when estimated rates of progression were compared across strategies, no systematic differences were found. This result aligns with recent literature.[20]

Within the central 10 degrees of the VF where 30 % of retinal ganglion cells are located, the 24–2 strategy only tests 12 locations.[21] Therefore, 24–2 VF tests alone may not be sufficient to effectively estimate glaucomatous damage to the macula and central defects may be missed.[2] In addition, central VF defects are seen more readily in patients with normal-pressure glaucoma,[5] which constitutes up to 92 % of open-angle glaucoma cases in the Japanese population.[6] A secondary aim of this study was to determine whether alternating between 24–2 and 10–2 tests could improve detection of macular progression compared to 24–2 testing alone. Although a very modest improvement was seen with interleaving, this difference was not statistically significant, and thus does not appear to offer clinical benefit over 24–2 alone in identifying macular progressors. This result should be interpreted in the context of limited early-stage 10–2 data, which constrains the ability to fully evaluate potential benefits of interleaved testing.

These findings serve to demonstrate that hierarchical modelling[19] is a feasible and flexible method for combining longitudinal data from different VF test types. However, limitations in the available data constrained a full evaluation of its potential effectiveness. In particular, real-world datasets often reflect clinical decision-making, where 10–2 testing is introduced only after signs of central damage are suspected or when 24–2 data becomes insufficient. This delayed initiation of 10–2

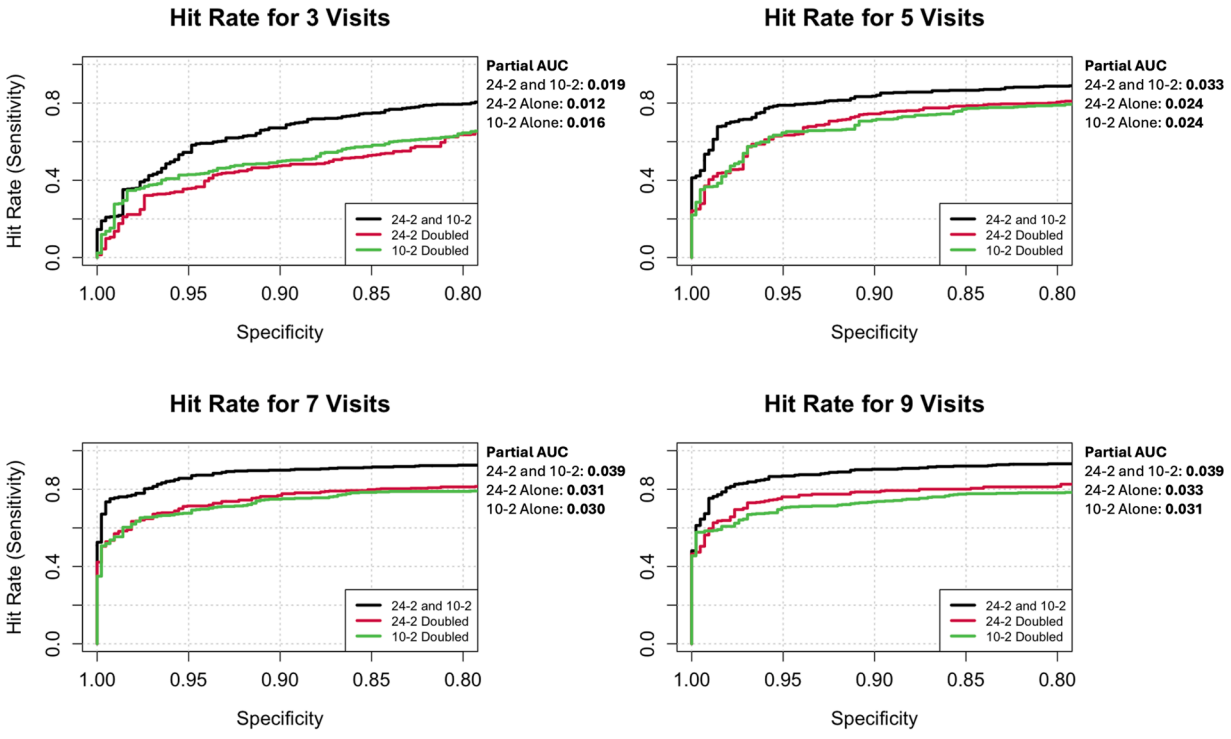

Fig. 3. ROC curves comparing detection (hit) rates to specificity of the simultaneous strategy (A, black) and 'doubled' versions of strategies B (red) and C (green).
Note: Combining the two tests gives better performance than repeating the test twice, so it's not just about more data.

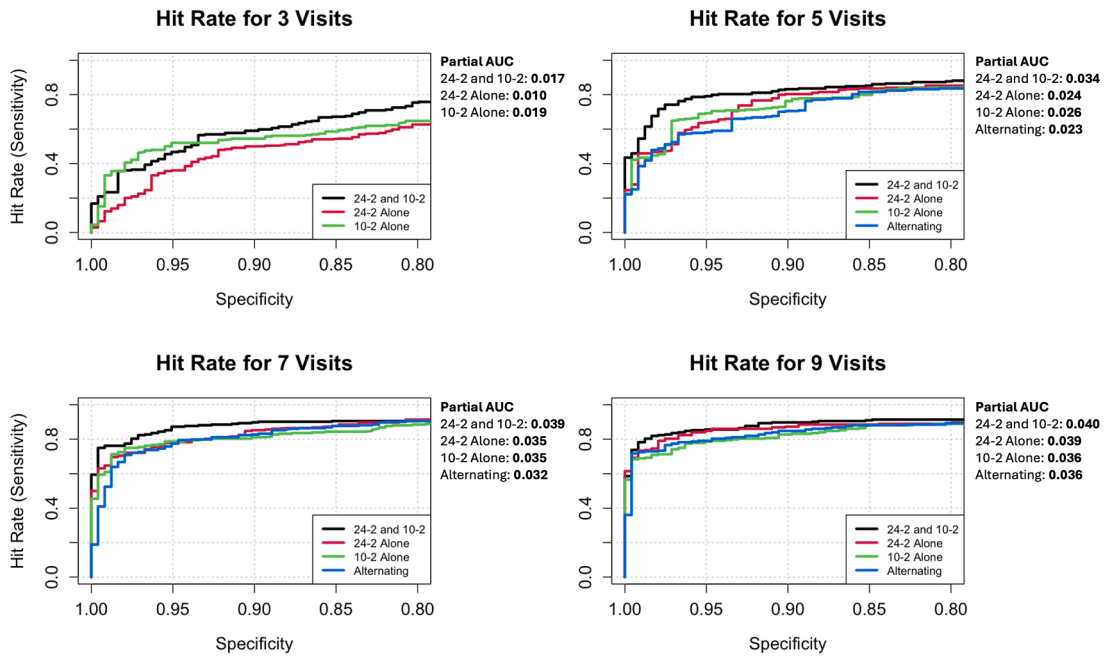

Fig. 4. ROC curves comparing detection (hit) rates to specificity of all strategies tested using subset of 244 participant's data with baseline MD >−10 on 24–2. **Note**: Strategy C is comparable to the simultaneous strategy (A) for first 6-months in early/mild disease, but this difference is lost at 12-months.

Table 6
Proportion of Progressors Identified, Detected by Strategies B and C.

Strategy		Visit 5 (12 m)	
		N (%) progressions detected out of 24–2 alone (B)	N (%) progressions detected out of 10–2 alone (C)
B	**24–2 Alone**	-	229 (80 %)
C	**10–2 Alone**	229 (80 %)	-
D	**Alternating** (starts 24–2)	246 (86 %)	230 (81 %)

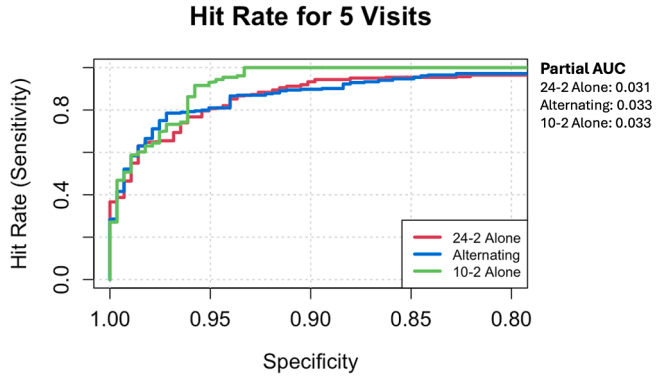

Fig. 5. ROC curves comparing detection (hit) rates to specificity of strategies B and D using subset of 284 participants identified as having macular progression. **Note:** No notable differences in number of macular progressors identified when comparing 24–2 with the alternating strategy at 12-months.

tests limits the availability of concurrent early-stage data and therefore narrows the window to test the efficacy of simultaneous strategy approaches in early glaucoma. Further research could explore the use of hierarchical modelling to combine longitudinal data from different VF test types collected prospectively, opposed to relying on available clinical data. This study also did not simulate the actual duration of tests or visit sequences and future work could evaluate the time costs associated with simultaneous versus interleaved testing strategies to help clinicians understand the trade-offs in detection speed and visit burden.

Additionally, while alternating between test types may intuitively seem advantageous by increasing spatial coverage, this approach slows the accumulation of data for each individual test type due to their non-overlapping locations. As a result, interleaved testing may hinder early detection by failing to build sufficient longitudinal series in either test type. Adaptive testing strategies that tailor the frequency and type of test to a person's previous results,[22] risk profile[23] or structural measurements,[24] may offer a more efficient solution. For example, the Australian Reduced Range Extended Spatial Test (ARREST) algorithm[22] does not re-test locations where visual thresholds are below 16 dB, as reliable measurements are often not possible in these areas.[25] Instead, it focuses neighbouring locations where thresholds are above 16 dB and results in a faster test time and better spatial characterisation of VF defects when compared to the standard 24–2 test.[22] Another approach which utilises machine learning is the Variational Bayes Linear Regression (VBLR) model,[26] which learns from the spatial and temporal patterns of progression in a training dataset and was ultimately shown to outperform conventional linear regression. These kinds of personalised strategies could help make testing more efficient and better targeted.

Prior research recommends that six VF tests should be conducted in two years using standard clinical testing methods[27] to accurately measure progression. Yet, this is often not possible in already over-stretched ophthalmology clinics. From a patient-centered perspective, simultaneous testing strategies conducted within a single visit may help streamline care and reduce visit burden. By increasing the likelihood of early progression detection, this approach also supports better-informed clinical decisions and potentially earlier interventions, which may preserve visual function and quality of life.

A key strength of this study is that the simulation framework allowed for the creation of a more homogeneous dataset with equally spaced visits, which is rarely possible in routine care. This balance made it possible to explore testing strategies under standardised conditions while still reflecting the visual thresholds of real patients. Furthermore, this is the first study to the authors knowledge that has used this hierarchical modelling methodology to assess and compare hybrid testing

strategies.

A key limitation of this study is that the analyses are based on simulated VF series derived from retrospectively collected clinical data, and results should be interpreted in that context. Further limitations around this include the underrepresentation of early-stage eyes with concurrent 10–2 and 24–2 data, and the lack of prospective consistency in test administration. In routine clinical practice, patients typically transition from 24–2 to 10–2 testing as disease advances, which limits the availability of longitudinal datasets containing overlapping 24–2 and 10–2 measurements. This likely constrained the evaluation of hierarchical modelling used in this analysis and may have led to an underestimation of its potential performance. In addition, 30–2 tests were adapted into 24–2 for this analysis. Although the central test locations overlap, the protocols differ slightly in fixation strategy, which may introduce minor inconsistencies compared to 'true' 24–2 tests. Another limitation is that clinical factors, like treatment effects, were not incorporated into the LMM when estimating progression slopes. Intraocular pressure (IOP) was not available for all the included eyes. Integrating such time-varying covariates would also be complex, but could provide more realistic progression estimates and could be the subject of future work. Lastly, the analyses did not stratify eyes by baseline VF severity, and it is possible that eyes with advanced VF loss may show different progression detection patterns across testing strategies. A prospective study design is needed to understand how simultaneous and interleaved 10–2 and 24–2 testing strategies perform across different stages and types of glaucoma, and whether they improve outcomes for patients over the long term.

In summary, the data suggests alternating 24–2 and 10–2 tests offers no meaningful advantage over 24–2 alone for global or macular progression, while 24–2 alone remains superior to 10–2 alone for overall progression detection. Based on simulated VF series derived from retrospectively collected clinical data, these findings suggest that for practical clinical purposes, 24–2 testing alone provides robust monitoring, and interleaved strategies may offer limited additional benefit for central progression, particularly in routine follow-up.

Table of contents statement

This study investigates whether alternating 24–2 and 10–2 visual field testing improves detection of glaucoma progression compared with individual strategies. Simulated series were generated from retrospective data from 426 eyes using hierarchical models to estimate pointwise slopes. Four testing strategies were evaluated over 24-months, with sensitivity and partial AUC assessed via cluster-adjusted ROC analyses. Results indicate 24–2 alone remains effective, alternating testing offers no clear advantage, and simultaneous testing maximizes detection but is clinically impractical.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Bethany E Higgins reports financial support and travel were provided by Japan Society for the Promotion of Science. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding/Support: This work was supported by the JSPS Summer Programme Award 2024. [Grant No. SP24111.]

Financial Disclosures: The authors report no financial relationships directly related to this work. Outside of the submitted work:

BH reports no relevant financial disclosures.

GM has consulted for CenterVue-iCare, Omikron, Alcon, Théa, Enavate, Dompé, and Astellas, and holds ownership interest in Relayer.

MT reports financial relationships with Santen, Ushio, Senju, Kowa, Otsuka, Chugai, Sandoz, Rohto-Nitten, Glaukos Japan, Viatris, Novartis, FiberTech, Inami, JFC SalesPlan, Nidek, Tomey Corporation, Zeiss, Topcon, AMO Japan, and Hoya.

SM reports no relevant financial disclosures.

KM has consulted for Santen and reports financial relationships with Santen, Senju, Alcon, Kowa, Otsuka, Rohto-Nitten, Nitto-Medic, Glaukos Japan, Inami, and Topcon.

KS reports no relevant financial disclosures.

TY reports no relevant financial disclosures.

KH has received honoraria from Chugai, Santen, and Kowa; a research grant from Kowa; serves on the Kitasato University Research Ethics Committee and the Japan Imaging and Perimetry Society; and holds stock in Merck & Co., Inc., Organon & Co., GSK plc, Oncolys BioPharma Inc., and Tsubota Laboratory, Inc.

NS reports financial relationships with Otsuka Pharmaceutical, Alcon Japan, Senju Pharmaceutical, Santen Pharmaceutical, HOYA Corporation, Kowa Company, and Ogura Optical.

DC has consulted for AbbVie/Allergan, Apellis, Théa, Glaukos, and Janssen; reports financial relationships with AbbVie/Allergan, Apellis, and Santen; receives support from Santen; and holds ownership interest in Irida Health.

RA reports financial relationships with Reichert Technologies, Oculus, Kowa, CenterVue, and Nidek, and has received research support from Carl Zeiss Meditec, Inc.

Other Acknowledgments: N/A

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ajoint.2026.100231.

References

1. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA*. 2014;311:1901. https://doi.org/10.1001/JAMA.2014.3192.
2. De Moraes CG, Hood DC, Thenappan A, Girkin CA, Medeiros FA, Weinreb RN, et al. 24-2 Visual fields miss Central defects shown on 10-2 tests in glaucoma suspects, ocular hypertensives, and early glaucoma. *Ophthalmology*. 2017;124:1449–1456. https://doi.org/10.1016/J.OPHTHA.2017.04.021.
3. Montesano G, McKendrick AM, Turpin A, Brusini P, Oddone F, Fogagnolo P, et al. Do additional testing locations improve the detection of macular perimetric defects in Glaucoma? *Ophthalmology*. 2021;128:1722–1735. https://doi.org/10.1016/j.ophtha.2021.06.012.
4. Blumberg DM, De Moraes CG, Prager AJ, Yu Q, Al-Aswad L, Cioffi GA, et al. Association between undetected 10-2 visual field damage and vision-related quality of life in patients with Glaucoma. *JAMA Ophthalmol*. 2017;135:742–747. https://doi.org/10.1001/JAMAOPHTHALMOL.2017.1396.
5. Araie M. Pattern of visual field defects in normal-tension and high-tension glaucoma. *Curr Opin Ophthalmol*. 1995;6:36–45. https://doi.org/10.1097/00055735-199504000-00007.
6. Iwase A, Suzuki Y, Araie M, Yamamoto T, Abe H, Shirato S, et al. The prevalence of primary open-angle glaucoma in Japanese: the Tajimi Study. *Ophthalmology*. 2004;111:1641–1648. https://doi.org/10.1016/J.OPHTHA.2004.03.029.
7. Crabb DP, Russell RA, Malik R, Anand N, Baker H, Boodhna T, et al. Frequency of visual field testing when monitoring patients newly diagnosed with glaucoma: mixed methods and modelling. *Health Serv Deliv Res*. 2014;2:1–102. https://doi.org/10.3310/HSDR02270.
8. Malik R, Baker H, Russell RA, Crabb DP. A survey of attitudes of glaucoma subspecialists in England and Wales to visual field test intervals in relation to NICE guidelines. *BMJ Open*. 2013;3, e002067. https://doi.org/10.1136/BMJOPEN-2012-002067.
9. Wu Z, Medeiros FA. Development of a visual field simulation model of longitudinal point-wise sensitivity changes from a clinical glaucoma cohort. *Transl Vis Sci Technol*. 2018;7:22. https://doi.org/10.1167/TVST.7.3.22.
10. Fujino Y, Asaoka R, Murata H, Miki A, Tanito M, Mizoue S, et al. Evaluation of Glaucoma progression in large-scale clinical data: the Japanese Archive of multicentral Databases in Glaucoma (JAMDIG). *Invest Ophthalmol Vis Sci*. 2016;57:2012–2020. https://doi.org/10.1167/IOVS.15-19046.
11. Yohannan J, Wang J, Brown J, Chauhan BC, Boland MV, Friedman DS, et al. Evidence-based criteria for assessment of visual field reliability. *Ophthalmology*. 2017;124:1612–1620. https://doi.org/10.1016/j.ophtha.2017.04.035.

12. Bryan SR, Eilers PHC, Lesaffre EMEH, Lemij HG, Vermeer KA. Global visit effects in point-wise longitudinal modeling of glaucomatous visual fields. *Invest Ophthalmol Vis Sci*. 2015;56:4283–4289. https://doi.org/10.1167/IOVS.15-16691.
13. Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci*. 2000;41:417–421.
14. O'Leary N, Chauhan BC, Artes PH. Visual field progression in glaucoma: estimating the overall significance of deterioration with permutation analyses of pointwise linear regression (PoPLR). *Invest Ophthalmol Vis Sci*. 2012;53:6776–6784. https://doi.org/10.1167/IOVS.12-10049.
15. Zhu H, Crabb DP, Ho T, DF Garway-Heath. More accurate modeling of visual field progression in Glaucoma: ANSWERS. *Invest Ophthalmol Vis Sci*. 2015;56:6077–6083. https://doi.org/10.1167/IOVS.15-16957.
16. Marín-Franch I, Artes PH, Turpin A, Racette L. Visual field progression in Glaucoma: comparison between PoPLR and ANSWERS. *Transl Vis Sci Technol*. 2021;10:13. https://doi.org/10.1167/TVST.10.14.13. –13.
17. O'Leary N, Chauhan BC, Artes PH. Visual field progression in glaucoma: estimating the overall significance of deterioration with permutation analyses of pointwise linear regression (PoPLR). *Invest Ophthalmol Vis Sci*. 2012;53:6776–6784. https://doi.org/10.1167/IOVS.12-10049.
18. Zhu H, Russell RA, Saunders LJ, Ceccon S, Garway-Heath DF, Crabb DP. Detecting changes in retinal function: analysis with non-stationary Weibull error regression and spatial enhancement (ANSWERS). *PLoS One*. 2014;9, e85654. https://doi.org/10.1371/JOURNAL.PONE.0085654.
19. Montesano G, Garway-Heath DF, Ometto G, Crabb DP. Hierarchical censored Bayesian analysis of visual field progression. *Transl Vis Sci Technol*. 2021;10. https://doi.org/10.1167/TVST.10.12.4.
20. Medeiros FA, Malek DA, Tseng H, Swaminathan SS, Boland MV, Friedman DS, et al. Short-term detection of fast progressors in Glaucoma: the fast progression assessment through clustered evaluation (Fast-PACE) study. *Ophthalmology*. 2024;131:645–657. https://doi.org/10.1016/J.OPHTHA.2023.12.031.
21. Park SC, Kung Y, Su D, Simonson JL, Furlanetto RL, Liebmann JM, et al. Parafoveal scotoma progression in glaucoma: Humphrey 10-2 versus 24-2 visual field analysis. *Ophthalmology*. 2013;120:1546–1550. https://doi.org/10.1016/j.ophtha.2013.01.045.
22. Turpin A, Morgan WH, McKendrick AM. Improving spatial resolution and test times of visual field testing using ARREST. *Transl Vis Sci Technol*. 2018;7:35. https://doi.org/10.1167/TVST.7.5.35.
23. Park SC, De Moraes CG, Teng CCW, Tello C, Liebmann JM, Ritch R. Initial parafoveal versus peripheral scotomas in glaucoma: risk factors and visual field characteristics. *Ophthalmology*. 2011;118:1782–1789. https://doi.org/10.1016/j.ophtha.2011.02.013.
24. Russell RA, Malik R, Chauhan BC, Crabb DP, Garway-Heath DF. Improved estimates of visual field progression using Bayesian linear regression to integrate structural information in patients with ocular hypertension. *Invest Ophthalmol Vis Sci*. 2012;53:2760–2769. https://doi.org/10.1167/IOVS.11-7976.
25. Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. *Ophthalmology*. 2014;121:1359–1369. https://doi.org/10.1016/j.ophtha.2014.01.020.
26. Murata H, Araie M, Asaoka R. A new approach to measure visual field progression in glaucoma patients using variational bayes linear regression. *Invest Ophthalmol Vis Sci*. 2014;55:8386–8392. https://doi.org/10.1167/IOVS.14-14625.
27. Chauhan BC, Garway-Heath DF, Goñi FJ, Rossetti L, Bengtsson B, Viswanathan AC, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthal*. 2008;92:569–573. https://doi.org/10.1136/BJO.2007.135012.