

The Frontloading Fields Study (FFS): Detecting Changes in Mean Deviation in Glaucoma Using Multiple Visual Field Tests Per Clinical Visit

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

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

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

Correspondence: Jack Phu, Centre for Eye Health, University of New South Wales, Gate 14 Barker Street, Rupert Myers Building (South Wing), Sydney 2052, New South Wales, Australia.

e-mail: jack.phu@unsw.edu.au

Received: July 29, 2021

Accepted: October 11, 2021

Published: November 15, 2021

Keywords: SITA; SITA-Faster; static automated perimetry; Humphrey Field Analyzer

Citation: 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(13):21, <https://doi.org/10.1167/tvst.10.13.21>

Purpose: To determine the impact of different numbers of visual field tests per visit for detecting mean deviation changes over time in patients with early glaucoma or suspected glaucoma and to identify a practical approach to maximize change detection.

Methods: Intrasession ($n = 322$) and intersession ($n = 323$) visual field results for patients with glaucoma or suspected glaucoma were used to model mean deviation change in 10,000 progressing and 10,000 non-progressing computer-simulated patients over time. Variables assessed in the model included follow-up intervals (0.5, 1, or 2 years), reliability rates (70%, 85%, or 100%) and number of visual field tests performed at each visit (one to four).

Results: Two visual field tests per session compared with one provided higher case detection rates at 2 years (99%–99.8% vs. 34.7%–76.3%, respectively), reduced time to detection (three or four visits vs. six to 10, respectively), and more positive mean deviation score (−4 dB vs. −10 dB, respectively) at the point of mean deviation change identification, especially in the context of unreliable results. Performing two tests per visit offered similar advantages compared with more tests. False positive change detection rates (<2.5%), were similar across all conditions. Patients followed up 6 monthly had less severe mean deviation loss at follow-up compared to 1-year and 2-year follow-up intervals.

Conclusions: Performing two tests per clinical visit at 6 months is practical using SITA-Faster and provides higher detection rates of mean deviation change in comparison with only one test performed per visit and more spaced-out intervals.

Translational Relevance: This model provides guidance for selecting the number of tests per visit to detect mean deviation change.

Introduction

Current management paradigms for glaucoma, a leading cause of irreversible blindness worldwide, emphasize two main goals: first, to preserve vision and prevent irreversible vision loss, and, second, to preserve quality of life by balancing treatment decisions against the potential impact of disease.^{1–3} Individualizing the glaucoma treatment plan involves interpreting the clinical findings to understand the individual's disease trajectory. Disease progression in glaucoma is often

quantified using visual field testing over time, as perimetry results have been strongly correlated with an individual's functionality and quality of life.⁴

Interpretation of visual field progression occurs contextually, incorporating the period over which monitoring has occurred, the variability of the results, and the fidelity of the measurements. Ideally, the greater the amount of data available, the greater the confidence in identifying true glaucomatous change. Recommendations for the number of visual field results within specific periods of time and at different levels of test variability have emphasized the need for larger

volumes of clinical data⁵ than what are typically obtained in routine clinical practice.⁶

Such recommendations often assume that tests are performed visit by visit (evenly spaced follow-ups) and further assume that all results are reliable and useable. Since the publication of such recommendations, Crabb and Garway-Heath have proposed an approach that clusters visual field tests at baseline with a more distant follow-up visit.⁷ Rather than performing a similar number of visual field tests at regular intervals (such as 6 months), this approach has been suggested to provide greater power to detect glaucoma progression over time.

Although seemingly effective, the practical implementation clustering (or frontloading) approach had previously been impeded by the time required to conduct sufficient visual field tests per visit due to test algorithms such as SITA-Standard taking 5 to 7 minutes per test. More recently, we and others have demonstrated the effectiveness of SITA-Faster at returning results similar to those of SITA-Standard but with a reduction in test of time of over 50%.^{8–10} Thus, in the time typically required for one SITA-Standard test per eye, a clinician could obtain four SITA-Faster results, two per eye.¹¹

The increasing evidence regarding the effectiveness of SITA-Faster has prompted further questions regarding the incorporation of this algorithm into clustering or frontloading approaches for identifying disease progression. Aside from the need to recognize slightly greater variability resulting from a faster algorithm, there is also the need to acknowledge the propensity of SITA-Faster to produce a greater rate of results that do not meet current recommendations of reliability, with an almost tripling of this rate compared with SITA-Standard (see below for further discussion).^{8,12}

In the present study, we incorporated real-world patient data into computer simulations to determine the detection of statistically significant mean deviation change in glaucoma suspect and early glaucoma. The variables of interest were different numbers of visual field tests per visit and different follow-up intervals, after incorporating estimates of inter- and intrasession mean deviation variability and reliability levels of SITA-Faster. Another objective of our study was to determine the amount of time spent doing visual field testing per clinical visit under each testing paradigm. This would provide practical guidance for implementation in clinics aiming for early detection of significant changes in mean deviation, a commonly used metric for measuring glaucoma progression and its prognostication. We hypothesized that there is an optimal and practical number of visual

field tests conducted per session at specific intervals to best identify patients exhibiting statistically significant mean deviation change.

Methods

This study was a retrospective study using patients derived from the centralized database of the Centre for Eye Health, University of New South Wales. Ethics approval for the study was provided by the Human Research Ethics Committee of the University of New South Wales. The study adhered to the tenets of the Declaration of Helsinki. All subjects (“intrasession” and “intersession,” as described below) provided written informed consent prior to inclusion in the study.

Visual Field Data

We used the visual field data of patients seen in the glaucoma clinic at the Centre for Eye Health, University of New South Wales, to characterize the variability of mean deviation results at different levels of glaucoma severity. Mean deviations from their visual field data were retrospectively extracted from their medical records and were used to construct distributions of mean deviation variability. Mean deviation was used because it is a conventionally used marker in clinical practice for the purposes of staging and monitoring glaucoma.^{13,14} Note that these data only served to inform the variability of mean deviation values. The simulation patients were an otherwise naïve group (see below).

Although mean deviation represents a global index of visual field integrity that is sometimes regressed linearly in commercial progression analysis software, the numerical value may represent a diverse range of possible glaucomatous scotomata, such as deep and localized defects or generalized depression. Accordingly, there is also a diversity of associated measurement mean deviation variability attributable to the individual’s sensitivity values. Recently, Wu and Medeiros¹⁵ proposed a method for transforming simulated, correlated sensitivity values into mean deviation to obtain estimates of mean deviation variability. Broadly, this provides a strategy for incorporating the change in mean deviation variability as it worsens over time. Acknowledging the heteroskedastic nature of mean deviation variability, we implemented a method using our present data to incorporate the change in mean deviation variability in our follow-up model (see further below).

Patient Cohorts

Two cohorts of patients were used, patients who had undergone frontloading (two visual field tests) within the same clinical visit (the “intrasession” cohort) and patients who had historical longitudinal visual field data (the “intersession” cohort, which had at least one reliable visual field test per visit). Both cohorts were comprised of patients who were seen within the clinic as glaucoma suspects or as patients with manifest glaucoma.

The diagnosis of glaucoma was made as per current clinical guidelines.² In short, this required the presence of glaucomatous structural defects (for example, cupping, diffuse or focal rim thinning, adjacent retinal nerve fiber layer defects) with or without accompanying reproducible concordant visual field defects on the 24-2 test grid, in the absence of other retinal or neurological pathologies. Glaucoma-suspect subjects were those in whom one or more signs of glaucoma were present but their combination was insufficient for a diagnosis of glaucoma requiring therapeutic intervention. As per the clinical protocols of the Centre for Eye Health, the diagnosis was made by one examining clinician and by a remote review by another clinician.¹⁶ For the patients with glaucoma, we selected the eye with the worse stage of glaucoma; for non-glaucoma patients, we randomly selected one eye for inclusion. Our goal was to focus on patients with early to moderate glaucoma due to the nature of the clinic from which patient data were derived, which limited the number of cases of advanced glaucoma included in the present study (defined as a mean deviation score worse than -12 dB). Nonetheless, the composition of the cohort suited the purpose of the study, because in more advanced stages of glaucoma other strategies for detecting functional change may have to be used (such as changing to the 10-2 test grid), as well as the consideration of factors such as the measurement floor effect.

The “intrasession” cohort included patients who had undergone two SITA-Faster tests per eye within the same clinical visit. These included a subset of patients who had been previously reported in our previous study (the Frontloading Fields Study).¹¹ For the purposes of the present study, we included only patients who had reliable results in both tests: $<15\%$ false positive rate, no seeding point errors (one or more primary seeding points with artificially reduced sensitivity in the absence of known pathology), and $<20\%$ of gaze tracker deviations exceeding 6° , as we have previously defined^{11,12}. We note that there is debate regarding the use of “traditional” reliability indices, such as recent work demonstrating the low contri-

bution of elevated false positive rates to measurement variability.¹⁷ For the purposes of the present study, the above criteria were chosen to reflect both the protocols of the clinic in which the data were collected and the automatic exclusion criteria from the commercially available Guided Progression Analysis linked to the Humphrey Field Analyzer hardware (Carl Zeiss Meditec, Jena, Germany). However, for the ensuing Methods and Discussion sections, we note that the “reliability” nomenclature does not refer to specific criteria, such as a specific false positive rate, which may become antiquated with time and emerging evidence. Instead, references to “reliable” and “unreliable” results represent occurrences in which user-based criteria, based on the best evidence available at the time, can be applied to afford a clinical judgment for interpretability. The resultant difference in mean deviation results for test one and test two was calculated for each patient, and the distribution of these differences characterized the intrasession test–retest mean deviation variability.

The “intersession” cohort included patients who had been seen more than once in the clinic. We extracted the visual field test results for patients who were clinically stable (including either glaucoma or suspect patients). Clinical stability was defined through clinical examination, with no evidence of structural or functional deterioration and thus no modification to the treatment plan, if applicable. This was used to reflect the impression of real-world clinicians with the available clinical data. We did not use a prespecified quantitative cut-off value for inter-test differences to identify clinically stable patients, as this would introduce bias. Thus, the quantitative differences observed between visits from the extracted data not only would reflect instrument-based factors but would also represent measurements captured in clinical practice. We extracted reliable SITA-Faster results from two adjacent visits for one eye. The difference in mean deviation results between the results from visit one and visit two was calculated for each patient, and the distribution of these differences characterized intersession test–retest mean deviation variability.

For both intrasession and intersession cohorts, we extracted and included the files of consecutive patients attending the glaucoma service to reduce the probability of selection bias within a 6-month period. Exclusion criteria included age <18 years, history of ocular trauma or surgery (aside from uncomplicated cataract surgery or selective laser trabeculoplasty), and the presence of macular or retinal pathology affecting the visual field (including age-related macular degeneration and diabetic retinopathy).

For both intrasession and intersession mean deviation variability, we also assessed its change as a function of average mean deviation severity, as previous reports have demonstrated increases in mean deviation variability with worsening stage of glaucoma.^{18,19} The change in mean deviation variability was incorporated into the model for the purposes of identifying significant progression.

Modeling Progression Over Time Using Computer Simulation

Each naïve simulated patient started off with no significant visual field defect, defined by a mean deviation score of 0 dB. This was to represent the earliest stage of glaucoma or glaucoma suspects as defined by current visual field staging systems,^{2,14} in accordance with our purpose of examining the detection of mean deviation change in the earliest stages of glaucoma.

Our variables were progression rate (-0.5 , -1 , and -2 dB/yr), number of visual field tests per visit (1–4), follow-up intervals (6 months, 1 year, and 2 years), and proportion of reliable visual field results (100% reliable or 0% unreliable, 85% reliable or 15% unreliable, and 70% reliable or 30% unreliable, reflecting a range at which rates of low test reliability may occur in clinical practice using SITA-Faster^{8,12,20}).

A notable difference between the present study and previous modeling exercises^{7,21,22} is the incorporation of low test reliability and its confounding effects. Note that these probabilities of low test reliability do not reflect specific cut-offs or parameters of reliability (for example, a specific false positive rate or gaze tracker deviations) but instead represent an approach where clinicians are compelled to discard unusable clinical data.

For the conditions where the proportion of low test reliability tests was $>0\%$, we also assessed a “one in hand” approach, where one repeat test was conducted to overcome an instance of an unreliable result. In this instance, another test was conducted, replacing the result with reliability indices outside expected limits. All unreliable results were otherwise considered to be a “missing” data point for the purposes of the linear regression analysis (see below); thus, it was possible to have visits where the analysis was not performed and where change was undetected. Such patients were retained in the analysis and followed until change was detected or until the end of the follow-up period.

At each “visit,” a patient’s mean deviation score was adjusted by mean deviation variability, beginning with the baseline visit, which can then increase with a worsening mean deviation score. To model this,

we first determined the normality of distribution of mean deviation differences (intersession and intrasession). This would allow us to use a suitable model to extract out a random value from the underlying distribution. As each simulated patient was followed naively from baseline, they did not individually possess multiple visual field results from which individual variability indices could be derived. Instead, we relied upon population-based estimates of mean deviation variability as a function of mean deviation score to use for simulation purposes (see Discussion for further details).²³

The second step was to determine the relationship between the underlying average mean deviation score and the difference between tests (inter- and intrasession). A Bland–Altman analysis would be able to visualize the change in difference, but the linear regression result would not adequately capture the heteroskedasticity. Instead, we used a sliding-window analysis to overcome the limitations associated with “gaps” in mean deviation score within our cohort. We have previously used this method to overcome similar gaps in age-related analyses,²⁴ where granular, independent-variable, real-world data (in the present study, mean deviation) are impractical to obtain. In brief, we ordered the average mean deviation result consecutively and generated standard deviation values from groups of 10 adjacent difference values. The resultant standard deviations were plotted as a function of average mean deviation result, and their relationship was used to generate adjusted mean deviation variability scores for the model.

In combination, the simulated mean deviation score at each visit was the combination of the ground-truth score (baseline plus the product of rate of change and follow-up duration) and the intra- or intersession mean deviation variability, adjusted by the severity of mean deviation score. Then, for each patient, a linear regression analysis was performed at each visit to determine if at that visit there was a significant downward trend in mean deviation score, defined as a negative slope statistically significantly different to 0 using an *F*-test at the $P < 0.05$ level of significance (two-sided, which would return a predicted false positive level of 0.025 for a negative slope). Although we recorded the slope value for mean deviation over time, we did not require that the slope be at or statistically significantly lower than the ground-truth rate of change (specified as our variable progression rates of -2 , -1 , or -0.5 dB/yr, or 0 dB/yr for false positives) to be identified as changed, only that it be negative. This was to reflect clinical practice in which it would be desirable to identify a patient who demonstrates any statistically significant visual field deterioration

at follow-up visits, given certain rates of underlying mean deviation change. Another difference with conventional automated analyses is the minimum data requirement for change analysis, with some commercial tools requiring or recommending a minimum number of data points. We only required that a linear regression analysis and corresponding extractable P value was available.

Modeling the Time Required for Each Visual Field Test

The second part of the modeling process was to understand the amount of time spent undertaking perimetric testing for each paradigm. The test duration for each visual field test was extracted for each patient. Visual field test durations were separated by reliable and unreliable results to determine their respective distributions. In addition, we examined the change in visual field test duration as a function of severity of glaucoma to account for potentially increasing

durations over time. The total test duration for each visit was calculated as the sum of all tests per eye.

We simulated 10,000 patients for each combination of conditions. The simulations were performed using a custom written program written in MATLAB 2019b (The MathWorks, Natick, MA). The outputs of interest were the number of patients demonstrating statistically significant mean deviation change; number of instances where reliable results could not be obtained, thus confounding change analysis; and the distribution of progression rate across the cohort (mean and standard deviation). The cases of change detected were plotted as a function of time of follow-up, with the resultant data being fitted with asymmetric sigmoid functions in Prism 8 (GraphPad Software, La Jolla, CA) to obtain functions with which to estimate and compare case detection over time. An example of the way we fitted the data and extracted key parameters of interest is shown in Figure 1. Two scenarios are shown. The first is time to case detection determined by extracting the intersection of a horizontally fixed line (e.g., $y = 0.95$ for 95% case detection). The second is

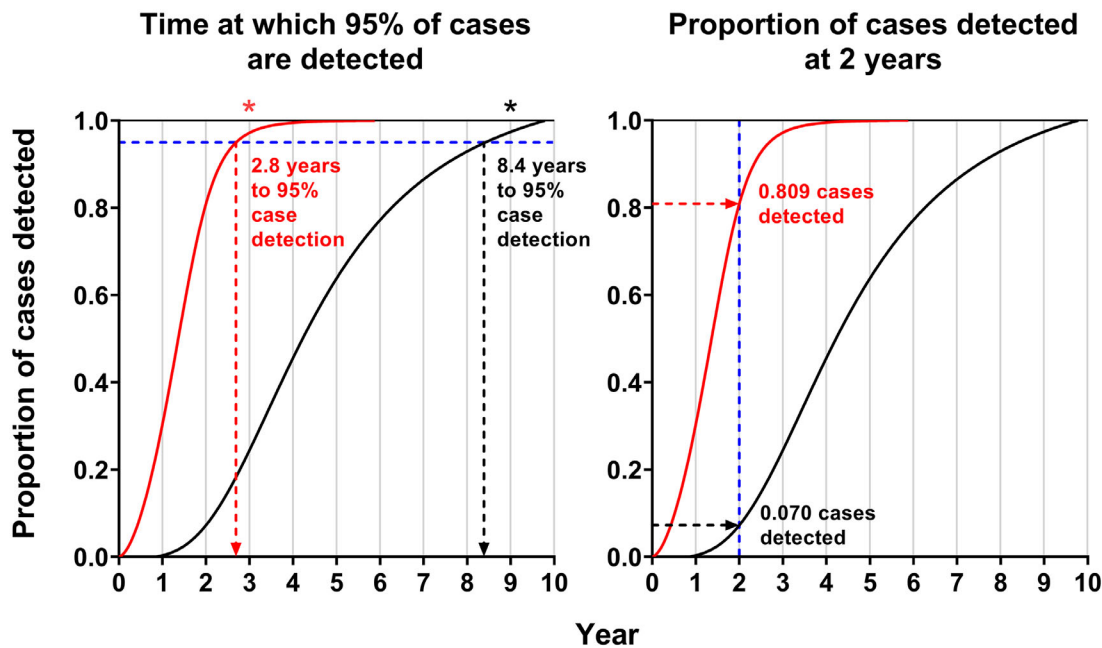


Figure 1. A schematic showing the way we extracted key parameters of interest in the present study based on the sigmoidal functions fitting the simulated cases identified as a function of follow-up time. This example represents a set of conditions with a progression rate of -2 dB per year, yearly follow up, with 70% of the data noted as reliable. (Left) A case detection criterion is set using the horizontal blue dashed line, and the time to identify this proportion of cases is determined by the intersection of the line and the curves (e.g., detection of 95% of cases is set at $y = 0.95$, intersecting the red curve at 2.8 years and the black curve at 8.4 years). (Right) A time criterion is set using the vertical blue dashed line, and the proportion of cases identified at this time point is determined by the intersection of the line and the curves (e.g., the proportion of cases identified a 2 years is set at $x = 2$, intersecting the red curve at 0.809 cases and the black curve at 0.070 cases). The red curve (two tests per visit) and the black curve (one test per visit) indicate two different numbers of tests per clinical visit, and we show two contrasting situations for clarity in this example. The gray vertical lines identify points at which an actual clinical visit took place. For example, the red asterisk at 3 years for the two tests per year condition in the left panel (time at which 95% of cases are detected) indicates the actual visit of detection, despite the model intersecting the blue dashed line at 2.8 years (similarly, the black asterisk indicates the visit for the one test per visit condition).

Table 1. Key Assumptions Made and Parameters Used in the Present Study for Modeling Purposes

Assumption or Parameter	Real-World Application	Limitation of Assumption
Mean deviation is a useful visual field metric for assessing disease progression.	Mean deviation is used as a global measurement of progression in some automated progression analyses.	Pointwise assessment of sensitivity and probability maps, as well as other global indices (e.g., pattern standard deviation, glaucoma hemifield test) may be used at various glaucoma stages for assessing disease progression.
Disease progression assessed using mean deviation occurs monotonically and linearly.	Disease progression assessed using mean deviation is depicted using linear trend analysis on some commercial software, which facilitates simple clinical interpretation (including a slope for rate of change).	Disease progression, in the long term, may be nonlinear and may vary at different stages of the disease with the use of current analysis methods.
Variability characteristics of two SITA-Faster frontloaded visual field tests may be extrapolated into more than two tests, and reliability metrics remain consistent over time.	This assumption would also inherently benefit fewer tests performed, as the benefits of increasing numbers of tests would be expected to diminish with more tests in practice. Furthermore, it would become increasingly impractical in the real world to perform an increasing number of tests (due to clinic flow and patient acceptance).	No current study has assessed more than two intra-visit tests and reported on the variability characteristics. Thus, this assumption requires empirical testing.
This simulation focused on mean deviation change occurring in the early stages of glaucoma, beginning with 0 dB at baseline.	Early and moderate glaucoma tend to be the most frequently seen in clinical practice. Advanced stages of glaucoma require incorporation of greater variability and the effect of the measurement floor. Typically, advanced stages of glaucoma are assessed using alternative methods such as the 10-2 test grid. Furthermore, visual field results tend to be less reliable in advanced disease, which would add further sources of variability to the model.	This model does not make inferences about the early detection of progression in glaucoma in the context of advanced visual field loss at baseline.

See Discussion for more details and relevant citations.

the number of cases detected at a given time period by extracting the intersection of a vertically fixed line (e.g., $x = 2$ years for cases detected at 2 years).

Several assumptions were made to develop the computer simulation model, and these are summarized in Table 1.

Results

Table 2 shows the age, gender, ethnicity, and diagnosis distributions of the present cohort. The average mean deviation result from the Humphrey Field Analyzer across both tests for each subject in each cohort are also shown in Table 2. We used all data to determine the distribution of mean deviation differences (test 1 – test 2) for intersession and intrasession

tests (Supplementary Figs. S1A, S1B). Extra sum-of-squares F -tests showed differences in the best-fit Gaussian curve between distributions, $F(3, 30) = 27.30$, $P < 0.0001$. The linear regression analysis applied to the Bland–Altman plots revealed a small but significant effect of visual field severity on intrasession mean deviation difference ($P = 0.0054$), but no significant effect on intersession ($P = 0.2935$) difference in mean deviation within this range of glaucoma stages (early to moderate no worse than -12 dB) (Supplementary Figs. S1C, S1D). There was notable heteroskedasticity, as observed in subjects with worse mean deviation scores. Monte Carlo simulation applied to both linear regression analyses returned similar results. The borderline and lack of statistically significant effects for intrasession and intersession differences were most likely a product of limiting the range of patients to those with

Table 2. Demographics of Subjects Included in the Present Study That Inform the Distribution of Mean Deviation Differences Between Tests 1 and 2

Demographic	Intrasession Cohort (<i>n</i> = 322)	Intersession Cohort (<i>n</i> = 323)
Age (yr), median (interquartile range)	62.8 (54.1–69.2)	65.1 (58.1–71.8)
Gender, <i>n</i> (%)		
Male	176 (54.7%)	191 (59.1%)
Female	146 (45.3%)	132 (40.9%)
Ethnicity, <i>n</i> (%)		
Caucasian	164 (50.9%)	182 (56.3%)
East Asian	128 (39.8%)	112 (34.7%)
South Asian	21 (6.5%)	18 (5.6%)
Central or South American	4 (1.2%)	5 (1.5%)
African	2 (0.6%)	2 (0.6%)
Aboriginal or Torres Strait Islander	3 (0.9%)	2 (0.6%)
Mixed	0 (0%)	2 (0.6%)
Diagnosis, <i>n</i> (%)		
Glaucoma	133 (41.3%)	140 (43.3%)
Glaucoma suspect	189 (58.7%)	183 (56.7%)
Humphrey Field Analyzer average mean deviation (dB)		
Median and interquartile range	−1.32 (−3.17 to −0.16)	−0.91 (−2.33 to +0.05)
Full range	−29.31 to +2.86	−22.92 to +2.52

early-to-moderate glaucoma or suspected glaucoma, with a greater number of subjects with smaller magnitude mean deviation scores. Sliding window analysis performed on the distributions of intersession and intrasession mean deviation values showed an increase in variability of mean deviation measurements with a lower (worse) mean deviation score (Supplementary Figs. S1E, S1F). As described in Methods, we incorporated the change in mean deviation variability as a function of mean deviation score into the change analysis models below.

Cases of Mean Deviation Changes Detected at Each Time Point

Figure 2 shows the proportion of cases detected by each number of tests per session as a function of follow-up time for a simulated progression rate of -2 dB per year. Each combination of reliability (rows) and follow-up interval (columns) conditions is shown in a separate panel. Because there was no substantial improvement in detection with the one-repeat condition, these results are shown in Supplementary Figures S2 to S4. Extra sum-of-squares *F*-tests applied to each condition showed that no single asymmetric sigmoidal function fit all of the data ($P < 0.0001$ for all); thus, there were significant differences in the times at which cases were detected. We show the individual curves

for each progression rate separated by follow-up intervals (6 months, 1 year, and 2 years) in Supplementary Figures S2 to S4.

The time (in years) to detect 95% of cases within the cohort was plotted as a function of the number of visual field tests per visit for each condition (Fig. 3). Similarly, the number of cases detected at 2 years was also plotted as a function of tests per visit for each progression rate (-2 , -1 , and -0.5 dB/yr, cumulative cases detected) and for false positives (at the 2-year visit) (Fig. 4). Because the 2-year follow-up interval resulted in significant delays in case detection, especially when one visual field test was done per visit, we have not shown these data in Figures 3 and 4. Several common themes are evident from both of the figures. First, shorter intervals between follow-up visits, faster progression rates, and higher reliability rates reduced the time to detect mean deviation changes and increased the number of cases detected at 2 years. There was some improvement in case detection when the “one in hand” approach (repeat testing of unreliable results) was used, but the difference between the one test per visit and the frontloaded approaches remained similar (Supplementary Figs. S2–S4).

Second, there was a notable plateau effect in benefit with an increasing number of visual field tests performed per visit, with two to four tests per visit demonstrating similar times to detection and

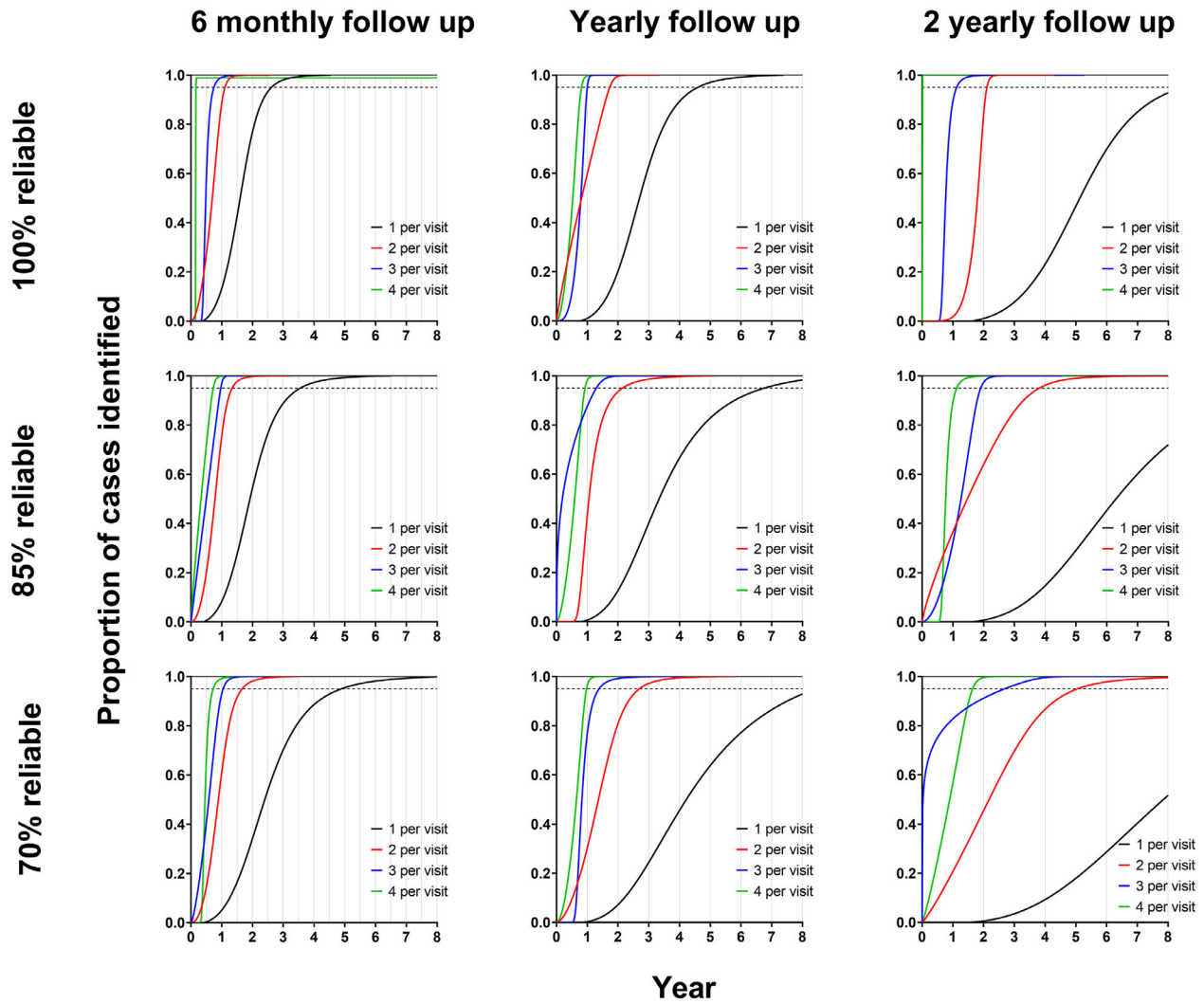


Figure 2. Proportion of cases identified as having a significant mean deviation change as a function of the follow-up year progressing at -2 dB per year. Each column represents a different follow-up interval (6 months, 1 year, and 2 years), and each row represents a different reliability condition (100% reliable, 85% reliable, and 70% reliable); note that the 85% reliable but allowing one repeat test and 70% reliable but allowing one repeat test did not show significant differences and are therefore shown in the Supplementary Figures. The curves indicate the asymmetric sigmoidal function through the simulated data, with each curve indicating a different number of visual field tests conducted at each visit (*black*, one per visit; *red*, two per visit; *blue*, three per visit; *green*, four per visit). Because each column indicates a different follow-up interval, the *gray vertical lines* identify points at which an actual clinical visit took place, allowing for like-for-like comparisons across the conditions. Therefore, the sigmoidal functions are not representative of granular clinical visits but rather present a clearer method for visualizing the data. The *black dotted line* indicates the point at which 95% of the cohort was detected as progressing. Each x-axis has been truncated to year 8 to expand the range of visible data. The truncation resulted in some of the y-values being hidden but illustrates the difference in detection rates at the specified time points.

proportions of case detection. For example, close to 100% of cases are predicted to be detected with two, three, or four tests per visit for -2 -dB/yr progressors that demonstrate 100% test reliability. The benefits to performing more than two tests per visit were most apparent under conditions of lower rates of reliability, among slower progressors, and where longer intervals between visits occur. The overall false positive rates were near the predicted rate of 2.5% as described in

Methods. The rates were similar across conditions, with a tendency for slightly lower rates with more tests performed.

Distribution of Progression Rate for Each Visual Field Test Paradigm

The distribution of slopes for the combinations of reliability, follow-up interval, and progression rates are

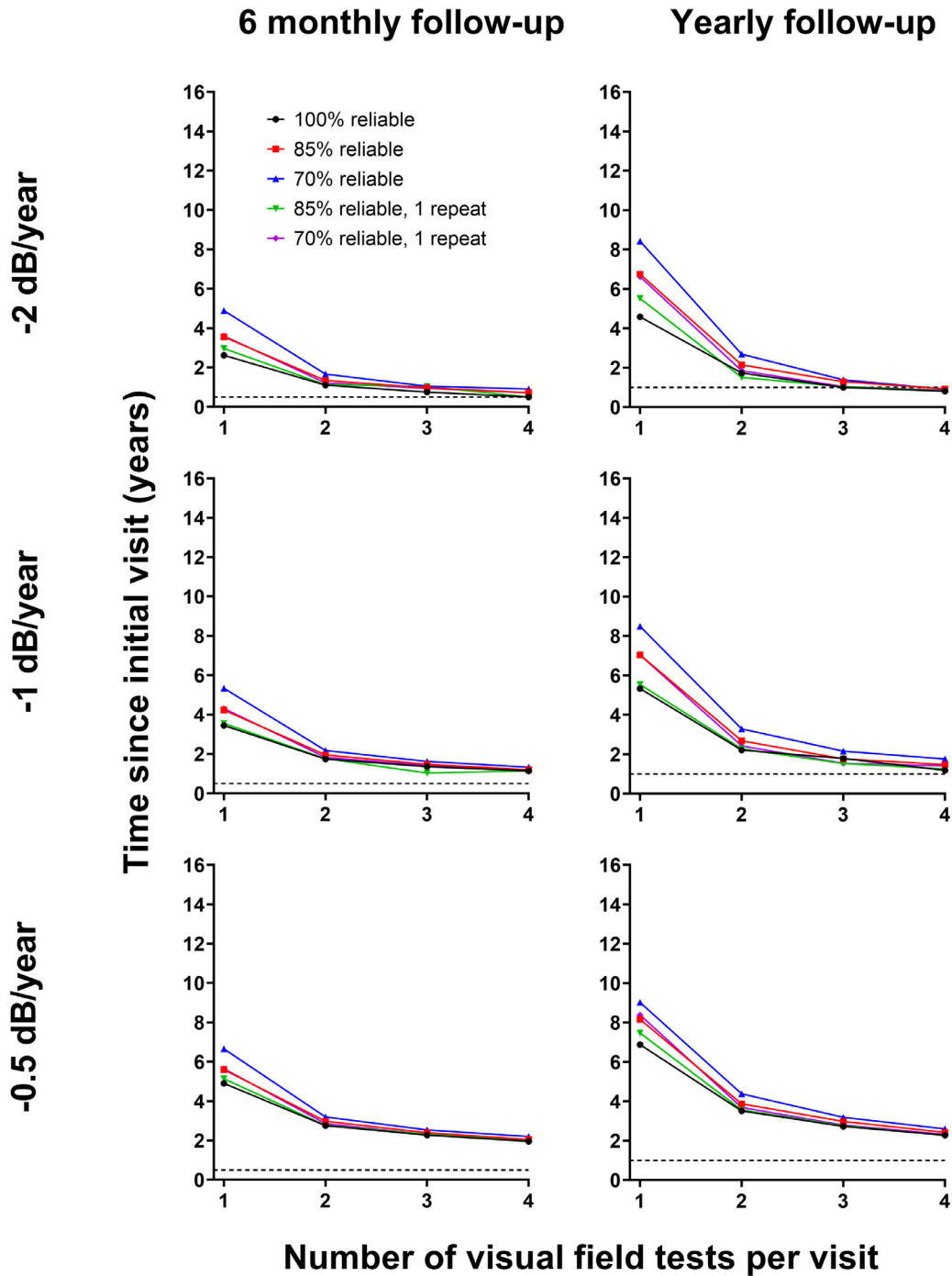


Figure 3. Time since the initial visit after which mean deviation change is identified in 95% of the cohort (years) as a function of the number of visual field tests per visit. Each column represents a different follow-up interval (6 months and 1 year; the 2-year results are not shown due to low detection rates), and each row represents a different progression rate (-2 dB/yr, top; -1 dB/yr, middle; -0.5 dB/yr, bottom). Each symbol and line within each panel indicates the result for a different reliability condition, allowing them to be compared across each progression rate and follow-up condition. A higher y-value indicates longer time since the initial visit at which the mean deviation change was identified. The black dashed line in each panel indicates the first time since the initial visit and thus the first time point at which detection was possibly identified using linear regression of mean deviation scores (except for one visual field test per year, which requires at least three visits).

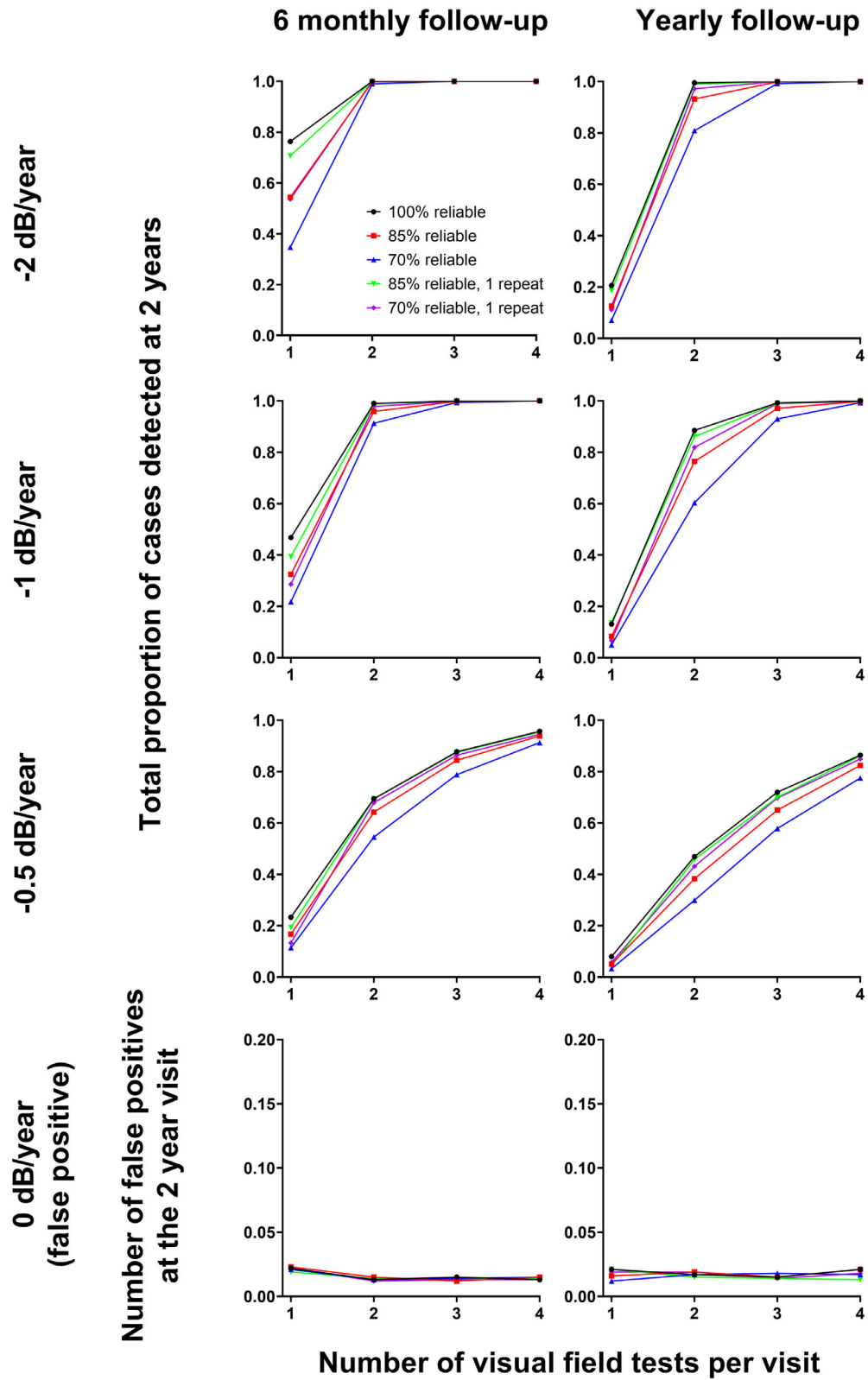


Figure 4. Proportion of cases detected at 2 years as a function of the number of visual field tests per visit. Each column, row, symbol, and line depicts progression rate, follow-up interval and reliability conditions as per Figure 3. A higher y-value indicates more cases detected, and a lower y-value indicates fewer cases detected. Note that, although the true “progressing” cases were reported as a cumulative (total) proportion of cases detected at 2 years, the false positive proportions are shown as the number detected at the 2-year visit (not cumulative). The false positive panels (*bottom*) have also had their y-axes adjusted for clarity.

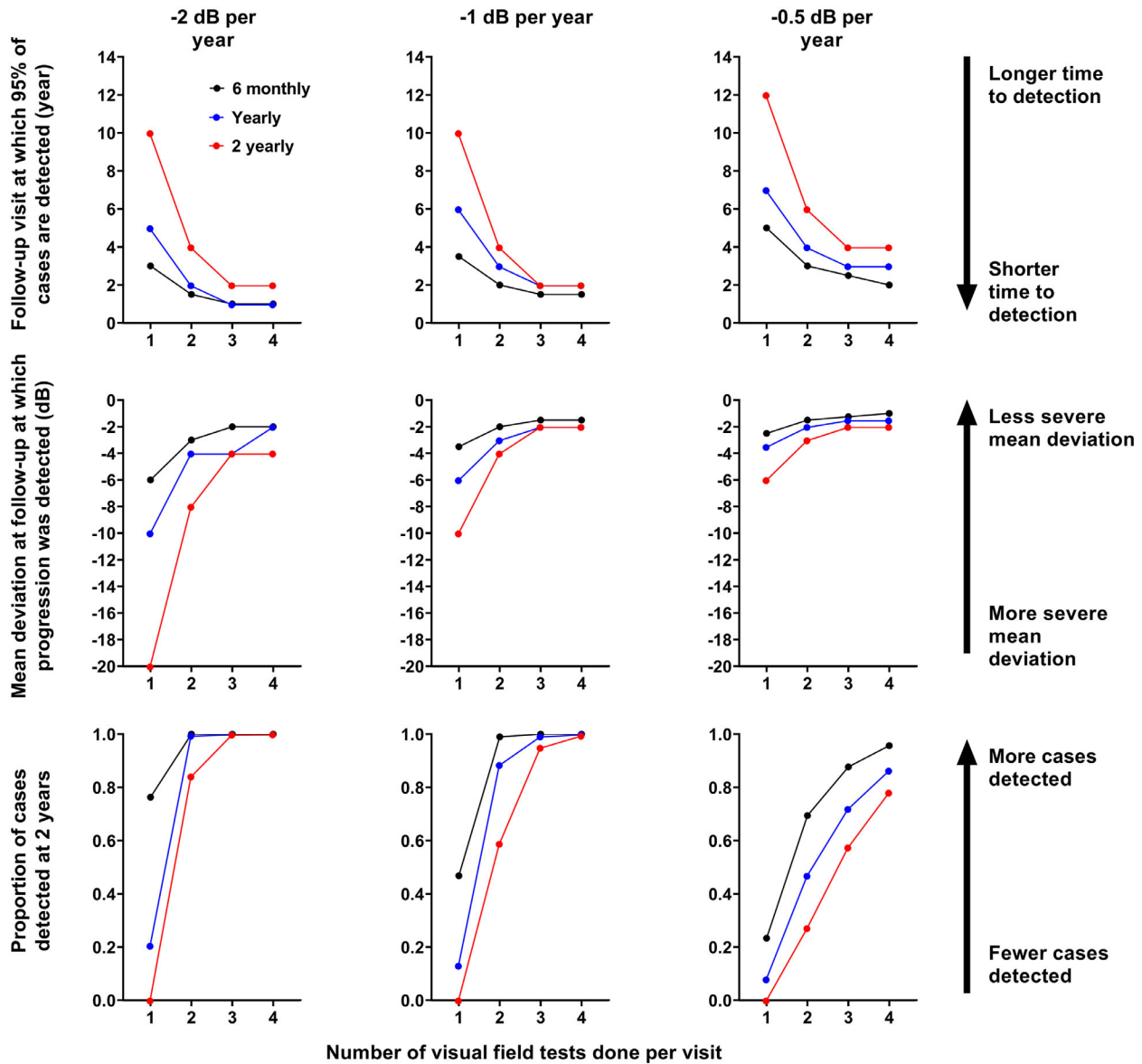


Figure 5. A summary of key clinical parameters of interest. (Top row) The practical follow-up visit at which 95% of cases were detected (year), where a higher y-value indicates a longer time to detection of change. (Middle row) The mean deviation at the follow-up visit at which change was detected (dB), where a higher y-value indicates less mean deviation loss and a lower y-value indicates more mean deviation loss at the time of change detection. (Bottom row) The proportion of cases detected at 2 years, where a high y-value indicates more cases and a lower y-value indicates fewer cases detected. The three progression rates are shown in each column. As the follow-up visit time point was determined by the interval, in practice this constrains the time at which detection can be identified (e.g., 6 months, 1 year, 2 years, and so on). The mean deviation prediction was calculated by multiplying the follow-up visit by the progression rate. Within each panel, the three follow-up intervals are shown (black, 6 months; blue, 1 year; red, 2 years).

shown in Supplementary Figures S5 to S7. When only one test was done per visit, the distributions of slope values were flatter and wider in comparison to when at least two visual field tests were done per visit. Under conditions of greater rates of low test reliability, the distributions became wider across all test conditions. The distributions for non-progressing patients (i.e., false positives) are shown in Supplementary Figure S8,

which demonstrates a similar tendency to be wider under conditions of low test reliability.

Average Test Time Per Session

Across the simulated cohort, the average test times per eye per session were 2.7 minutes, 5.5 minutes,

8.2 minutes, and 11.0 minutes for one, two, three, and four tests per visit, respectively.

Summary of Clinically Relevant Parameters

Figure 5 provides a summary of three key clinically relevant parameters: (1) the actual predicted follow-up visits at which 95% of cases are detected (rounded up to the next nearest time point); (2) the mean deviation values at which changes were detected; and (3) proportion of cases detected at 2 years. Following from Figures 3 to 5, there was generally little difference when doing two, three, or four visual field tests per visit, but at least two tests performed per visit offered time advantages over only one across all progression rates and resulted in less mean deviation loss at detection, most pronounced with the 2-year follow-up interval. However, mean deviation loss before detection of change was less when using a 6-month follow-up interval compared with longer intervals, and there was little difference among two, three, or four visual field tests per visit due to the longer time elapsed between follow-up visits. Although doing one or two tests per visit resulted in fewer cases detected at 2 years for slow progressors (-0.5 dB/yr) compared with three or four tests per visit, the eventual point of detection was similar with small differences in mean deviation severity.

The combination of results shown in Figures 2 to 5 therefore suggests that performing two visual field tests per visit appears to offer a practical method for testing that largely preserves diagnostic ability; this approach had the least severe mean deviation score at the visit where change was detected. Table 3 provides a summary of key parameters when comparing one or two visual field tests performed per visit (a complete list of the outcome parameters is shown in Supplementary Tables S1 and S2). Due to the complexity of the table, we also provide an encoded Excel spreadsheet calculator that allows a quick comparison of expected outcomes for one to four tests per visit with the input variables assessed in the present study (available for download online).

Discussion

We conducted a computer simulation study that comprehensively described variations in the number of visual field tests performed per visit for optimizing detection of glaucoma functional change from its earliest stages (at 0 dB) across different clinically determined progression rates and reliability levels for different follow-up intervals. Obtaining two or more

results per clinical visit significantly shortens the time required for case detection compared with one test per visit, as is conventionally performed in clinical practice when using SITA-Faster. Specifically, we note that these benefits were pronounced in situations where the results returned by the perimetric algorithm may not meet a predefined set of reliability criteria, and where such results may be excluded from trend analysis. There are diminishing returns in case detection with more visual field tests performed per visit, with two performed per visit having detection rates similar to those for three or four being performed per visit.

Is More Better? Conducting More Visual Field Tests Per Visit to Overcome Testing Limitations

Visual field testing has several well-known limitations, many of which stem from its dependence upon the patient's ability to reliably perform the test to obtain high-fidelity and accurate results. Factors such as procedural learning,^{25,26} fatigue,^{27,28} attention,^{12,29,30} and disease severity³¹ play important roles in the variability and repeatability of the sensitivity measurements. As such, intersession and intrasession variance can mask true changes in visual field sensitivity. The frontloading approach addresses limitations pertaining to variability by creating robust data points consisting of the average of several test results.

The previous clustering approach proposed by Crabb and Garway-Heath illustrated the benefits of grouping data points at two "ends" of the follow-up interval compared to even, clinically routine follow-up.⁷ The next issue is whether the benefits of frontloading may be further enhanced by shortening the follow-up interval, such as was suggested by Wu et al.²¹ The most notable benefit of shortening the follow-up interval from 2 years to 6 months was the much less severe mean deviation loss measured at the case detection visit, similar to the results found by Anderson and colleagues.³² Thus, clinicians should be cognizant of the expected mean deviation result at the end of the follow-up interval they use.

The benefits conferred by the combination of more tests per visit and more frequent reviews did not come at the cost of higher rates of false positives. The slightly more positive mean sensitivity results (median, 0.34 dB) (Supplementary Fig. S1A) returned by visual field tests after the first result within the same visit in the frontloading conditions meant that false positives were less likely to be detected when performing multiple tests per visit. Analogously, a similar (albeit less noticeable) tendency for quantitatively shorter follow-up

Table 3. Summary of Key Clinical Parameters for Each Combination of Progression Rate, Reliability Rate, and Follow-Up Interval

Progression Rate	Reliability	Number of Visual Field Tests per Follow-Up	Cases Detected at 2 Years			Proportion of Patients With Unreliable Results That Prevented Progression Analysis at 2 Years			Follow-Up Visit at Which 95% of Cases Were Detected			Mean Deviation at the Point of Progression Detection of 95% of Cases (dB)		
			6M	1Y	2Y	6M	1Y	2Y	6M	1Y	2Y	6M	1Y	2Y
-2 dB per year	100%	1	0.763	0.206	0.000	0	0	0	6	5	5	-6	-10	-20
		2	1.000	0.995	0.842	0	0	0	3	2	2	-3	-4	-8
		1	0.544	0.126	0.000	0.003	0.062	0.285	7	7	6	-7	-14	-24
	85%	2	0.998	0.932	0.640	0	0	0.024	3	3	2	-3	-6	-8
		1	0.347	0.070	0.000	0.03	0.217	0.513	10	9	8	-10	-18	-32
		2	0.990	0.809	0.455	0	0.007	0.089	4	3	3	-4	-12	-12
	85%, 1 repeat	1	0.707	0.186	0.000	0	0.001	0.048	6	6	5	-6	-12	-20
		2	1.000	0.990	0.812	0	0	0.0001	3	2	2	-3	-4	-8
		1	0.536	0.112	0.000	0.0001	0.024	0.175	8	7	6	-8	-14	-24
	70%, 1 repeat	2	1.000	0.972	0.738	0	0	0.004	3	2	2	-3	-4	-8
		1	0.468	0.131	0.000	0	0	0	7	6	5	-3.5	-6	-10
		2	0.990	0.885	0.589	0	0	0	4	3	2	-3	-4	-8
85%	1	0.325	0.083	0.000	0.002	0.056	0.279	9	7	7	-4.5	-7	-14	
	2	0.959	0.764	0.433	0	0.0002	0.024	4	3	3	-2	-3	-6	
	1	0.218	0.050	0.000	0.027	0.215	0.517	11	9	8	-5.5	-9	-16	
70%	2	0.913	0.603	0.304	0	0.005	0.09	5	4	3	-2.5	-4	-6	
	1	0.393	0.136	0.000	0	0.001	0.045	6	6	6	-4	-6	-12	
	2	0.988	0.861	0.562	0	0	0.0002	4	3	2	-2	-3	-4	
85%, 1 repeat	1	0.286	0.067	0.000	0.001	0.019	0.177	9	8	6	-4.5	-8	-12	
	2	0.978	0.819	0.507	0	0	0.003	4	3	2	-2	-3	-4	
	1	0.233	0.080	0.000	0	0	0	10	7	6	-2.5	-3.5	-6	
-0.5 dB per year	100%	2	0.694	0.469	0.272	0	0	0	6	4	3	-1.5	-2	-3
		1	0.167	0.051	0.000	0.002	0.285	0.275	12	9	7	-3	-4.5	-7
		2	0.642	0.383	0.212	0	0.024	0.023	6	4	3	-1.5	-2	-3
85%	2	0.114	0.033	0.000	0.033	0.513	0.518	14	9	8	-3.5	-4.5	-8	
	1	0.545	0.299	0.153	0	0.089	0.083	7	5	4	-1.75	-2.5	-4	
	2	0.694	0.458	0.279	0	0.048	0.044	11	8	7	-2.8	-4	-7	
85%, 1 repeat	1	0.193	0.045	0.000	0	0.0001	0.001	6	4	3	-1.5	-2	-3	
	2	0.694	0.458	0.279	0	0.048	0.044	11	8	7	-2.8	-4	-7	
	1	0.133	0.057	0.000	0.0004	0.175	0.171	12	9	7	-3	-4.5	-7	
70%, 1 repeat	2	0.679	0.431	0.239	0	0.004	0.004	6	4	3	-1.5	-2	-3	
	1	0.022	0.021	0.000	0	0	0	—	—	—	—	—	—	
	2	0.013	0.017	0.017	0	0	0	—	—	—	—	—	—	
0 dB per year (false positives)	100%	2	0.023	0.016	0.000	0.055	0.059	0.276	—	—	—	—	—	—
		1	0.015	0.019	0.020	0.008	0.001	0.022	—	—	—	—	—	—
		2	0.021	0.012	0.000	0.216	0.217	0.502	—	—	—	—	—	—
85%	1	0.013	0.017	0.011	0.0003	0.005	0.087	—	—	—	—	—	—	
	2	0.019	0.021	0.000	0.001	0.002	0.044	—	—	—	—	—	—	
	1	0.014	0.015	0.019	0	0	0.0002	—	—	—	—	—	—	
70%, 1 repeat	2	0.022	0.019	0.000	0.022	0.026	0.0176	—	—	—	—	—	—	
	1	0.012	0.019	0.025	0	0	0.004	—	—	—	—	—	—	
	2	0.012	0.019	0.025	0	0	0.004	—	—	—	—	—	—	

As described in the text, this table compares the results for one and two visual field tests performed per visit. For a complete list of results, including three and four visual field tests performed per visit, refer to Supplementary Table S1. 6M, 6 months; 1Y, 1 year; 2Y, 2 years.

intervals producing lower false positive rates was due to the slight skew in the intersession mean deviation in the positive direction (0.04 dB) (Supplementary Fig. S1B). Situations with a propensity for fewer data points for performing regression analysis (such as low test reliability and longer test intervals) tended to have relatively higher false positive rates (close to the predicted 2.5% value) at the 2-year visit, but the overall differences in false positive rates across conditions was small and were not likely clinically significant. The discussion below further highlights the consequences of data exclusion.

More Testing Overcomes the Impact of Low Test Reliability

Previous models predicting the number of visual field tests required for detecting glaucomatous change have identified test variability as a potential source for error.⁵ More recently, there have been reports that as many as 30% of SITA-Faster visual field tests return results that do not meet specific, predefined “reliability” criteria.^{8,11,12} It is important to note that such criteria evolve with time and the available evidence. For example, recent work by Heijl and colleagues¹⁷ demonstrated the minimal contribution of elevated false positive rates to measurement variability, and the authors recommended a revision to historical cut-off values used as an indicator to discard results. Therefore, when identifying issues pertaining to the usability of clinical data, clinicians need to exercise caution and recognize the importance of data fidelity.

Irrespective of the criteria applied to determine reliability, the presence of results that do not meet specific reliability criteria consequently delay change identification across all test paradigms. As expected, the probability of correctly identifying visual field mean deviation change increases with the number of tests performed by exercising a “brute force” strategy of overcoming the anticipated frequency of results discarded due to specific reliability criteria.

In practice, clinicians and their technicians are trained to identify signs of low test reliability and, upon seeing such results, have a tendency to repeat the test to attempt to obtain a more reliable result (simulated by using the “one in hand” approach). However, although this condition slightly improved the proportion of case detection under situations of excluded data, the large differences in delays in case detection between the one test per visit condition and frontloaded paradigms remained similar. One interpretation is that criteria for discarding potentially useful data may have to be revised, as injudicious exclusion

of clinical data may lead to delays in case detection. The practical concerns of “reliability” in the results returned by SITA-Faster relate to the interpretation of its reliability metrics, often conducted automatically and with arbitrary cut-offs,¹⁷ which can unfortunately cause automated analyses to exclude data that may still be useful. Overall, these models suggest that an approach of consistently performing at least two visual field tests per visit may provide benefits across most patients. A consistent intrasession repeat testing approach remains clinically efficient and well-tolerated by patients and technicians,³³ with the present model suggesting only 5.5 minutes for two tests per eye, similar to the times that we have previously reported.¹¹

Diminishing Returns With More Intrasession Tests: Practical Recommendations

As expected, a brute-force approach improves case detection. However, the benefit of multiple tests per visit for case detection and mean deviation at detection appeared to diminish beyond two field tests per session. As we have previously shown, most patients who are glaucoma suspect or have manifest glaucoma are able to complete visual field testing twice per eye within 15 to 20 minutes¹¹; thus, instead of attempting three or more visual field tests per eye, it appears to be most practical to perform two tests to achieve similar results in detection.

In addition to a practical recommendation regarding tests per visit, this framework allows clinicians to exercise prudence in determining the follow-up interval, especially relevant in the context of the Covid-19 pandemic. By increasing the likelihood of detecting cases over fewer clinical visits or shorter overall follow-up time, a frontloading approach may confer further benefits of reducing unnecessary review appointments for patients with suspected or manifest glaucoma.

Limitations

Similar to previous modeling work,⁷ the present study makes several assumptions for the purposes of this modeling exercise (see Table 1). For example, the model assumes that glaucoma functional change measured using the mean deviation result is linear and monotonic. In reality, patients may demonstrate abrupt change following an exponential decay due to the course of their disease and its management. Nonlinear, exponential worsening of visual field indices has been previously discussed by others.^{34,35} In the context of our models, it would follow that nonlinear models may predict more severe long-term visual field loss, and this

would expectedly favor shorter follow-up intervals to “catch” abrupt and rapid changes before they significantly affect the patient. Similar to other modeling work, this requires empirical evaluation to determine its real-world validity and translation.

Another assumption made is that the variability characteristics for two frontloaded visual field tests can be extrapolated to more than two tests. It is possible that additional variability factors such as fatigue may have magnified contributions. If that were the case, we postulate that the diminishing returns with more than two results per visit would be more apparent, supporting our main finding that two frontloaded visual fields may be sufficient. This hypothesis would require a different study design to fully test.

Our model used patients with suspected or early glaucoma with an assigned mean deviation score of 0 dB as they progressed toward moderate and more advanced stages of glaucoma, as this cohort of patients is most commonly seen in clinical practice,³⁶ and among these patients changes in a 24-2 global index such as mean deviation may be more relevant. In more advanced stages of glaucoma, more variables must be considered, such as increased variability at pointwise locations and the measurement floor effect,³¹ as well as the potential need to change to alternative testing grids such as the 10-2.³⁷ Despite its limitations, mean deviation remains a metric that is correlated with other important patient-related outcomes, such as quality of life.³⁸ Other useful progression analysis methods may focus on pointwise change (such as change probability maps),^{39,40} in part captured by the mean deviation slope, but this is a focus of other simulation studies. As described in Methods, modeling the correlations in pointwise sensitivity may be useful for deriving robust mean deviation scores¹⁵ and may also advantageously reflect a greater diversity of scotomata expected in glaucoma.

Furthermore, mean deviation variability estimates were derived from the overall population, rather than from the individual patient. Understanding the variability characteristics at the individual level may have an impact on estimates of change trajectory, due to the heterogeneity in variability within the population.²³ In real-world practice, this could be obtained from patients with pre-existing clinical data. We note that, as a framework, the variables in the present model can be adjusted to further assess different levels of mean deviation and their associated variability and unreliability rates.

Although we report on practical recommendations for number of tests per visit and visit interval, there are jurisdictional differences in recommendations for eye examination frequency,^{2,41,42} which at times may

not align with the intervals that we examined, including the need to perform other glaucoma-related tests that may not be coincident with visual field testing. Our goal was to provide guidelines as to when clinicians might predict a significant change to occur based on the progression rate and number of tests conducted.

We assessed a commercially available algorithm, SITA-Faster, for the implementation of frontloading. By understanding the benefits of having multiple sensitivity readings within a session, future alternative thresholding algorithms could incorporate methods for “interleaving” presentations to return such data, without the need to repeat the test. This is an area of ongoing interest in developments in perimetry.

Conclusions

Our model provides guidance for SITA-Faster visual field testing schedules to overcome limitations arising from test variability and low test reliability when identification of a specific mean deviation rate of change is used as an endpoint of interest. This guidance can be titrated based on the clinician’s need by referring to our Excel spreadsheet comparing key variables assessed in the present study.⁴³ The combination of findings suggests that two visual field tests conducted per clinical visit spaced 6 months apart appears to be practical and achievable, in addition to being capable of case detection of mean deviation change sooner and with less mean deviation loss compared to performing only one test per visit.

Acknowledgments

Supported in part by a grant from National Health and Medical Research Council Ideas (1186915 to MK and JP), and a University of New South Wales Faculty of Science Early Career Academics Network Seeding Grant (JP). Guide Dogs NSW/ACT provided funding for the clinical services enabling data collection for this study. The funding organization had no role in the design or conduct of this research.

Disclosure: **J. Phu**, None; **M. Kalloniatis**, None

References

1. Phu J, Agar A, Wang H, Masselos K, Kalloniatis M. Management of open-angle glaucoma by primary eye-care practitioners: toward a

- personalised medicine approach. *Clin Exp Optom.* 2021;104(3):367–384.
2. Prum BE, Jr, Rosenberg LF, Gedde SJ, et al. Primary Open-Angle Glaucoma Preferred Practice Pattern guidelines. *Ophthalmology.* 2016;123(1):P41–P111.
 3. EGS Foundation. European Glaucoma Society Terminology and Guidelines for Glaucoma, 4th Edition - Chapter 3: treatment principles and options supported by the EGS Foundation: Part 1: foreword; introduction; glossary; Chapter 3 Treatment principles and options. *Br J Ophthalmol.* 2017;101(6):130–195.
 4. De Moraes CG, Liebmann JM, Levin LA. Detection and measurement of clinically meaningful visual field progression in clinical trials for glaucoma. *Prog Retin Eye Res.* 2017;56:107–147.
 5. Chauhan BC, Garway-Heath DF, Goni FJ, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol.* 2008;92(4):569–573.
 6. Fung SS, Lemer C, Russell RA, Malik R, Crabb DP. Are practical recommendations practiced? A national multi-centre cross-sectional study on frequency of visual field testing in glaucoma. *Br J Ophthalmol.* 2013;97(7):843–847.
 7. Crabb DP, Garway-Heath DF. Intervals between visual field tests when monitoring the glaucomatous patient: wait-and-see approach. *Invest Ophthalmol Vis Sci.* 2012;53(6):2770–2776.
 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;208:251–264.
 9. Heijl A, Patella VM, Chong LX, et al. A new SITA perimetric threshold testing algorithm: construction and a multicenter clinical study. *Am J Ophthalmol.* 2019;198:154–165.
 10. Lavanya R, Riyazuddin M, Dasari S, et al. A comparison of the visual field parameters of SITA-Faster and SITA-Standard strategies in glaucoma. *J Glaucoma.* 2020;29(9):783–788.
 11. 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.
 12. 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;221:115–130.
 13. Ismail R, Azuara-Blanco A, Ramsay CR. Consensus on outcome measures for glaucoma effectiveness trials: results from a Delphi and nominal group technique approaches. *J Glaucoma.* 2016;25(6):539–546.
 14. Mills RP, Budenz DL, Lee PP, et al. Categorizing the stage of glaucoma from pre-diagnosis to end-stage disease. *Am J Ophthalmol.* 2006;141(1):24–30.
 15. 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(3):22.
 16. 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(7):795–804.
 17. Heijl A, Patella VM, Flanagan JG, et al. False positive responses in standard automated perimetry [published online ahead of print]. *Am J Ophthalmol.* 2021, <https://doi.org/10.1016/j.ajo.2021.06.026>.
 18. Blumenthal EZ, Sample PA, Berry CC, et al. Evaluating several sources of variability for standard and SWAP visual fields in glaucoma patients, suspects, and normals. *Ophthalmology.* 2003;110(10):1895–1902.
 19. Artes PH, Hutchison DM, Nicoleta MT, et al. Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma. *Invest Ophthalmol Vis Sci.* 2005;46(7):2451–2457.
 20. Phu J, Kalloniatis M. Ability of 24-2C and 24-2 grids to identify central visual field defects and structure-function concordance in glaucoma and suspects. *Am J Ophthalmol.* 2020;219:317–331.
 21. Wu Z, Saunders LJ, Daga FB, et al. Frequency of testing to detect visual field progression derived using a longitudinal cohort of glaucoma patients. *Ophthalmology.* 2017;124(6):786–792.
 22. Susanna FN, Melchior B, Paula JS, et al. Variability and power to detect progression of different visual field patterns [published online ahead of print]. *Ophthalmol Glaucoma.* 2021, <https://doi.org/10.1016/j.ogla.2021.04.004>.
 23. Montesano G, Quigley HA, Crabb DP. Improving the power of glaucoma neuroprotection trials using existing visual field data. *Am J Ophthalmol.* 2021;229:127–136.
 24. Phu J, Tong J, Zangerl B, et al. Cluster analysis reveals patterns of age-related change in anterior chamber depth for gender and ethnicity: clinical implications. *Ophthalmic Physiol Opt.* 2020;40(5):632–649.

25. Gardiner SK, Demirel S, Johnson CA. Is there evidence for continued learning over multiple years in perimetry? *Optom Vis Sci*. 2008;85(11):1043–1048.
26. Wild JM, Dengler-Harles M, Searle AE, O'Neill EC, Crews SJ. The influence of the learning effect on automated perimetry in patients with suspected glaucoma. *Acta Ophthalmol (Copenh)*. 1989;67(5):537–545.
27. Heijl A. Time changes of contrast thresholds during automatic perimetry. *Acta Ophthalmol (Copenh)*. 1977;55(4):696–708.
28. Heijl A, Drance SM. Changes in differential threshold in patients with glaucoma during prolonged perimetry. *Br J Ophthalmol*. 1983;67(8):512–516.
29. Phu J, Kalloniatis M, Khuu SK. Reducing spatial uncertainty through attentional cueing improves contrast sensitivity in regions of the visual field with glaucomatous defects. *Transl Vis Sci Technol*. 2018;7(2):8.
30. Wall M, Woodward KR, Brito CF. The effect of attention on conventional automated perimetry and luminance size threshold perimetry. *Invest Ophthalmol Vis Sci*. 2004;45(1):342–350.
31. 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(7):1359–1369.
32. Anderson AJ, Asokan R, Murata H, Asaoka R. Detecting glaucomatous progression with infrequent visual field testing. *Ophthalmic Physiol Opt*. 2018;38(2):174–182.
33. Phu J, Kalloniatis M. Patient and technician perspectives following the introduction of frontloaded visual field testing in glaucoma assessment [published online ahead of print]. *Clin Exp Optom*. 2021, <https://doi.org/10.1080/08164622.2021.1965461>.
34. Pathak M, Demirel S, Gardiner SK. Nonlinear, multilevel mixed-effects approach for modeling longitudinal standard automated perimetry data in glaucoma. *Invest Ophthalmol Vis Sci*. 2013;54(8):5505–5513.
35. Pathak M, Demirel S, Gardiner SK. Nonlinear trend analysis of longitudinal pointwise visual field sensitivity in suspected and early glaucoma. *Transl Vis Sci Technol*. 2015;4(1):8.
36. Boodhna T, Crabb DP. Disease severity in newly diagnosed glaucoma patients with visual field loss: trends from more than a decade of data. *Ophthalmic Physiol Opt*. 2015;35(2):225–230.
37. Rao HL, Begum VU, Khadka D, Mandal AK, Senthill S, Garudadri CS. Comparing glaucoma progression on 24-2 and 10-2 visual field examinations. *PLoS One*. 2015;10(5):e0127233.
38. McKean-Cowdin R, Varma R, Wu J, Hays RD, Azen SP, Los Angeles Latino Eye Study Group. Severity of visual field loss and health-related quality of life. *Am J Ophthalmol*. 2007;143(6):1013–1023.
39. Heijl A, Leske MC, Bengtsson B, Bengtsson B, Hussein M, Early Manifest Glaucoma Trial Group. Measuring visual field progression in the Early Manifest Glaucoma Trial. *Acta Ophthalmol Scand*. 2003;81(3):286–293.
40. Garway-Heath DF, Crabb DP, Bunce C, et al. Latanoprost for open-angle glaucoma (UKGTS): a randomised, multicentre, placebo-controlled trial. *Lancet*. 2015;385(9975):1295–1304.
41. National Health and Medical Research Council. NHMRC guidelines for the screening, prognosis, diagnosis, management and prevention of glaucoma. Available at: <https://healthinfonet.ecu.edu.au/key-resources/resources/20018/>. Accessed October 22, 2021.
42. National Institute for Health and Care Excellence. *Glaucoma: Diagnosis and Management*. London, UK: National Institute for Health and Care Excellence; 2019.
43. Phu J, Kalloniatis. MD Change Calculator - Excel Spreadsheet. CFEH Research Tools and Downloads. Centre for Eye Health, Sydney, Australia. Available online: <https://www.centreforeyehealth.com.au/research-tools-and-downloads/>.