Factors That Influence Standard Automated Perimetry Test Results in Glaucoma: Test Reliability, Technician Experience, Time of Day, and Season

Francisco G. Junoy Montolio,[1] *Christiaan Wesselink,*[1] *Marijke Gordijn,*[2] *and Nomdo M. Jansonius*[1,3]

PURPOSE. To determine the influence of several factors on standard automated perimetry test results in glaucoma.

METHODS. Longitudinal Humphrey field analyzer 30-2 Swedish interactive threshold algorithm data from 160 eyes of 160 glaucoma patients were used. The influence of technician experience, time of day, and season on the mean deviation (MD) was determined by performing linear regression analysis of MD against time on a series of visual fields and subsequently performing a multiple linear regression analysis with the MD residuals as dependent variable and the factors mentioned above as independent variables. Analyses were performed with and without adjustment for the test reliability (fixation losses and false-positive and false-negative answers) and with and without stratification according to disease stage (baseline MD).

RESULTS. Mean follow-up was 9.4 years, with on average 10.8 tests per patient. Technician experience, time of day, and season were associated with the MD. Approximately 0.2 dB lower MD values were found for inexperienced technicians ($P < 0.001$), tests performed after lunch ($P < 0.001$), and tests performed in the summer or autumn ($P < 0.001$). The effects of time of day and season appeared to depend on disease stage. Independent of these effects, the percentage of false-positive answers strongly influenced the MD with a 1 dB increase in MD per 10% increase in false-positive answers.

CONCLUSIONS. Technician experience, time of day, season, and the percentage of false-positive answers have a significant influence on the MD of standard automated perimetry. (*Invest Ophthalmol Vis Sci.* 2012;53:7010–7017) DOI:10.1167/iovs.12-10268

Standard automated perimetry (SAP) is an invaluable test for the diagnosis and monitoring of glaucoma. However, because SAP is a subjective test that aims to measure a sensitivity threshold in a living organism, SAP test results are prone to variability. Knowing the sources of the variability might enable reducing it, and this may improve progression detection in glaucoma. With less variability, smaller changes can be picked up and fewer tests are needed.

The variability of SAP test results depends on many factors. A well-known factor is disease stage.[1,2] Other factors are patient motivation and technician performance, the latter via appropriateness of refraction and patient instruction, reassurance before the test, and patient monitoring during the test.[3-5]

Circadian rhythms and seasonal influences may also contribute to the variability. These periodicities may influence SAP test results in at least two different ways. First, patients need a good cognitive function to perform perimetry. In an extensive review, Blatter et al.[6] described a daily variation in cognitive function in humans. Cognitive function deteriorated over the day in the elderly (at the typical age of a glaucoma patient), whereas the opposite is the case in young subjects. Second, SAP measures a threshold sensitivity of the visual system. Earlier studies in humans showed a daily variation in retinal visual sensitivity, suggesting lower sensitivities in the early morning.[7-9] Besides a daily variation, a seasonal (circannual) variation might also be present. After all, exposure to light varies per season and prior light history affects light sensitivity.[10] Indeed, retinal sensitivity seems to be highest in the spring.[11,12]

Variability is entangled with reliability, which is commonly assessed by reliability indices and displayed as the percentages of fixation losses (FL) and false-positive (FP) and false-negative (FN) answers. Two issues concerning reliability are relevant to this study. First, whether reliability as assessed by the reliability indices influences variability. Second, whether other factors that influence variability act through the reliability or have their own, independent influence.

The aim of this study was to determine the influence of several factors on SAP test results in glaucoma. For this purpose, we analyzed the influence of technician experience, time of day, season, and the percentages of FL and FP and FN answers on the mean deviation (MD), a commonly used summary measure of SAP test results.

METHODS

Study Population

The present study was performed using data from the Groningen Longitudinal Glaucoma Study (GLGS), a prospective observational cohort study performed in a clinical setting.[13,14] In the GLGS, all glaucoma patients and glaucoma suspects who visited the glaucoma

outpatient service of the University Medical Center Groningen between July 1, 2000 and June 30, 2001, and who provided informed consent, were included in an observational study with conventional perimetry, frequency-doubling perimetry (frequency-doubling technology [FDT]; Carl Zeiss Meditec AG, Jena, Germany), and laser polarimetry (GDx; Laser Diagnostic Technologies, San Diego, CA). The study comprised both glaucoma patients and glaucoma suspect patients and was originally designed as a health technology assessment study. The original aim was to determine if it was possible to replace, in glaucoma patients and/or glaucoma suspects, the lengthy and cumbersome conventional perimetry by FDT and/or GDx. After the initial health technology assessment study, we continued performing conventional perimetry in glaucoma patients and moved to FDT/GDx in glaucoma suspects in our regular care. The GLGS continued as an ongoing anonymous gathering of all information from all glaucoma patients and glaucoma suspects obtained during regular care.

For the present study, we used data from a subpopulation of the GLGS cohort: we included only patients who had (1) glaucoma at baseline (for criteria see the following text), (2) series of at least eight (nine with discarded learning test) visual field tests made with standard automated perimetry (Humphrey Field Analyzer [HFA]; Carl Zeiss Meditec Inc., Dublin, CA), and (3) a follow-up of at least 5 years.[15] The study protocol was approved by the ethics board of the University Medical Center Groningen and followed the tenets of the Declaration of Helsinki.

Perimetry and Glaucoma

Perimetry was performed using the HFA 30-2 Swedish interactive threshold algorithm (SITA) fast strategy. An abnormal test result was defined as any one of the following: a glaucoma hemifield test "outside of normal limits," a pattern standard deviation (PSD) with $P < 0.05$, or three adjacent nonedge points with $P < 0.05$ in the pattern deviation probability plot, of which at least one point reached $P < 0.01$, with all points being on the same side of the horizontal meridian. Because the influence of reliability was part of the outcome of the study, we did not exclude test results based on reliability indices.

For glaucomatous visual field loss at baseline, two consecutive perimetry test results had to be abnormal in at least one eye. Defects had to be in the same hemifield, and at least one depressed test point of these defects had to have exactly the same location on both fields. Moreover, defects had to be compatible with glaucoma and without any other explanation. This was judged by one of three glaucoma specialists involved in the baseline of the study.[13] A visual field test prior to the two baseline tests was discarded to reduce the influence of learning. Thus, at least three tests had to be performed at baseline before glaucomatous visual field loss could be diagnosed.

For being a glaucoma patient at baseline, glaucomatous visual field loss had to be present. Neither the intraocular pressure (IOP) nor the aspect of the optic disk was a formal part of the glaucoma diagnosis. However, the mean (standard deviation [SD]) pretreatment IOP of the presumed glaucoma patients was 30.3 (9.5) mm Hg and 90% of them had an abnormal optic disk according to the GDx (number > 29).[13,16]

During the follow-up period, perimetry was performed at a frequency of one test per year. Clinicians were allowed to increase the frequency of testing, for example, in case of suspected progression. This was a subjective decision; no formal tools or rules were given (observational study design).

Data Analysis

For this study, one eye per patient was included. If both eyes met the above-described criteria, one eye was chosen randomly. Parametric and nonparametric descriptive statistics were performed to describe the study population.

To determine the influence of the factors technician experience, time of day, and season, and of the reliability indices, we first performed a linear regression analysis of MD against time for all included series of visual fields and calculated the MD residuals (i.e., the distances between the measured MD values and the corresponding regression line). Subsequently, we selected one random visual field of each patient and performed a multiple linear regression analysis with the MD residual of the included fields as dependent variable and technician experience, time of day, season, and phase of follow-up (see the following text) as independent variables. We performed these analyses both with and without adjustment for the percentages of FL and FP and FN answers. In this way, we were able (1) to determine the influence of these indices on the MD and (2) to explore if potential effects of the other factors were caused by a direct influence or by an indirect influence, through the reliability. The random selection of one visual field per patient and subsequent multiple linear regression analysis was repeated 50 times. After applying this resampling technique, we presented the results as the mean effect estimates averaged over the 50 resamplings, with corresponding 95% confidence intervals (CIs), for all studied independent variables. As a secondary analysis, the analyses were repeated with the pattern standard deviation (PSD) residuals as dependent variable. A value of $P \leq 0.05$ was considered statistically significant (see Discussion section).

Categorical variables were recoded into dummy variables. For technician experience, the technicians were stratified into three categories: inexperienced, moderately experienced, and highly experienced. Twenty-eight technicians were involved in the study; for the stratification according to experience we ranked these technicians according to the number of visual fields they had performed. Subsequently, we divided the visual fields into three equally sized groups. This resulted in a group of 20 inexperienced technicians who performed a median of 22 tests (range: 1 to 82 tests) during the entire follow-up period, a group of 5 moderately experienced technicians who performed a median of 120 tests (range: 83 to 147 tests), and a group of 3 highly experienced technicians who performed a median of 171 tests (range: 161 to 193 tests) as part of the study. For time of day, the tests were stratified into four categories: performed before 10:00 AM, between 10:00 AM and noon, between noon and 2:00 PM, and after 2:00 PM. For season, the tests were also stratified into four categories, of 3 months each, based on the annual variation of retinal sensitivity as found by Sweeney et al.[11] The summer was classified as June, July, and August; the autumn as September, October, and November; and so on. The reliability indices were treated as continuous variables measured relative to their mean value in the concerned patient (e.g., 10% FP answers was coded as +5% in a patient who had on average 5% FP answers and as −5% in a patient who had on average 15% FP answers). Finally, we added an additional variable, which was the phase of follow-up. This variable reflects if a visual field belongs to the first, middle, or last part of the follow-up. This variable aims to adjust for possible deviations from the presumed linear relationship between MD and time. These deviations might occur, for example, due to prolonged learning or disease acceleration and might confound some of the studied relationships (see Discussion section).

The MD analyses were performed for the entire study population, with and without adjustment for the reliability indices. They were also performed after stratification according to glaucoma severity. Here, two strata were used: a baseline MD of −6 dB or better (early glaucoma) versus below −6 dB (moderate/severe glaucoma). This cutoff value yielded two, roughly equally sized groups. The PSD analyses were performed only for early glaucoma because the PSD is only for early glaucoma monotonically related to disease stage.

RESULTS

In all, 160 eyes of 160 patients, 89 males and 71 females, were included. The mean (SD) age at baseline was 63.6 (11.2) years. The baseline MD was skewed negatively, with a median value of −7.8 dB (range: −29.4 to −0.0 dB; interquartile range: −15.4 to −2.9 dB). There was a median MD decline of −0.2 dB/year (range: −3.2 to +0.7 dB/year; interquartile range: −0.5 to 0.0
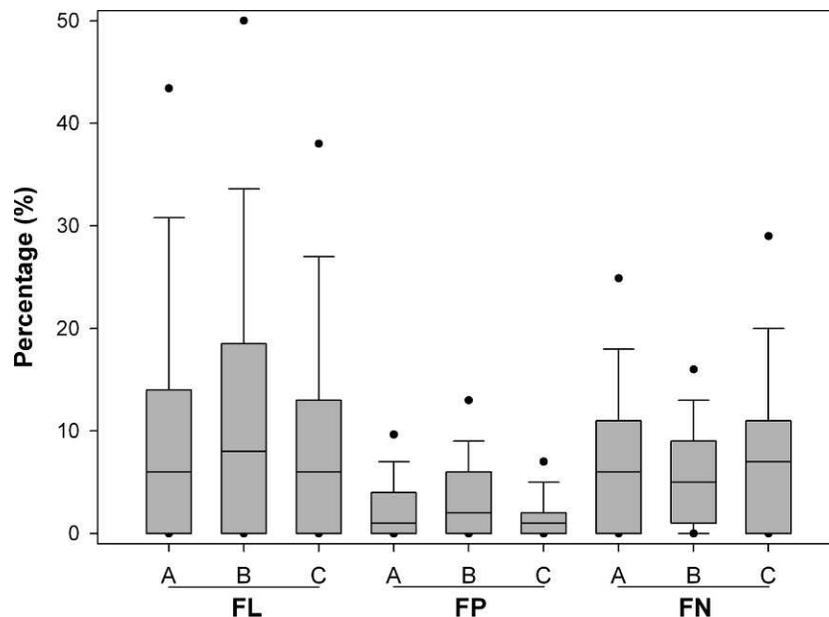
Figure 1.   Frequency distributions of the reliability indices FL, FP, and FN answer for all patients (**A**) and after stratification in early (**B**; mean deviation at baseline −6 dB or better) and moderate/severe (**C**) glaucoma. *Box plots* with *error bars* and *black dots* represent the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles.

dB/year; P90 −0.8 dB/year). The mean (SD) follow up was 9.4 (1.7) years, with a mean of 10.8 (2.2) visual field tests per patient. In total, 1735 tests were performed, of which 107 had to be excluded from further analysis because either no technician was registered, two technicians accompanied the test together as part of training, no fixation was monitored, or the FN answers were not available (which sometimes occurs in end-stage glaucoma; fields with FN answers not available had a mean [SD] MD of −27.0 [2.2] dB). Thus, 1628 visual field tests were included.

Figure 1 presents the frequency distributions of the FL and the FP and FN answers of all included 1628 visual field tests. The FL showed the highest percentages, followed by the FN and the FP answers. The percentages of FL and FP answers were significantly higher in early glaucoma compared with moderate/severe glaucoma (Mann–Whitney test, *P* < 0.001). The percentage of FN answers did not differ between early and moderate/severe glaucoma (*P* = 0.62). There was no significant correlation between the three reliability indices (Spearman's rank correlation coefficients < 0.30 for all three comparisons).

Table 1 presents the results of the multiple linear regression analysis for the entire study population, for all independent variables including the reliability indices. Table 2 shows the same analysis after the removal of the reliability indices. The effects of the various independent variables were essentially the same in both tables, indicating that these variables influence the MD directly rather than that they affect the reliability, with a subsequent effect on the MD. Figure 2 gives the results after stratification according to glaucoma stage.

Reliability Indices

As can be seen in Table 1, the percentage of FP answers had the greatest influence on the MD; a 1% increase in FP answers yielded an MD increase of approximately 0.1 dB (i.e., the MD of a visual field with 10% FP answers is overestimated by approximately 1 dB in comparison with a visual field with 0% FP answers). The influence of the percentages of FL and FN answers on the MD was also statistically significant, but the effect estimates were obviously lower than that of the

percentage of FP answers. Only in early glaucoma was a clinically relevant effect of the percentage of FN answers found. Here, the MD of a visual field with 10% FN answers is underestimated by approximately 0.5 dB in comparison with a visual field with 0% FN answers (Fig. 2).

Technician Experience

Technician experience appeared to be important, independently of glaucoma stage (Table 1; Fig. 2). Guidance by inexperienced technicians yielded MD values that were

Table 1.   Influence of Technician Experience, Time of Day, Season, Follow-up, and Test Reliability on the Mean Deviation (dB), in the Entire Study Population

	Beta	95% Confidence Interval	P Value
FL residuals	0.007	0.005 to 0.009	<0.001
FP residuals	0.086	0.078 to 0.095	<0.001
FN residuals	0.006	0.001 to 0.011	0.037
Technician:			
Highly experienced	Reference		
Medium experienced	−0.024	−0.108 to 0.060	0.571
Inexperienced	−0.176	−0.261 to −0.091	<0.001
Time of day:			
Before 10:00 AM	Reference		
10:00 to 11:59 AM	−0.033	−0.115 to 0.050	0.433
12:00 to 13:59 PM	−0.197	−0.286 to −0.109	<0.001
After 14:00 PM	−0.062	−0.152 to 0.028	0.174
Season:			
Spring	Reference		
Summer	−0.132	−0.225 to −0.038	0.006
Autumn	−0.189	−0.265 to −0.113	<0.001
Winter	−0.052	−0.146 to 0.041	0.266
Follow-up:			
Middle part	Reference		
First part	−0.240	−0.307 to −0.173	<0.001
Third part	−0.196	−0.275 to −0.117	<0.001

	Beta	95% Confidence Interval	*P* Value
Technician:			
Highly experienced	Reference		
Medium experienced	−0.063	−0.146 to 0.020	0.135
Inexperienced	−0.245	−0.333 to −0.157	<0.001
Time of day:			
Before 10:00 AM	Reference		
10:00 to 11:59 AM	−0.021	−0.100 to 0.058	0.589
12:00 to 13:59 PM	−0.158	−0.250 to −0.066	0.001
After 14:00 PM	−0.069	−0.155 to 0.017	0.112
Season:			
Spring	Reference		
Summer	−0.091	−0.182 to 0.001	0.053
Autumn	−0.152	−0.234 to −0.070	0.001
Winter	−0.072	−0.168 to 0.025	0.144
Follow-up:			
Middle part	Reference		
First part	−0.226	−0.295 to −0.157	<0.001
Third part	−0.103	−0.191 to −0.016	0.021

approximately 0.2 dB lower in comparison with their highly experienced colleagues ($P < 0.001$); for intermediately experienced technicians this difference was negligible.
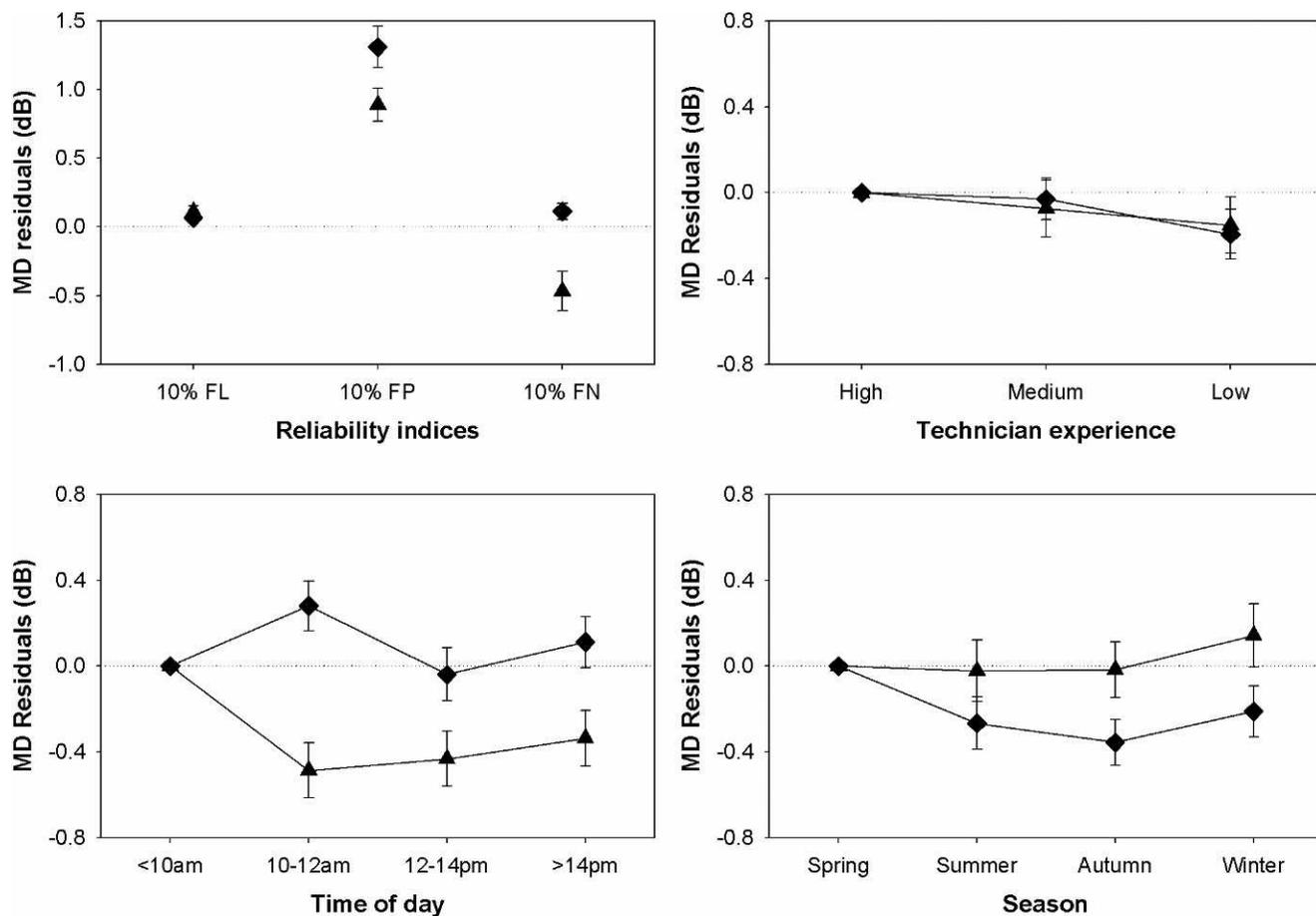
Time of Day

Time of day had a significant influence on the MD. Approximately 0.2 dB lower MD values were found directly after lunch ($P < 0.001$; Table 1). After stratification according to glaucoma stage, patients with early glaucoma performed significantly better in the early morning, with an approximately 0.4 dB better MD compared with the rest of the day. In moderate/severe glaucoma, the effect of time of day was less pronounced (Fig. 2).

Season

Interseasonal differences also appeared to play a significant role in visual field testing. Approximately 0.2 dB lower MD values were found in the summer/autumn compared with winter/spring ($P < 0.001$; Table 1). After stratification according to glaucoma stage, patients with early glaucoma appeared to have the highest sensitivity in the winter and patients with moderate/severe glaucoma in the spring (Fig. 2).

Follow-up

As can be seen in Table 1, the MD was significantly lower both in the beginning and at the end of the follow-up, compared

FIGURE 2. Influence of the test reliability (*top left*), technicians experience (*top right*), time of day (*bottom left*), and season (*bottom right*) on the MD after stratification according to glaucoma stage, being early (*triangles*; mean deviation at baseline −6 dB or better) versus moderate/severe glaucoma (*diamonds*). Results presented as average betas with 95% confidence intervals (dB).
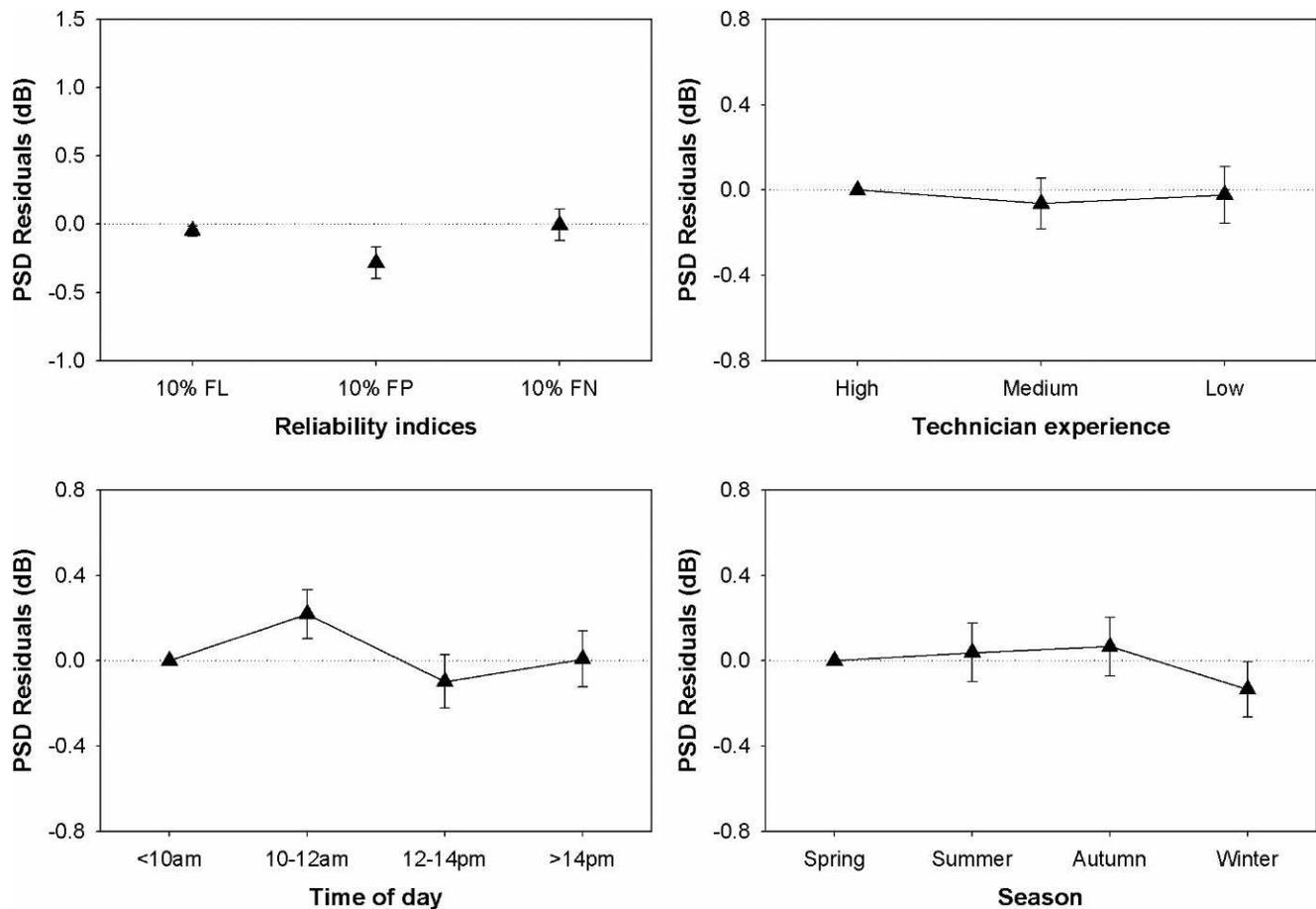
FIGURE 3. Influence of the test reliability (*top left*), technicians experience (*top right*), time of day (*bottom left*), and season (*bottom right*) on the PSD for patients with early glaucoma (mean deviation at baseline −6 dB or better). Results presented as average betas with 95% confidence intervals (dB).

with the middle part of the follow-up. This indicates a systematic deviation from the assumed linear decay.

Pattern Standard Deviation

Figure 3 shows the results of the multiple linear regression analysis with the PSD residuals as dependent variable, for patients with early glaucoma. Significant effects were found for the reliability indices (FP answers), time of day, and season. The effects were roughly opposite and generally smaller in comparison with the corresponding analysis with the MD residuals as dependent variable (Fig. 2, triangles).

Discussion

Technician experience, time of day, and season have a clinically relevant influence on the MD of SAP test results; together they may cause differences between tests of typically 0.5 dB. Of the three reliability indices, an excess of FP answers is the only serious threat to the test result; the MD is overestimated by 1 dB per 10% of FP answers.

Reliability Indices

In a clinical setting, Lee et al.[17] studied the influence of the reliability indices on the mean sensitivity variable (MS) of the 90° full-field projection perimeter (Octopus perimeter, Program G1 201; Interzeag, Schlieren, Switzerland). They used multiple linear regression analysis with the intertest difference

of the MS as the dependent variable and the difference of FP or FN catch trials as the independent variable (reliable visual field compared with unreliable visual field). They found an increase in MS of 1.5 dB for every 10% FP answers, whereas the MS decreased by 1.2 dB for every 10% FN answers. Their results are essentially in agreement with our findings, and this suggests that these findings are universal for SAP rather than specific to a single perimeter/strategy. Bengtsson[18] examined the associations between the reliability indices and the reproducibility of the MD with multiple linear regression analysis, using the HFA SITA standard strategy in a clinical setting. After disease stage, the percentage of FN answers had the highest, but nonsignificant, association. This result could be explained by the correlations between FN answers and disease stage and disease stage and reproducibility, which is in agreement with other studies and partially with our results.[18–20] In our study, the percentage of FN answers was of some clinical significance only in early glaucoma, whereas in moderate/severe glaucoma its influence was negligible (Fig. 2). Bengtsson did not find a significant influence of the percentage of FP answers. Two studies investigated the influence of FP answers by artificially adding random answers during full threshold[21] and SITA standard testing.[22] In healthy subjects, the addition of 33% FP answers resulted in an increase in the Humphrey STATPAC mean defect of 2.9 dB and an increase in the MD of 0.3 dB, respectively. In glaucoma patients, Newkirk et al.[22] found an increase in MD of 2.4 dB, which agrees with our findings. Our findings also indicate that FP answers cutoff point of 15%, as advocated for the SITA strategies,[23] may be not strict enough.

Technician Experience

The role of technicians cannot be underestimated. Proper instructions,[4] correct refraction,[3,5] and ensuring optimal conditions contribute to reliable test results, whereas supervision plays only a minor role.[24-26] As far as we know, no publications exist that examined the association between the degree of experience of technicians and the test result. Tables 1 and 2 and Figure 2 show that the test result was negatively influenced by inexperienced technicians, independently of glaucoma stage. Although the average influence of technician experience may be limited, it might be the case that the influence of individual technicians on the test results is substantially larger. Theoretically, this could be analyzed by putting all technicians individually as a dummy in the technician experience variable. However, there are not enough tests/observations per technician for such a detailed analysis. Thus, all we can conclude is that, on average, inexperienced technicians perform worse. In addition, inexperienced technicians seemed to have performed still a substantial number of tests, given that they also performed tests in patients not included in the study (see Methods section). In reality, those classified as inexperienced technicians are working in other parts of our department and visit the glaucoma service only incidentally, whereas those classified as experienced technicians have the glaucoma service as their default shop floor.

Time of Day

Circadian rhythms play a major role in the daily life of humans. These rhythms are endogenously generated by the circadian pacemaker, located in the hypothalamus. Adjustments to the circadian pacemaker are made by exposure to environmental light through intrinsically photosensitive retinal ganglion cells containing melanopsin.[27,28] Effects of time of day are difficult to interpret because they are influenced by interindividual differences in circadian period length, circadian phase, sleep duration, and the duration of prior wakefulness, vulnerability to sleep loss, age, and personality.[6] Several researchers examined circadian rhythms in visual thresholds.[7-9,12,29] Although the confidence intervals were wide, retinal sensitivity seemed to be the lowest in the early morning. We found the lowest sensitivity directly after lunch and, in a subgroup of patients with early glaucoma, the highest sensitivity in the early morning. Therefore, another explanation could be that, in perimetry, the influence of cognitive performance dominates that of retinal sensitivity. Cognitive performance seems to be best in the early morning at an age comparable to that of glaucoma patients, although this is influenced by many factors.[6,30] Lower performance in the afternoon is frequently observed and referred to as the "post lunch dip."[31] Especially in the elderly who suffer more often from a short and fragmented nocturnal sleep pattern, the severe post lunch dip is counteracted by daytime afternoon naps.[32] The lack of possibility to take the afternoon nap in patients if they have to visit the ophthalmology department may worsen performance during the afternoon.[33] In patients with moderate/severe glaucoma, the influence of time of day appeared to be less pronounced. This might be explained by a threshold effect, performance is always bad, or by glaucoma itself, which may disturb circadian rhythms by lesioning the nonimage-forming light-sensitive system (Lanzani MF, et al. *IOVS* 2011;52:ARVO E-Abstract 3471).[34-37]

Season

There is little information published on the influence of season on visual thresholds. Sweeney et al.[11] published in 1960 a study showing that, in healthy subjects, scotopic sensitivities were lowest in summer and gradually increased until spring, roughly following daily exposure to sunlight or prior light history.[10,38,39] Bassi and Powers[12] reported, in their study on circadian effects in healthy subjects, that subjects were slightly more sensitive during the winter. We found the highest sensitivity in winter and spring, in agreement with the findings of Sweeney et al.[11] (despite the fact that perimetry is not a scotopic task) and Bassi and Powers.[12] Very recently, a seasonal influence on the MD was reported in patients with ocular hypertension. Typically, 0.1 dB higher MD values were found in the winter compared with summer (Gardiner SK, et al. *IOVS* 2012;53:ARVO E-Abstract 1751). This is in agreement with our findings in patients with early glaucoma.

Pattern Standard Deviation

As a secondary analysis, we repeated our analyses with the PSD residuals instead of the MD residuals as dependent variable. These analyses were limited to patients with early glaucoma, because only for early glaucoma is the PSD monotonically related to disease stage. The significant effects on the PSD residuals we found (Fig. 3) were roughly opposite to and generally smaller than the corresponding effects on the MD residuals (Fig. 2, triangles). The opposite direction suggests, at least in early glaucoma, that the effects of the studied factors are larger in the diseased parts of the visual field than in the healthy parts. Obviously, this may change in more advanced disease, as indicated in Figure 2. Clinically, the smaller effects on the PSD residuals (Fig. 3) in comparison with the corresponding effects on the MD residuals (Fig. 2, triangles) tentatively indicate that the PSD is a more robust global index than the MD in early glaucoma.

Other Issues

To weigh the clinical relevance of our findings, the effects found in this study should be compared with the rate of change of the MD due to glaucoma and to the overall variability. As mentioned earlier in the Results section, the median MD decline in our study population was −0.2 dB/year. This indicates that the effects found in this study (together, typically 0.5 dB) are in the order of magnitude of the MD loss due to glaucoma after 2 to 3 years. Effects of typically 0.5 dB will generally not compromise the interpretation of a single test result. When series of test results are analyzed as part of progression detection, however, the effects may play a significant role. After all, the effects are nonnegligible when compared with the mean (SD) overall variability, expressed as the square root of the residual mean square of the MD (dB), which was 1.1 (0.7) dB in our study population.[40] In the case of event detection, additional variability may result in mixing-up stability with suspected progression or, after suspected progression has been observed, in mixing-up falsification with confirmation. In the case of trend analysis, additional variability will result in a longer period before a slope significantly different from zero can be detected,[41] and a less precise estimate of the actual slope can be made for a given follow-up.[15] Another clinically useful message is that only the percentage of false-positive answers has to be taken into account.

In our clinical setting, we adopted the SITA fast strategy because it was considered a time-saving improvement of the SITA standard strategy at the time we designed the study (1999). Later, it became clear that the strategies performed slightly differently. For example, more variability is found in the SITA fast strategy due to a higher error-related factor (ERF) at the end of the test related to a shorter data acquisition time.[42] Assuming that both SITA strategies reveal an unbiased estimate

of the MD (albeit the SITA fast strategy with a higher variability), factors that influence the MD by influencing the physiologic visual sensitivity should yield similar effects for both strategies. Here, it is interesting to note that Gardiner and colleagues found roughly similar seasonal effects (Gardiner SK, et al. *IOVS* 2012;53:ARVO E-Abstract 1751), using the full threshold strategy, as we did with the SITA fast strategy (discussed above; Season subsection). Factors that influence the MD through inattention or fatigue might have a more pronounced effect in longer strategies, such as SITA standard. The effects of the reliability indices seem to be universal for SAP rather than specific to a single perimeter/strategy (discussed above; Reliability Indices subsection).

The use of linear regression analysis is a common approach in glaucoma progression research.[41,43,44] However, if a systematic deviation from a linear deterioration would exist, false-positive associations might pop up in our analyses. This would occur for variables that are associated with the follow-up time point (e.g., if the glaucoma service opening hours would change from AM to PM during the study). To adjust for this possible confounding, we added a variable designated "follow-up." We found a lower sensitivity during the first and third tertiles of the follow-up, indicating a systematic deviation from a linear deterioration. This might be explained by prolonged learning[45,46] and/or the fact that glaucoma seems to have an accelerating character when using the MD as the outcome measure.[16] As a consequence, short series of visual fields might underestimate the rate of progression. We explored this possibility by comparing the first 3 years of follow-up to the entire follow-up in a subset of 104 patients with at least 9 years of follow-up (mean [SD] follow-up 10.4 [0.8] years). Table 3 shows the results. As can be seen in this table, the most obvious difference between the first 3 years and the entire follow-up is not a difference in the median rate of progression but rather a greater variability related to the shorter follow-up duration.[15,47] This suggests that both the phenomenon of "positive slopes" and the phenomenon of "rapid progressors" seem to be at least partially related to too short series of visual fields.

Another factor that might influence the MD is the development of cataract (or after cataract) or a cataract extraction (or capsulotomy) during follow-up. A change in MD slope due to a gradual development of cataract will not influence the residuals. Also, a deviation from a linear decay due to a faster development of cataract or a cataract extraction will not influence our findings. First, we also adjusted for these deviations with the variable designated "follow-up." Second, it is reasonable to assume that media opacity changes are not associated with the characteristics of the sampling (technician choice, time of day, and season). Thus, although the development of cataract or a cataract extraction during follow-up may increase MD variability, it does not influence our findings.

Tests performed by a single patient cannot be considered independent observations. We addressed this issue by the use of a resampling method, by randomly taking one test result of each patient followed by the multiple linear regression analysis, and repeating this 50 times. In multiple linear regression analysis, multiple hypothesis testing is already taken into account. It is not unequivocal if correction for multiple hypothesis testing would be needed with the resampling method applied in this study. However, the effects reported as being significant in this study all had $P < 0.001$, indicating that they would remain significant even with very conservative corrections for multiple hypothesis testing. We also addressed the dependence issue by taking the reliability indices relative to their mean value in the concerned patient (see Methods section). Finally, effects can be found only if the independent variables display sufficient variability (e.g., if all patients would

TABLE 3. Rate of Progression of the Mean Deviation (dB/year) as a Function of Follow-up Duration (Based on a Subset of 104 Patients with at Least 9 Years of Follow-up)

	First 3 Years of Follow-up	Entire Follow-up Period
P10	0.71	0.16
P25	0.24	−0.04
Median	−0.17	−0.16
P75	−0.68	−0.40
P90	−1.29	−0.63

visit the glaucoma service on a fixed time of day and year, no effects of time of day and year will be found). In a scientifically ideal situation, appointments would be allocated at random. Because this is not the case in an observational study, the observed effects might be smaller than the true effects. In our department, time of day is especially prone to limited variability. Of all patients, 65% had at least one test in all four time-of-day strata, 31% had no tests in one stratum, 4% had no tests in two strata, and 0% had all tests in a single stratum. For season, these percentages were 63, 34, 3, and 1%, respectively. These percentages suggest that there was sufficient within-subject variability for time of day and season to investigate their influences on perimetry test results.

Besides confounding, interactions between independent variables could also play a role. An interaction could be expected, for example, for time of day and season (there is little environmental light present in early morning during winter compared with summer). Interactions can be explored by stratification. If we stratified our analyses according to season instead of to stage (the latter was performed in Fig. 2), we found essentially similar effects of time of day for all four seasons, indicating no obvious interaction between time of day and season. The power (number of included subjects) prohibited more thorough interaction analyses.

In conclusion, in a clinical setting where individual patients are tested at the same time of day while technicians pay attention to patient instruction and reassurance, monitor fixation, and address apparent inattention, clinicians should pay attention only to the percentage of FP answers. Visual fields with more than 5% FP answers should be used for quantitative analyses only with caution. Compared with the MD, the PSD appeared to be less influenced by the studied factors and might thus be the preferred global index in early glaucoma.

References

1. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Invest Ophthalmol Vis Sci*. 2000;41:2201–2204.

2. 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;110:1895–1902.

3. Mutlukan E. The effect of refractive blur on the detection sensitivity to light offsets in the central visual field. *Acta Ophthalmol*. 1994;72:189–194.

4. Kutzko KE, Brito CF, Wall M. Effect of instructions on conventional automated perimetry. *Invest Ophthalmol Vis Sci*. 2000;41:2006–2013.

5. Anderson RS, McDowell DR, Ennis FA. Effect of localized defocus on detection thresholds for different sized targets in the fovea and periphery. *Acta Ophthalmol Scand*. 2001;79:60–63.

6. Blatter K, Cajochen C. Circadian rhythms in cognitive performance: methodological constraints, protocols, theoretical underpinnings. *Physiol Behav*. 2007;90:196–208.

7. Roenneberg T, Lotze M, Von Steinbuchel N. Diurnal variation in human visual sensitivity determined by incremental thresholds. *Clin Vis Sci*. 1992;7:83–91.

8. Tassi P, Pellerin N, Moessinger M, Eschenlauer R, Muzet A. Variation of visual detection over the 24-hour period in humans. *Chronobiol Int*. 2000;17:795–805.

9. Tuunainen A, Kripke DF, Cress AC, Youngstedt SD. Retinal circadian rhythms in humans. *Chronobiol Int*. 2001;18:957–971.

10. Hebert M, Martin SK, Lee C, Eastman CI. The effects of prior light history on the suppression of melatonin by light in humans. *J Pineal Res*. 2002;33:198–203.

11. Sweeney EJ, Kinney JA, Ryan A. Seasonal changes in scotopic sensitivity. *J Opt Soc Am*. 1960;50:237–240.

12. Bassi CJ, Powers MK. Daily fluctuations in the detectability of dim lights by humans. *Physiol Behav*. 1986;38:871–877.

13. Heeg GP, Blanksma LJ, Hardus PL, Jansonius NM. The Groningen Longitudinal Glaucoma Study. I. Baseline sensitivity and specificity of the frequency doubling perimeter and the GDx nerve fibre analyser. *Acta Ophthalmol Scand*. 2005;83: 46–52.

14. Wesselink C, Heeg GP, Jansonius NM. Glaucoma monitoring in a clinical setting: glaucoma progression analysis vs nonparametric progression analysis in the Groningen Longitudinal Glaucoma Study. *Arch Ophthalmol*. 2009;127:270–274.

15. Jansonius N. On the accuracy of measuring rates of visual field change in glaucoma. *Br J Ophthalmol*. 2010;94:1404–1405.

16. Wesselink C, Marcus MW, Jansonius NM. Risk factors for visual field progression in the Groningen Longitudinal Glaucoma Study: a comparison of different statistical approaches. *J Glaucoma*. In press.

17. Lee M, Zulauf M, Caprioli J. The influence of patient reliability on visual field outcome. *Am J Ophthalmol*. 1994;117:756–761.

18. Bengtsson B. Reliability of computerized perimetric threshold tests as assessed by reliability indices and threshold reproducibility in patients with suspect and manifest glaucoma. *Acta Ophthalmol Scand*. 2000;78:519–522.

19. Katz J, Sommer A, Witt K. Reliability of visual field results over repeated testing. *Ophthalmology*. 1991;98:70–75.

20. Birt CM, Shin DH, Samudrala V, Hughes BA, Kim C, Lee D. Analysis of reliability indices from Humphrey visual field tests in an urban glaucoma population. *Ophthalmology*. 1997;104: 1126–1130.

21. Cascairo MA, Stewart WC, Sutherland SE. Influence of missed catch trials on the visual field in normal subjects. *Graefes Arch Clin Exp Ophthalmol*. 1991;229:437–441.

22. Newkirk MR, Gardiner SK, Demirel S, Johnson CA. Assessment of false positives with the Humphrey Field Analyzer II perimeter with the SITA algorithm. *Invest Ophthalmol Vis Sci*. 2006;47:4632–4637.

23. Heijl A, Patella VM, eds. *Essential Perimetry. The Field Analyzer Primer*. Dublin, CA: Carl Zeiss Meditec, Inc.; 2002.

24. Johnson LN, Aminlari A, Sassani JW. Effect of intermittent versus continuous patient monitoring on reliability indices during automated perimetry. *Ophthalmology*. 1993;100:76–84.

25. Van Coevorden RE, Mills RP, Chen YY, Barnebey HS. Continuous visual field test supervision may not always be necessary. *Ophthalmology*. 1999;106:178–181.

26. Kramer BC, Musch DC, Niziol LM, Weizer JS. Reliability of simultaneous visual field testing. *Ophthalmology*. 2012;119: 304–307.

27. Hanifin JP, Brainard GC. Photoreception for circadian, neuroendocrine, and neurobehavioral regulation. *J Physiol Anthropol*. 2007;26:87–94.

28. Markwell EL, Feigl B, Zele AJ. Intrinsically photosensitive melanopsin retinal ganglion cell contributions to the pupillary light reflex and circadian rhythm. *Clin Exp Optom*. 2010;93: 137–149.

29. O'Keefe LP, Baker HD. Diurnal changes in human psychophysical luminance sensitivity. *Physiol Behav*. 1987;41:193–200.

30. Yoon C, May CP, Hasher L. Aging, circadian arousal patterns, and cognition. In: Schwarz N, Park D, Knauper B, Sudman S, eds. *Aging, Cognition, and Self Reports*. Washington, DC: Psychological Press; 1999:117–143.

31. Monk TH. The post-lunch dip in performance. *Clin Sports Med*. 2005;24:e15–e23.

32. Buysse DJ, Browman KE, Monk TH, Reynolds CF III, Fasiczka AL, Kupfer DJ. Napping and 24-hour sleep/wake patterns in healthy elderly and young adults. *J Am Geriatr Soc*. 1992;40: 779–786.

33. Takahashi M, Fukuda H, Arito H. Brief naps during post-lunch rest: effects on alertness, performance, and autonomic balance. *Eur J Appl Physiol Occup Physiol*. 1998;78:93–98.

34. Jean-Louis G, Zizi F, Lazzaro DR, Wolintz AH. Circadian rhythm dysfunction in glaucoma: a hypothesis. *J Circadian Rhythms*. 2008;6:Art 1.

35. Drouyer E, Dkhissi-Benyahya O, Chiquet C, et al. Glaucoma alters the circadian timing system. *PLoS One*. 2008;3:e3931.

36. Feigl B, Mattes D, Thomas R, Zele AJ. Intrinsically photosensitive (melanopsin) retinal ganglion cell function in glaucoma. *Invest Ophthalmol Vis Sci*. 2011;52:4362–4367.

37. Agorastos A, Huber CG. The role of melatonin in glaucoma: implications concerning pathophysiological relevance and therapeutic potential. *J Pineal Res*. 2011;50:1–7.

38. Smith KA, Schoen MW, Czeisler CA. Adaptation of human pineal melatonin suppression by recent photic history. *J Clin Endocrinol Metab*. 2004;89:3610–3614.

39. Jasser SA, Hanifin JP, Rollag MD, Brainard GC. Dim light adaptation attenuates acute melatonin suppression in humans. *J Biol Rhythms*. 2006;21:394–404.

40. Junoy Montolio FG, Wesselink C, Jansonius NM. Persistence, spatial distribution and implications for progression detection of blind parts of the visual field in glaucoma: a clinical cohort study. *PloS ONE*. 2012;7:e41211.

41. Chauhan BC, Garway-Heath DF, Goñi FJ, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol*. 2008;92:569–573.

42. Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from full threshold, SITA standard, and SITA fast strategies. *Invest Ophthalmol Vis Sci*. 2002;43:2654–2659.

43. McNaught A, Hitchings RA, Crabb D, Fitzke F. Modelling series of visual fields to detect progression in normal-tension glaucoma. *Graefes Arch Clin Exp Ophthalmol*. 1995;233:750–755.

44. Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression. *Am J Ophthalmol*. 2008;145: 343–353.

45. Wild J, Dengler-Harles M, Searle A, O'Neill E, Crews S. The influence of the learning effect on automated perimetry in patients with suspected glaucoma. *Acta Ophthalmol*. 1989;67: 537–545.

46. Gardiner SK, Demirel S, Johnson CA. Is there evidence for continued learning over multiple years in perimetry? *Optom Vis Sci*. 2008;85:1043–1048.

47. Nouri-Mahdavi K, Zarei R, Caprioli J. Influence of visual field testing frequency on detection of glaucoma progression with trend analyses. *Arch Ophthalmol*. 2011;129:1521–1527.