AMERICAN ACADEMY™
OF OPHTHALMOLOGY

Evidence-based Criteria for Assessment of Visual Field Reliability

Jithin Yohannan, MD, MPH,[1,2] *Jiangxia Wang, MS, MA,*[1,2] *Jamie Brown, MD,*[1,2] *Balwantray C. Chauhan, PhD,*[3] *Michael V. Boland, MD, PhD,*[1,2] *David S. Friedman, MD, MPH, PhD,*[1,2] *Pradeep Y. Ramulu, MD, MHS, PhD*[1,2]

Purpose: Assess the impact of false-positives (FP), false-negatives (FN), fixation losses (FL), and test duration (TD) on visual field (VF) reliability at different stages of glaucoma severity.

Design: Retrospective.

Participants: A total of 10 262 VFs from 1538 eyes of 909 subjects with suspect or manifest glaucoma and ≥5 VF examinations.

Methods: Predicted mean deviation (MD) was calculated with multilevel modeling of longitudinal data. Differences between predicted and observed MD (ΔMD) were calculated as a reliability measure. The impact of FP, FN, FL, and TD on ΔMD was assessed using multilevel modeling.

Main Outcome Measures: ΔMD associated with a 10% increment in FP, FN, and FL, or a 1-minute increase in TD.

Results: FL had little impact on ΔMD (<0.2 decibels [dB] per 10% abnormal catch trials), and no level of FL produced ≥1 dB of ΔMD at any disease stage. FP yielded greater than expected MD, with a 10% increment in abnormal catch trials associated with a ΔMD = 0.42, 0.73, and 0.66 dB in mild (MD >−6 dB), moderate (−6 ≤MD <−12 dB), and severe (−12 ≤MD ≤−20 dB) disease, respectively, up to 20% abnormal catch trials, and a ΔMD = 1.57, 2.06, and 3.53 dB beyond 20% abnormal catch trials. FNs generally produced observed MDs below expected MDs. FN were minimally impactful up to 20% abnormal catch trials (ΔMD per 10% increment >−0.14 dB at all levels of severity). Beyond 20% abnormal catch trials, each 10% increment in abnormal catch trials was associated with a ΔMD = −1.27, −0.53, and −0.51 dB in mild, moderate, and severe disease, respectively. |ΔMD| ≥1 dB occurred with 22% FP and 26% FN in early, 14% FP and 34% FN in moderate, and 16% FP and 51% FN in severe disease. A 1-minute increment in TD produced ΔMDs between −0.35 and −0.40 dB.

Conclusions: FL have little impact on reliability in patients with established glaucoma. FP, and to a lesser extent FNs and TD, significantly affect reliability. The impact of FP and FN varies with disease severity and over the range of abnormal catch trials. On the basis of our findings, we present evidence-based, severity-specific standards for classifying VF reliability for clinical or research applications. *Ophthalmology 2017;■:1–9 © 2017 by the American Academy of Ophthalmology*

Supplementary material is available at www.aaojournal.org

A number of seminal glaucoma studies, including the Ocular Hypertension Treatment Study,[1] Collaborative Initial Glaucoma Treatment Study,[2] Early Manifest Glaucoma Trial,[3] and Advanced Glaucoma Intervention Study,[4] have used automated visual field (VF) testing to assess the presence of glaucoma, gauge disease severity, and track progression. As such, automated VF testing remains the primary tool that glaucoma practitioners use to assess glaucoma-related visual damage, monitor progression,[5–7] and determine the impact of glaucoma on patient functionality.[8–11]

When performing VF testing, it is important to know whether the VF test was completed properly by the patient to determine how the results should be used to guide care. Classically, this question has been answered by using reliability measures based on the percentage of abnormal catch trials in metrics such as fixation losses (FL), false-positives (FP), and false-negatives (FN) to classify a VF as unreliable (untrustworthy and needing repetition) or reliable (usable for clinical decision making). The Humphrey Field Analyzer (HFA) software, for example, uses a cutoff of 33% FP or FN, and 20% for FL to define a field as unreliable.[12] Initially, such cutoffs were validated by testing normal subjects, ocular hypertensive patients, and patients with early glaucoma to determine the percentage of subjects who met or exceeded the cutoff levels. This early work showed that less than 0.5% of patients and normal subjects exceeded the 33% cutoff for FP and FN, but a large number (19%–35%) exceeded the cutoff of 20% or more FL,[13,14] and the suggestion was made to increase the cutoff for FL to 33%.[14] On the basis of these early data, VFs with FP, FN, or FL greater than 33% were not included in the Ocular Hypertension Treatment Study,[1] and these same cutoff values gained acceptance as a means

© 2017 by the American Academy of Ophthalmology
Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.ophtha.2017.04.035
ISSN 0161-6420/17

to judge whether a VF examination was unreliable. However, there are several limitations to this approach. First, the initial development of such cutoffs was based solely on how many patients exceeded the cutoff rather than an assessment of whether the fields exceeding these cutoffs values were unlikely to represent the true degree of VF loss. Second, cutoffs create a binary categorization of VF results (reliable or unreliable) that precludes consideration of *how unreliable* (i.e., how disparate from the true, unknown level of VF damage at the time of the examination) a VF is likely to be based on test parameters (i.e., FP, FN, FL).

To more meaningfully assess the impact of FP, FN, and FL on VF reliability, quantitative measures capturing the degree of VF reliability are needed to allow clinicians to make decisions based on the degree of error likely to be present in their patients' test results. Bengtsson[15] studied quantitative reliability in patients who performed VF tests twice in 1 week and found that the standard reliability indices were not significantly associated with threshold reproducibility, although severity of field loss was. Junoy Montolio et al[16] used a different method to measure quantitative reliability, first predicting what a particular VF mean deviation (MD) should be by using modeling and then calculating the difference between this predicted MD and the actual MD of the VF test to define a measure of reliability known as ΔMD. They then calculated the effect of FP, FN, and FL on ΔMD and found that FP had the largest impact on ΔMD with a 10% increment in FP associated with a 1.5 decibels (dB) higher ΔMD. FN and FL, when compared with FP, had less dramatic effects on ΔMD. Although this study moved toward a quantitative assessment of VF reliability, it had several limitations, including (1) lack of a complete investigation of the relative impact of FP, FN and FL at different stages of disease severity; (2) equal weighting of each additional 10% FP, FN, or FL (thus suggesting that going between 0% and 10% FP and between 30% and 40% FP has an equivalent impact on ΔMD); and (3) a limited sample of 160 patients.

In this study, we build on the work of Junoy Montolio[16] and use a large VF database to determine the quantitative impact of FP, FN, FL, and test duration (TD) on VF reliability as defined by ΔMD, the difference between observed MD values and those predicted by our regression models. In addition, we attempt to assess how disease severity and the range of abnormal catch trials where additional FP, FN, and FL occur affect the relationship between these indices and VF reliability. Finally, we use our data to propose evidence-based criteria to quantitatively judge VF reliability in clinical practice and research settings.

Methods

The study protocol was approved 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 to obtain information via chart review.

Study Participants

Patients aged 18 years or older who were evaluated at the Wilmer Eye Institute Glaucoma Center of Excellence between 2002 and 2012 were eligible to be included in the analysis if they had a glaucoma-related diagnosis (glaucoma suspect or any other form of glaucoma). Eyes for which 5 or more VFs were obtained with the HFA II (Carl Zeiss Medical Technologies Inc., Dublin, CA) and the 24-2 Swedish Interactive Threshold Algorithm (SITA) protocol were analyzed. Patients could have 1 or both eyes included in analyses. Because the current study was designed to evaluate the impact of abnormal catch trials and other VF metrics on VF reliability, no eyes or VFs were excluded because of poor reliability. Only VFs with an MD >-20 were included in the analysis.

The VF data were retrieved for eyes meeting the inclusion criteria. Mean deviation (MD) was extracted as the measure of disease severity. Visual field metrics potentially affecting measured MD were extracted, including the test duration and the percentage of abnormal FL, FP, and FN catch trials. Finally, the date and time of each VF test were obtained. Time of day was categorized as early morning (7−10 AM), late morning (10 AM to noon), early afternoon (12 noon to 2 PM), or late afternoon (2−5 PM). The date of VF testing was used to determine the day of the week and the season of the year (spring, summer, fall, or winter).

Patient age was determined directly from VF output. A chart review was performed to determine patient sex, race, and the additional variables presented in Table 1.

Modeling of Reliability

Reliability of MD was computed as the difference between observed and predicted MD ($MD_{observed} - MD_{predicted}$, referred to as ΔMD), and was derived in a 3-step process as summarized in Figure 1. First, predicted MDs were calculated for eligible VF tests in the database with linear mixed-effects regression models. The dependent variable in this model was the MD for each eligible VF test in the database, whereas the independent variables included time and the features described in Table 1. Because the baseline disease condition categories generated from the first VF MD and the eye-specific average VF MD were used as covariates, first VFs were not excluded from the sample used in the regression model. A linear mixed-effects regression model approach was used to account for clustering between eyes within the same patient and VFs performed on the same eye. The model used random intercepts, random slopes, and an unstructured variance-covariance matrix. Second, 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 mixed effects model from the actual observed MD for that VF test ($MD_{observed} - MD_{predicted}$). Third, predictors of reliability (ΔMD) were identified with a multilevel linear mixed-effects model using random intercepts but not random slopes, because ΔMD was not expected to vary over time. In this final multivariate model, the dependent variable was ΔMD (representing reliability) and predictors that were used to explain ΔMD included FL, FP, FN, TD, time of day, day of week, and season. Interaction terms between the severity of VF loss and FLs, FPs, FNs, and TD were used to account for the fact that the effects of FL, FP, FN, and TD on ΔMD vary by the severity of VF loss. The estimated regression coefficients represent the effect of each factor on ΔMD assuming all other factors are held constant.

Derived regression coefficients were used to define the TD or percentage of abnormal FL, FP, and FN catch trials required to produce various degrees of unreliability at different stages of disease severity. Acceptable levels of ΔMD were defined as (1) an absolute level (i.e., >1 dB or <-1 dB) or (2) a level defined

Download English Version:

https://daneshyari.com/en/article/5705049

Download Persian Version:

https://daneshyari.com/article/5705049

Daneshyari.com