tvst

Article

Factors Predicting a Greater Likelihood of Poor Visual Field Reliability in Glaucoma Patients and Suspects

Inas F. Aboobakar[1], Jiangxia Wang[1], Balwantray C. Chauhan[2], Michael V. Boland[1], David S. Friedman[1], Pradeep Y. Ramulu[1], and Jithin Yohannan[1]

[1] Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[2] Department of Ophthalmology and Visual Sciences, Dalhousie University, Halifax, Nova Scotia, Canada

Correspondence: Jithin Yohannan, Wilmer Eye Institute, Johns Hopkins Hospital, 600 N. Wolfe St., Baltimore, MD 21287, USA. e-mail: jithin@jhmi.edu

Received: May 9, 2019
Accepted: November 15, 2019
Published: January 30, 2020

Keywords: visual field; visual field reliability; false positives; false negatives; fixation losses

Citation: Aboobakar IF, Wang J, Chauhan BC, Boland MV, Friedman DS, Ramulu PY, Yohannan J. Factors predicting a greater likelihood of poor visual field reliability in glaucoma patients and suspects. Trans Vis Sci Tech. 2020;9(1):4, https://doi.org/10.1167/tvst.9.1.4

Purpose: Identify factors predicting worse or better than expected visual field (VF) performance.

Methods: A total of 10,262 VFs from 1538 eyes of 909 subjects with manifest or suspected glaucoma were analyzed. Linear mixed-effects models predicted mean deviation (MD) at each timepoint. Differences between observed and predicted MD (ΔMD) were calculated and logistic regression identified factors predicting lower than expected (ΔMD <-1 dB) or higher than expected (ΔMD >1 dB) sensitivity.

Results: Both higher and lower than expected sensitivity were more likely in VFs with severe compared with mild damage (relative risk [RR] >1.3, $P < 0.05$). Higher than expected sensitivity was more likely in VFs with moderate damage (RR $= 2.57$, $P < 0.001$). False-positive (FP) errors increased the likelihood of higher than expected sensitivity at all disease stages (RR >2.1 per 10% increase, $P < 0.001$), whereas false-negative (FN) errors increased the likelihood of lower than expected sensitivity in mild and moderate disease (RR >1.19 per 10% increase, $P < 0.05$). Fixation loss errors slightly increased the likelihood of higher than expected VF sensitivity in moderate and severe disease (RR >1.1 per 10% increase, $P < 0.01$). Longer test duration increased likelihood of lower than expected sensitivity at all disease stages (RR >1.36 per minute increase, $P < 0.001$). Lower than expected sensitivity was more likely in late afternoon tests (RR $= 1.27$, $P < 0.01$). A total of 26.6% of VFs had higher or lower than expected sensitivity in the absence of FPs, FNs, or fixation losses.

Conclusions: FPs, test duration, and FNs are the primary measures predicting if a VF is likely to be reliable, although tests with normal reliability measures may still be unreliable. Our results help clinicians judge VF reliability and highlight the need to integrate reliability measures with other clinical data when making treatment decisions.

Translational Relevance: This likelihood model derived from a large dataset helps clinicians identify VFs that may either falsely suggest disease progression or mask true worsening, thereby improving the utility of VFs in clinical practice.

Introduction

Visual fields (VFs) are critical for glaucoma diagnosis, gauging disease severity and monitoring for progression.[1–3] The utility of VFs in clinical decision making is dependent on test reliability. This is particularly crucial when deciding whether to make treatment decisions (i.e., advancement of therapy) based on VF data.

Previous studies have aimed to identify criteria for assessing whether a VF is reliable. Early researchers took a qualitative approach to VF reliability, classifying VFs as either reliable or unreliable.[1,4] This binary assessment was based on cutoffs for the percent abnormal catch trials of false positives (FPs), false negatives (FNs), or fixation losses (FLs) for a particular VF. However, these binary classifications preclude a more nuanced assessment of whether a VF should be trusted. We recently developed a quantitative model for the

degree to which a VF was likely to deviate from its true value using a large dataset of VFs from glaucoma patients and suspects.[5] In that study, FPs, FNs, and test duration (TD) all significantly affected reliability; FLs, on the other hand, had little effect on reliability.[5] Based on those data, disease severity-specific standards for quantifying VF reliability were proposed.

Although our previous work helps the clinician understand the degree to which a VF deviates from its true value, in day-to-day clinical practice it is often more important to know how likely a VF is to deviate significantly from its true value. In other words, with what degree of FPs, FNs, FLs, or TD is a VF mean deviation (MD) likely to be 1 dB or more away from its expected value? In addition, it is important to know in which direction a VF deviates from expected (i.e., better than expected or worse than expected). A VF that is worse than expected has overly poor sensitivity that could falsely suggest disease progression; this is important to identify because acting on these results could subject a patient to unnecessary risks of treatment. Conversely, a better-than-expected VF has overly good sensitivity that can mask true disease progression.

In this study, we build on our prior work and create a predictive model that provides information on the likelihood of lower- or higher-than-expected sensitivity.

Methods

The study protocol was approved prospectively by the Johns Hopkins institutional review board and adhered to the tenets of the Declaration of Helsinki. A waiver of consent was obtained to review VF data and obtain patient information via chart review. The study was Health Insurance Portability and Accountability Act compliant.

Study Participants

Patients included in this study were 18 years of age or older and were evaluated at the Wilmer Eye Institute Glaucoma Center of Excellence between 2002 and 2012, as previously described.[5] All study participants had a glaucoma-related diagnosis (glaucoma suspect or any other form of glaucoma). All eyes that were analyzed had five or more VFs that were obtained with the Humphrey Field Analyzer (HFA II, Carl Zeiss Medical Technologies Inc., Dublin, CA) using the Swedish Interactive Threshold Algorithm (SITA) standard test protocol and the 24-2 pattern. Patients could have either one or both eyes included in the analyses. Only VFs with a MD greater (better) than -20 dB were included in the analyses.

Table 1. Demographic and Visual Field Characteristics of Study Patients

Total No. of Subjects	909
Total no. eyes	1,538
Mean no. of fields per eye (SD)	6.7 (1.8)
Total no. of fields	10,262
Age at first VF, mean (SD)	64.9 (11.8)
Gender	
Male (%)	497 (54.7%)
Female (%)	412 (45.3%)
Race	
White (%)	594 (65.3%)
Black (%)	232 (25.5%)
Asian (%)	83 (9.1%)
Baseline disease severity (no. of eyes)	
Mild (MD >-6 dB) (%)	1,110 (72.2%)
Moderate ($-12 <$MD ≤-6) (%)	262 (17.0%)
Severe ($-20 <$ MD ≤-12) (%)	166 (10.8%)

SD, standard deviation.

VF and Clinical Data Collection

VF data were retrieved for eyes meeting the inclusion criteria listed previously. MD was used as the measure of disease severity: MD >-6 was defined as mild disease, $-12 <$ MD ≤-6 as moderate disease, and $-20 <$ MD ≤-12 as severe disease. VF metrics potentially affecting the reliability of measured MD were examined, including the TD and the percentage of FL, FP, and FN. The time and date of each VF test was obtained and categorized as early morning (7-10 AM), late morning (10 AM to noon), early afternoon (noon to 2 PM), or late afternoon (2-5 PM). The quarter of the year was also recorded (1 = winter, 2 = spring, 3 = summer, 4 = fall). A chart review was performed to determine patient age, sex, race, and clinical characteristics for all study subjects (Table 1 and Supplementary Table S1).

Modeling of VF Reliability

VF reliability was computed as the difference between observed and predicted MD ($MD_{observed} - MD_{predicted}$) and referred to as ΔMD. The process for deriving ΔMD is shown in Figure 1. First, the predicted MDs were calculated for eligible VF tests using a linear mixed-effects regression model. The dependent variable in this model was the MD for each eligible VF test in the database; the independent

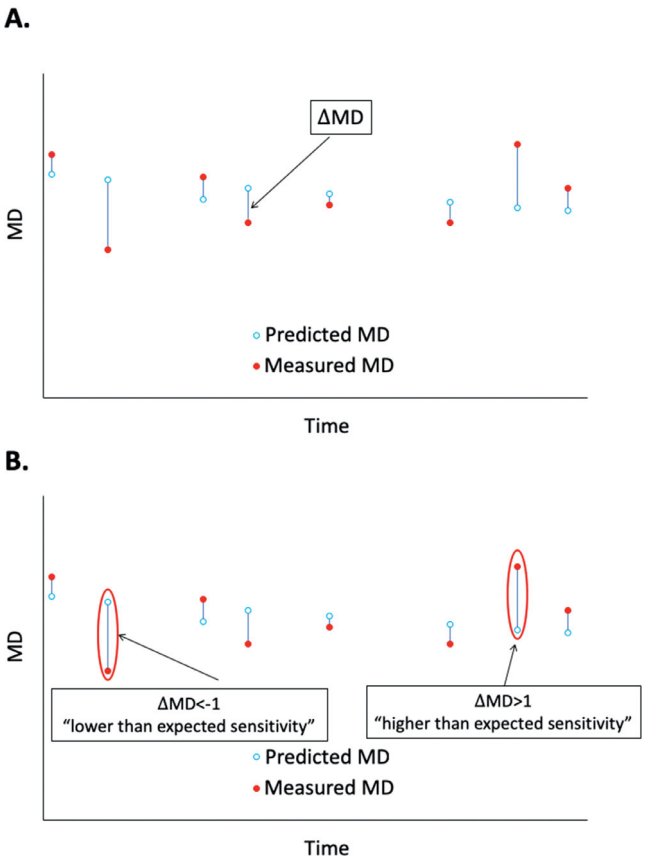

Figure 1. Depiction of the difference in measured versus predicted MD and identification of visual fields with higher than expected and lower than expected sensitivity. (A) Schematic shows how ΔMD was calculated. The predicted MD at each time point was calculated using a mixed-effects linear regression model using the set of input variables listed in Supplementary Table S1. The ΔMD was obtained by subtracting the predicted MD from the observed MD at that point ($\Delta MD = MD_{observed} - MD_{predicted}$). (B) Visual fields with lower than expected sensitivity were those in which the observed MD was worse than predicted ($\Delta MD < -1$), whereas VFs with higher than expected sensitivity were those in which the observed MD was better than predicted ($\Delta MD > 1$). (Figure adapted from Yohannan et al.[5])

variables are summarized in Supplementary Table S1. Because the baseline disease condition categories generated from the first VF MD and the eye-specific average VF MD were used as covariates, first VF MDs were not included in the sample used in the regression model. A linear mixed-effects regression model approach was used to account for clustering of different VF tests within the same eye within the same patient. The model used random intercepts, random slopes, and an unstructured variance-covariance matrix. Next, ΔMD was calculated as a continuous, directional measure of reliability for each VF test included in the study by subtracting the predicted MD obtained from the linear mixed effects model from the actual observed MD for that VF test ($\Delta MD = MD_{observed} - MD_{predicted}$). VFs with an observed MD value within 1 dB of the predicted MD ($|\Delta MD| \leq 1$ dB) were defined as having an "expected" sensitivity. VFs with an observed MD at least 1 dB worse than predicted ($\Delta MD < -1$ dB) had lower than expected sensitivity (worse-than-expected VF). VFs with an observed MD at least 1 dB greater than predicted ($\Delta MD > 1$ dB) had higher than expected sensitivity (better-than-expected VF).

Modeling Higher or Lower Than Expected Sensitivity

Predictors of lower ($\Delta MD < -1$ dB) and higher ($\Delta MD > 1$ dB) than expected sensitivity were identified with multinomial logistic mixed-effects regression models with a random intercept for patients and a random intercept for eyes to account for the correlations at the patient and eye level. In these models, the dependent variable was the category of ΔMD (-1 dB $\leq \Delta MD \leq 1$ dB, $\Delta MD < -1$ dB, or $\Delta MD > 1$ dB); the reference group for the dependent variable was the $|\Delta MD| \leq 1$ group (i.e., the group with observed MD close to that predicted). The independent variables included age, day of the week, quarter of the year, time of day, and the interaction terms between the baseline severity of VF loss and FLs, FPs, FNs, and TD to account for the fact that the effects of FL, FP, FN, and TD on ΔMD vary by the severity of VF loss. Relative risk (RR) was calculated for the effect of each factor on ΔMD reliability assuming all other factors are held constant. For FP, FN, and FL, the RRs were calculated per 10% increase in abnormal catch trials. For TD, the RRs were calculated per 1-minute increase in TD. All statistical analyses were performed in STATA statistical software: release 14 (StataCorp LP, College Station, TX).

Results

A total of 10,262 VFs from 1538 eyes of 909 subjects with manifest or suspect glaucoma were examined. Subjects performed a mean of 6.7 (standard deviation, 1.8) VFs per eye. Demographic data are presented in Table 1. There were 6842 VFs with an expected sensitivity ($|\Delta MD| \leq 1$ dB), 1775 VFs with lower than expected sensitivity, and 1645 VFs with higher than expected sensitivity (Table 2).

Factors predicting lower or higher than expected VF sensitivity

There were significant differences among the three groups across all measured VF metrics in a bivariate

Table 2. Characteristics of Visual Fields Grouped by Normal, Lower than Expected, or Higher than Expected Sensitivity

	$\lvert\Delta MD\rvert \leq 1$ Expected Sensitivity N = 6842	$\Delta MD < -1$ Lower than Expected Sensitivity N = 1775	$\Delta MD > 1$ Higher than Expected Sensitivity N = 1645	*P* Value
Baseline disease severity				<0.001
Mild (MD > −6 dB) (%)	5409 (79.1%)	1051 (59.2%)	1002 (60.9%)	
Moderate (−12 < MD ≤ −6) (%)	905 (13.2%)	431 (24.3%)	418 (25.4%)	
Severe (−20 < MD ≤ −12) (%)	528 (7.7%)	293 (16.5%)	225 (13.7%)	
Fixation losses				<0.001
0∓10%	3643 (53.3%)	914 (51.6%)	687 (41.9%)	
10∓20%	1318 (19.3%)	379 (21.4%)	330 (20.1%)	
>20%	1872 (27.4%)	479 (27.0%)	624 (38.0%)	
False positives				<0.001
0∓10%	6311 (92.2%)	1638 (92.3%)	1355 (82.4%)	
10∓20%	434 (6.3%)	111 (6.3%)	151 (9.2%)	
>20%	97 (1.4%)	26 (1.5%)	139 (8.4%)	
False negatives				<0.001
0∓10%	6201 (90.6%)	1347 (75.9%)	1279 (77.8%)	
10-20%	564 (8.2%)	326 (18.4%)	284 (17.3%)	
>20%	77 (1.1%)	102 (5.7%)	82 (5.0%)	
Test duration, min				<0.001
<6	4325 (63.2%)	510 (28.7%)	707 (43.0%)	
6-8	2216 (32.4%)	979 (55.2%)	719 (43.7%)	
>8	301 (4.4%)	286 (16.1%)	219 (13.3%)	
Time of day of VF				0.001
7∓10 AM	1861 (27.2%)	409 (23.0%)	438 (26.6%)	
10 AM∓12 PM	1802 (26.3%)	463 (26.1%)	439 (26.7%)	
12∓2 PM	1549 (22.6%)	407 (22.9%)	371 (22.6%)	
2∓5 PM	1630 (23.8%)	496 (27.9%)	397 (24.1%)	
Quarter of the year				0.032
First (winter)	1520 (22.2%)	381 (21.5)	409 (24.9)	
Second (spring)	1800 (26.3)	460 (25.9)	449 (27.3)	
Third (summer)	1647 (24.1)	464 (26.1)	409 (24.9)	
Fourth (fall)	1875 (27.4)	470 (26.5)	378 (23.0)	

Note: The *P* values are derived from mixed effects models for the VF reliability measures.

analysis (Table 2). In the multivariate analysis, compared with VFs with mild damage (MD > −6), those with severe (−20 < MD ≤ −12) damage were more likely to have lower than expected sensitivity (RR = 1.33, $P < 0.05$) (Table 3). Moderate (−12 < MD ≤ −6) and severe disease also increased the likelihood of a higher than expected sensitivity (RR = 2.57 and 2.44, respectively; $P < 0.001$).

FPs decreased the likelihood of a lower than expected sensitivity at all stages of disease severity (RR = 0.47, 0.52, and 0.49 per 10% FPs for mild, moderate, and severe disease; $P < 0.001$) (Table 3 and Fig. 2A). On the other hand, FPs increased the likelihood of a higher than expected sensitivity at all disease severities (RR = 2.14, 2.93, and 2.99 per 10% FPs for mild, moderate, and severe disease; $P < 0.001$) (Table 3 and Fig. 2B).

FNs increased the likelihood of a lower than expected sensitivity in mild (RR = 1.5 per 10% FN, $P < 0.001$) and moderate (RR = 1.19 per 10% FN, $P = 0.02$) disease, but not in severe disease (RR = 1.10, $P = 0.2$) (Table 3 and Fig. 2C). FNs did not

Table 3. Patient and Test Factors Influencing the Likelihood of Lower Than Expected and Higher Than Expected Visual Field Sensitivity

		ΔMD <−1 (Lower Than Expected) vs. |ΔMD| ≤1 (Expected)		ΔMD >1 (Higher Than Expected) vs. |ΔMD| ≤1 (Expected)	
Reliability Measure	Subgroups	RR (95% CI)	*P*	RR (95% CI)	*P*
Disease severity	Mild (MD >−6)	Ref	-	Ref	-
	Moderate (−12 < MD ≤−6)	1.01 (0.83–1.23)	0.910	2.57 (2.07–3.20)	**<0.001**
	Severe (−20 < MD ≤−12)	1.33 (1.02–1.73)	**0.036**	2.44 (1.77–3.38)	**<0.001**
False positives (10% increase)	None/mild	0.47 (0.40–0.56)	**<0.001**	2.14 (1.85–2.48)	**<0.001**
	Moderate	0.52 (0.41–0.66)	**<0.001**	2.93 (2.28–3.77)	**<0.001**
	Severe	0.49 (0.34–0.70)	**<0.001**	2.99 (2.06–4.36)	**<0.001**
False negatives (10% increase)	None/mild	1.50 (1.28–1.75)	**<0.001**	0.98 (0.82–1.18)	0.841
	Moderate	1.19 (1.02–1.38)	**0.023**	1.02 (0.86–1.22)	0.802
	Severe	1.10 (0.95–1.27)	0.222	1.18 (0.98–1.42)	0.074
Fixation losses (10% increase)	None/mild	0.91 (0.88–0.95)	**<0.001**	1.01 (0.97–1.05)	0.622
	Moderate	0.96 (0.90–1.02)	0.213	1.11 (1.03–1.19)	**0.005**
	Severe	0.96 (0.88–1.04)	0.287	1.15 (1.05–1.26)	**0.003**
Test duration (1-min. increase)	None/mild	2.12 (1.95–2.30)	**<0.001**	0.90 (0.81–1.00)	**0.048**
	Moderate	1.58 (1.41–1.78)	**<0.001**	0.72 (0.62–0.82)	**<0.001**
	Severe	1.36 (1.18–1.56)	**<0.001**	0.76 (0.63–0.91)	**0.003**
Time of day of VF	7am-10am	Ref	-	Ref	-
	10am-12pm	1.09 (0.93–1.27)	0.275	0.97 (0.82–1.15)	0.758
	12pm-2pm	1.07 (0.91–1.62)	0.399	0.92 (0.77–1.10)	0.371
	2pm-5pm	1.27 (1.09–1.74)	**0.003**	0.94 (0.79–1.13)	0.513
Quarter of the year	First (winter)	Ref	-	Ref	-
	Second (spring)	1.08 (0.92–1.26)	0.362	0.95 (0.81–1.13)	0.583
	Third (summer)	1.19 (1.01–1.39)	**0.033**	0.89 (0.75–1.06)	0.186
	Fourth (fall)	1.08 (0.92–1.26)	0.352	0.74 (0.62–0.88)	**0.001**

Bold face type in table indicates statistically significant results ($p < 0.05$).

significantly affect the likelihood of a higher than expected sensitivity ($P > 0.05$ for all stages of disease) (Table 3 and Fig. 2D).

FLs decreased the likelihood of lower than expected sensitivity in mild disease (RR = 0.91 per 10% FL, $P < 0.001$), and increased the likelihood of higher than expected sensitivity in moderate (RR = 1.11 per 10% FL, $P < 0.01$) and severe (RR = 1.15 per 10% FL, $P < 0.01$) disease, though the effect sizes were small (Table 3 and Figs. 2E, 2F). Of note, 687 (6.7%) of all VFs had zero FPs, FNs, or FLs; among these VFs, 193 (26.6%) had either lower or higher than expected sensitivity (data not shown in tables).

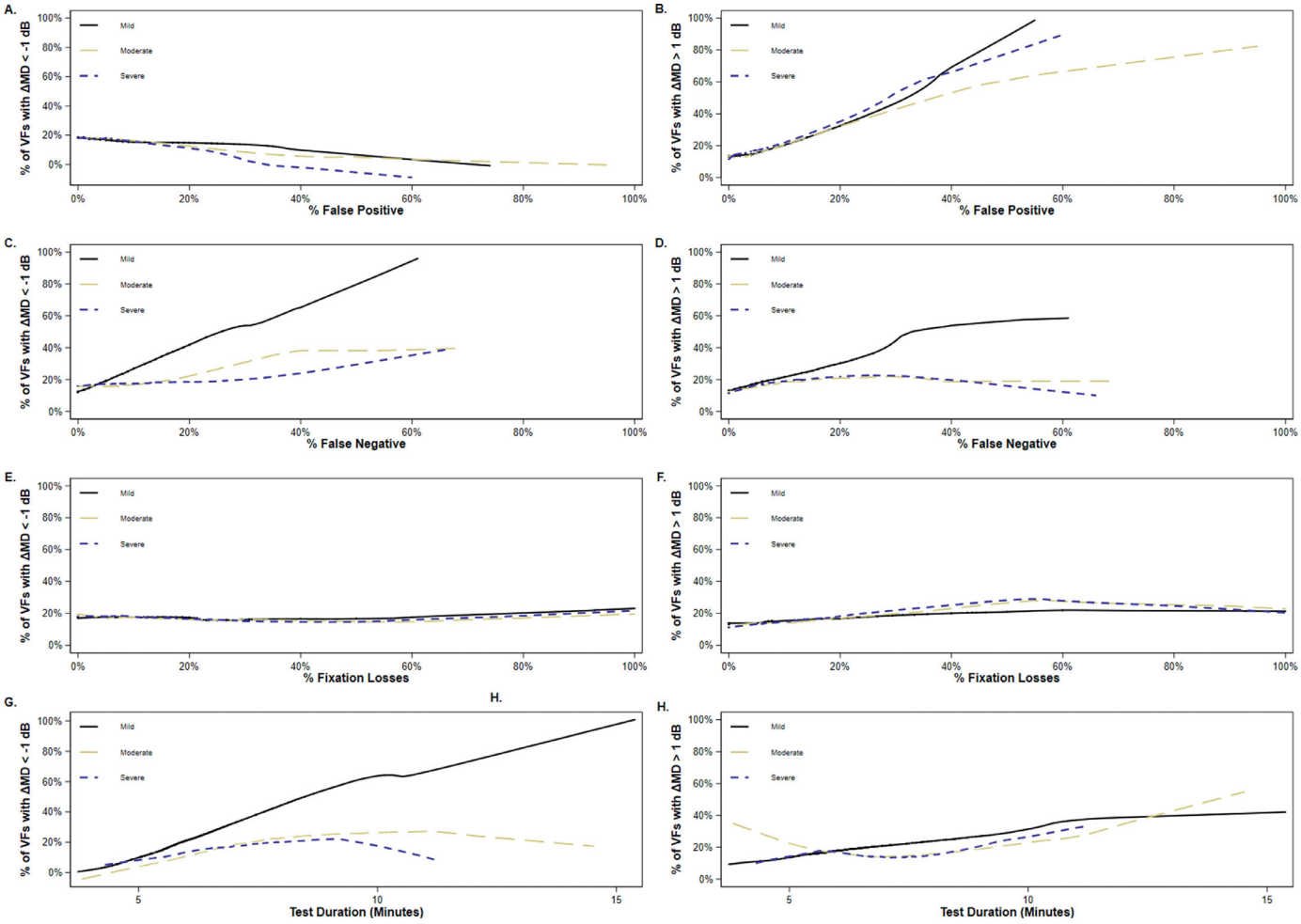

Figure 2. Influence of false positives, false negatives, fixation losses, and test duration on the likelihood of lower or higher than expected visual field sensitivity across the range of observed values. (A) False positives decrease the likelihood of a lower-than-expected sensitivity at all stages of disease severity. (B) False positives also increase the likelihood of a higher-than-expected sensitivity at all stages of disease severity. (C) False negatives increase the likelihood of a lower-than-expected sensitivity in mild and moderate disease. There was no significant effect in severe disease ($P > 0.05$). (D) False negatives have no significant effect on the likelihood of higher-than-expected sensitivity ($P > 0.05$). (E) Fixation losses slightly decrease the likelihood of lower-than-expected sensitivity in early disease. There was no significant effect in moderate and severe disease. (F) Fixation losses have a mild effect on the likelihood of higher-than-expected sensitivity in moderate and severe disease. (G) Longer test duration increases the likelihood of a lower-than-expected sensitivity at all disease stages. (H) Longer test duration decreases the likelihood of a higher-than-expected sensitivity at all stages of disease severity.

Longer TD increased the likelihood of a lower than expected sensitivity at all stages of disease severity (RR = 2.12, 1.58, and 1.36 per 1-minute increase in TD for mild, moderate, and severe disease; $P < 0.001$) (Table 3 and Fig. 2G). Also, VFs with short TD almost never had lower than expected sensitivity (Fig. 2G). Increased TD also decreased the likelihood of higher than expected sensitivity at all disease severities (RR = 0.9, 0.72, and 0.76 per 1-minute increase in TD for mild, moderate, and severe disease; $P < 0.05$) (Table 3 and Fig. 2H). The trend for an association between increased TD and lower than expected sensitivity was also confirmed within each of the disease severity bins. A 1-minute increase in TD results in a 1-dB decrease in ΔMD in the mild glaucoma group (95% confidence interval [CI], −1.06, −0.93), a 0.76-dB decrease in ΔMD in the moderate glaucoma group (95% CI, −0.87, −0.65), and a 0.33-dB decrease in ΔMD in the severe glaucoma group (95% CI, −0.47, −0.19).

VF tests taken in the late afternoon (2-5 PM) were more likely to have a lower than expected sensitivity compared with early morning tests (7-10 AM) (RR = 1.27, $P < 0.01$) (Table 3). Tests taken in late morning (10 AM-12 PM) or early afternoon (12-2 PM) did not differ significantly from early morning tests (7-10 AM) ($P > 0.05$). No time of day was associated with a higher

likelihood of a higher than expected sensitivity ($P > 0.05$). VFs taken in the summer tended to have a higher likelihood of lower than expected sensitivity compared with VFs taken in the winter (RR: 1.19, $P = 0.03$), and VFs taken in the fall had a lower likelihood of higher than expected sensitivity compared with those taken in the winter (RR: 0.74, $P = 0.01$) (Table 3).

Discussion

In our models designed to gauge the likelihood of VF reliability, FNs and longer TD increased the likelihood of a worse-than-expected MD, whereas FPs increased the likelihood of a better-than-expected MD. Late afternoon tests also increased the likelihood of a worse-than-expected MD compared with early morning tests. Finally, severe VF damage increased variability of MD in both directions, and thus were noted to have a greater likelihood of both a worse-than-expected and better-than-expected MD. Even in VFs without any FPs, FNs, and FLs, 26.6% were either better or worse than expected. Our findings are important to bear in mind when interpreting VFs because acting on a series of worse-than-expected VFs can lead to unnecessary treatment, whereas better-than-expected VFs could mask true progression.

The reliability measures identified in this study hold important practical implications for day-to-day clinical practice. First, they highlight the importance of taking into consideration FPs, TD, FNs, and FLs when interpreting VFs. In our previous work, we showed that FPs have the strongest impact on the ΔMD for a given VF, with greater effects noted in more severe disease.[5] The results of this study show that FPs have the greatest effect on the likelihood of poor VF reliability, with a particularly pronounced effect in more severe disease. Similar to this study, TD had the second largest effect on the ΔMD in our prior study,[5] followed by FNs and FLs, as judged by the magnitude of RR. The direction of effect of each of these variables on the ΔMD in our prior study (i.e., increase or decrease in ΔMD) also correlates with the direction of effect on the likelihood of VF reliability (i.e., better than expected or worse than expected). FLs, while demonstrating a statistically significant effect on reliability in some models, do not appear to have a clinically meaningful effect in this population of experienced VF takers given the small odds ratios, consistent with our prior work.[5] The relative unimportance of FLs may be due to the fact that FLs can occur frequently when the blind spot is mismapped, with small head tilts and other changes in patient positioning during VF testing, each of which would not be expected to also produce significant errors in overall sensitivity.[5–7] Of note, although mismapping of the blind spot would not be expected to significantly affect the MD of the entire VF, it may cause local impacts in the area of mismapping (affecting the reliability of sensitivities of specific points, which might be important in pointwise analyses).

We decided to consider poor reliability from a better-than-expected and worse-than-expected MD as separate outcomes because the clinical implications of these two outcomes are distinct. Specifically, VFs with a better-than-expected sensitivity (i.e., from FPs or, to a much lesser extent, FLs) could mask true disease progression. These tests may therefore need to be repeated, particularly if there is a concern for progression based on other clinical parameters (i.e., intraocular pressure, visual symptoms, Optical Coherence Tomography findings, trajectory of the other eye). On the other hand, VFs likely to have a lower than expected sensitivity (i.e., from a long TD or a high number of FNs), particularly in the context of mild or moderate disease, may be repeated if it suggests actionable progression, but would not need to be repeated were it to indicate stable VF damage. Of note, FNs do not affect the likelihood of a worse-than-expected VF in severe disease, likely because of the variable response of diseased locations in advanced disease[8] and that VF MD likely reaches a floor in advanced disease and therefore reduces the marginal impact of FNs in such cases. Therefore, high FNs in severe disease should not generally serve as grounds for discounting a VF suggestive of progression.

Interestingly, our results also demonstrate that 26.6% of VFs that have no FPs, FNs, or FLs still demonstrate a better or worse than expected MD. Thus, it is important to not consider changes in the VF MD in isolation. Rather, these changes should be considered in the context of other findings, including where the worse points in the VF are located, intraocular pressure, patient symptoms, and optic nerve structure.[9,10]

Another notable finding in our study was the effect of time of day on the likelihood of VF reliability because tests taken in late afternoon were 30% more likely to be worse than expected compared with early morning VF tests. These results corroborate the findings of a prior study that demonstrated that post-lunchtime VFs were 0.2 dB worse on average compared with early morning tests.[7] Given these findings, it may be worthwhile to consider performing a morning VF test before making treatment decisions in a patient who shows worsening on an afternoon VF compared with a previously obtained morning VF. Additionally, scheduling a given patient to take VFs at one consistent time of day may reduce variability in testing results.

Our results also demonstrate a mild seasonal variation in the likelihood of a better- or worse-than-expected VF outcome. VFs done in the winter tend to have a higher likelihood of higher-than-expected sensitivity than VFs taken in the fall. Additionally, they have a lower likelihood of lower-than-expected sensitivity than VFs taken in the summer. Our prior work[5] and the work of other authors[7] support this idea of mild seasonal variance in VF performance, with VF performance being highest in the winter months. Prior experimental work has revealed that visual sensitivity tends to be higher in the winter in healthy subjects, which may help explain this mild association.[11,12]

One limitation of this work is that the predicted MD that was used to calculate ΔMD was based on modeling and there is no way to know the "true" level of expected VF loss at any point in time. The generalizability of our study to different perimeters and testing algorithms other than the SITA standard strategy and 24-2 pattern may also be limited; prior studies have shown, for instance, that there are overall significantly fewer FNs and FPs in tests incorporating SITA compared with a full threshold algorithm.[13,14] Our analyses also did not include some factors that have been shown to predict the reliability of VFs, including technician comments, patient experience with VF testing, or gaze tracking.[15,16] Last, our study did not use serial repeated measures of VFs to assess reliability; Bengtsson's approach of repeated measures (having patients repeat VF within 1 week) to measure reliability is certainly a viable alternative methodology, although it is impractical to apply to this methodology to a large sample of clinical patients.[17]

In summary, we have created a predictive model for the likelihood of poor VF reliability using a large database of VFs from glaucoma patients and suspects. Utilization of these results in day-to-day practice will allow a clinician to best use VF results for glaucoma management.

Acknowledgments

Supported by grants from the National Institutes of Health R01 EY022976, Research to Prevent Blindness, and Doris Duke Foundation.

Disclosure: **I.F. Aboobakar**, None; **J. Wang**, None; **B.C. Chauhan**, None; **M.V. Boland**, None; **D.S. Friedman**, None; **P.Y. Ramulu**, None; **J. Yohannan**, None

References

1. Nelson-quigg JM, Twelker JD, Johnson CA. Response properties of normal observers and patients during automated perimetry. *Arch Ophthalmol*. 1989;107:1612–1615.
2. Advanced Glaucoma Intervention Study. Visual field test scoring and reliability. *Ophthalmology*. 1994;101:1445–1455.
3. Chauhan BC, Malik R, Shuba LM, Rafuse PE, Nicolela MT, Artes PH. Rates of glaucomatous visual field change in a large clinical population. *Invest Ophthalmol Vis Sci*. 2014;55:4135–4143.
4. Bickler-bluth M, Trick GL, Kolker AE, Cooper DG. Assessing the utility of reliability indices for automated visual fields. Testing ocular hypertensives. *Ophthalmology*. 1989;96:616–619.
5. Yohannan J, Wang J, Brown J, et al. Evidence-based criteria for assessment of visual field reliability. *Ophthalmology*. 2017;124:1612–1620.
6. Tan NYQ, Tham YC, Koh V, et al. The effect of testing reliability on visual field sensitivity in normal eyes: the Singapore Chinese Eye Study. *Ophthalmology*. 2018;125:15–21.
7. Junoy Montolio FG, Wesselink C, Gordijn M, Jansonius NM. Factors that influence standard automated perimetry test results in glaucoma: test reliability, technician experience, time of day, and season. *Invest Ophthalmol Vis Sci*. 2012;53:7010–7017.
8. Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. *Ophthalmology*. 2014;121:1359–1369.
9. Leske MC, Heijl A, Hussein M, et al. Factors for glaucoma progression and the effect of treatment: the early manifest glaucoma trial. *Arch Ophthalmol*. 2003;121:48–56.
10. Wollstein G, Schuman JS, Price LL, et al. Optical coherence tomography longitudinal evaluation of retinal nerve fiber layer thickness in glaucoma. *Arch Ophthalmol*. 2005;123:464–470.
11. Sweeney EJ, Kinney JAS, Ryan A. Seasonal changes in scotopic sensitivity. *J Opt Soc Am*. 1960;50:237.
12. Bassi CJ, Powers MK. Daily fluctuations in the detectability of dim lights by humans. *Physiol Behav*. 1986;38:871–877.
13. Johnson CA, Sherman K, Doyle C, Wall M. A comparison of false-negative responses for full

threshold and SITA standard perimetry in glaucoma patients and normal observers. *J Glaucoma*. 2014;23:288–292.
14. Wall M, Doyle CK, Brito CF, Woodward KR, Johnson CA. A comparison of catch trial methods used in standard automated perimetry in glaucoma patients. *J Glaucoma*. 2008;17:626–630.
15. Johnson LN, Aminlari A, Sassani JW. Effect of intermittent versus continuous patient monitoring on reliability indices during automated perimetry. *Ophthalmology*. 1993;100:76–84.
16. Ishiyama Y, Murata H, Asaoka R. The usefulness of gaze tracking as an index of visual field reliability in glaucoma patients. *Invest Ophthalmol Vis Sci*. 2015;56:6233–6236.
17. 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.
18. The AGIS Investigators. The Advanced Glaucoma Intervention Study (AGIS): 7. The relationship between control of intraocular pressure and visual field deterioration. *Am J Ophthalmol*. 2000;130:429–440.
19. Heijl A, Leske MC, Bengtsson B, Hyman L, Bengtsson B, Hussein M. Early Manifest Glaucoma Trial Group. Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial. *Arch Ophthalmol*. 2002;120:1268–1279.
20. Lichter PR, Musch DC, Gillespie BW, et al. CIGTS Study Group. Interim clinical outcomes in the Collaborative Initial Glaucoma Treatment Study comparing initial treatment randomized to medications or surgery. *Ophthalmology*. 2001;108:1943–1953.