AMERICAN ACADEMY™ OF OPHTHALMOLOGY

The Effect of Testing Reliability on Visual Field Sensitivity in Normal Eyes

The Singapore Chinese Eye Study

Nicholas Y.Q. Tan, MA, BM BCh,[1] Yih-Chung Tham, PhD,[1] Victor Koh, MMed (Oph),[1,2] Duc Quang Nguyen, BEng,[1] Carol Y. Cheung, PhD,[3] Tin Aung, MBBS, PhD,[1,2,4] Tien Yin Wong, MD, PhD,[1,2,4] Ching-Yu Cheng, MD, PhD[1,2,4]

Purpose: To quantitatively determine how the reliability indices in standard automated perimetry (SAP) affect the global indices of visual field (VF) results in nonglaucomatous eyes.

Design: Observational, cross-sectional study.

Participants: A total of 830 adults aged 40 to 80 years, without visual impairment, glaucoma, significant cataract, and major eye diseases, were selected from the population-based Singapore Chinese Eye Study (SCES).

Methods: Study participants underwent a comprehensive and standardized ocular examination and VF assessment using a Humphrey Field Analyzer II (Carl Zeiss Meditec, Inc., Dublin, CA). The effects of the test reliability, as indicated by the false-negative (FN), false-positive (FP), and fixation loss (FL) rates, on global indices, as indicated by the mean deviation (MD) and pattern standard deviation (PSD), were analyzed with multivariable regression models.

Main Outcome Measures: The MD and PSD.

Results: A total of 1828 VF results from 1235 normal eyes of 830 study subjects were included in the analyses. The multivariable regression analyses adjusted for age, gender, best-corrected visual acuity, and test duration showed that at lower frequencies of false answers (<15%), FNs decreased the MD (β [change in decibels {dB} per 5% increment in false answers] = −0.71 dB; $P < 0.001$), whereas FPs increased the MD (β = 0.65 dB; $P < 0.001$). At higher frequencies (≥15%), the false answers influenced the MD to a greater extent, where the β for the associations with FN and FP rates was −1.15 and 1.26 dB, respectively (both $P < 0.001$). We also found that when FN rate was <15%, higher FN rate increased the PSD (β = 0.51 dB; $P < 0.001$), and the effect was slightly larger when FN rate was ≥15% (β = 0.71 dB; $P < 0.001$). The effect of FPs on PSD was observed only when FP rate was <15% (β = −0.22 dB; $P < 0.001$). The FL had no associations with the MD, and had minimal effects on the PSD.

Conclusions: We quantified the effect of unreliable responses on the MD and PSD in SAP. Our study may allow clinicians to estimate how VF results are affected by varying degrees of unreliability, instead of relying on cutoff values for reliability indices. *Ophthalmology 2018;125:15-21 © 2017 by the American Academy of Ophthalmology*

Glaucoma is a major cause of vision loss, and standard automated perimetry (SAP) using the Humphrey Field Analyzer 24-2 Swedish Interactive Threshold Algorithm (SITA) standard (Carl Zeiss Meditec, Inc., Dublin, CA) is one of the most widely used perimetric tests to assess visual field (VF) in glaucoma.[1–3] However, SAP is still a subjective assessment and is heavily influenced by the reliability of patient performance.[4] Traditional perimetric reliability indices, namely, the false negative (FN), false positive (FP), and fixation loss (FL), are used to judge the quality or reliability of VF test results.

In the original full threshold test procedure, the cutoff values for the reliability indices were entirely arbitrarily set at 33% for FNs and FPs, and 20% for FL. Subsequently, further reports[5,6] suggested that the cutoff limit for FP rates may be tightened further to reach a compromise between limiting the extent to which FP responses may affect the mean deviation (MD) and pattern standard deviation (PSD), and avoiding too strict an FP reliability criterion that might exclude too many "unreliable" VF results. It was also suggested that the FN responses may be discounted in the consideration of VF reliability because FNs may represent true VF loss rather than unreliable testing.[7] Therefore, in the current SITA test, the manufacturer guidelines suggest a reliability limit of 20% for FLs, 15% for FP responses, and no limit for FN responses.[8]

However, setting a cutoff for reliability indices is still inherently problematic because labeling VFs as "reliable" or

http://dx.doi.org/10.1016/j.ophtha.2017.08.002
ISSN 0161-6420/17

"unreliable" represents a false dichotomy because reliability—and its indices—falls along a continuum with a multitude of different degrees of dependability, rather than just two. Therefore, taking the results of "reliable" tests at face value and completely disregarding "unreliable" tests is an oversimplistic approach: on one hand, reliable VFs may still be misleading; on the other hand, unreliable VFs may contain enough useful information to establish or exclude disease, or to detect progression.

Thus, the aim of this study was to quantify in normal eyes how each of the 3 reliability indices (FN, FP, and FL) affect the global indices of the VF test, namely, the MD and PSD. This would allow the clinician to more accurately estimate to what degree unreliable responses may affect the VF globally.

Methods

Study Population

The study subjects were recruited as part of the Singapore Chinese Eye Study (SCES), a population-based, cross-sectional study of eye diseases in Chinese adults, aged 40 to 80 years, residing in the southwestern part of Singapore between February 2009 and December 2011. The methodology of the SCES has been reported in detail elsewhere.[9] Written informed consent was obtained from each participant. The study adhered to the tenets of the Declaration of Helsinki, and ethics committee approval was obtained from the SingHealth Centralised Institutional Review Board. In total, 3353 subjects (72.8% response rate) participated in the SCES between February 2009 and December 2011.

Ocular Examination

All subjects underwent a standardized and comprehensive interview and ophthalmic examination, including measurement of logarithmic minimal-angle resolution best-corrected visual acuity (BCVA), refraction, axial length measurement, intraocular pressure (IOP) measurement, and fundus examination. The IOP was measured with a Goldmann applanation tonometer (Haag-Streit, Bern, Switzerland) before pupil dilation. Fundus examination included the evaluation of the optic disc with a 78-diopter (D) lens at ×16 magnification with a measuring graticule during dilated ophthalmoscopy for the calculation of the vertical cup-to-disc ratio (VCDR). The refraction of each eye was measured using an autorefractor (Canon RK 5 Auto Ref-Keratometer; Canon Inc., Ltd., Tochigiken, Japan). Axial length was measured with noncontact partial coherence laser interferometry (IOLMaster version 3.01; Carl Zeiss Meditec AG, Jena, Germany), and the mean of 5 measurements was used in the analysis. Lens opacity was assessed from lens photographs using the Wisconsin Cataract Grading System.[10] In addition, participants also underwent VF assessment, which is further described below.

To include only normal eyes in this study, we excluded eyes that presented with the following conditions: (1) BCVA >0.1 for participants aged <50 years, or BCVA >0.2 for participants aged ≥50 years; (2) refractive error exceeding ±5 D sphere or 2.5 D cylinder; (3) axial length >26.5 mm; (4) presence of any corneal opacity or grade 2 or 3 pterygia[11]; (5) presence of a significant cataract, defined as any nuclear cataract grade 4 or more, cortical cataract ≥25% of the total lens area, or posterior subcapsular cataract ≥5% of the total lens area[12]; (6) presence of ptosis; (7) presence of macular, vitreoretinal, or optic nerve pathology; (8) abnormal anterior segment deposits consistent with pseudoexfoliation or pigment dispersion syndromes, or other signs consistent with secondary glaucomas; (9) narrow anterior chamber angle or peripheral anterior synechiae; and (10) suspected glaucoma or glaucoma.

We defined suspected glaucoma as (1) IOP >21 mmHg or (2) VCDR >0.6 or VCDR asymmetry >0.2.[13] We defined glaucoma according to the International Society for Geographical and Epidemiological Ophthalmology classification.[14]

Visual Field Assessment

Assessment of VF was conducted using the Humphrey Field Analyzer II (Carl Zeiss Meditec, Inc., Dublin, CA) with the SITA Standard 24-2 program. Experienced technicians explained the test instructions to the participants in their own language, and all VF assessments were performed using the appropriate near refractive error correction for each eye. Where the VF result was flagged as unreliable (according to the manufacturer's guidelines of FP >15% or FL >20%) or outside normal limits on the Glaucoma Hemifield Test,[15] subjects were encouraged to repeat the VF assessment for that eye a second time. In other instances, VFs may be repeated if the subject reported any problems during their first attempt.

For our analyses, we excluded VFs if they exhibited the following features: (1) lid or rim artefacts (n = 54); (2) consistent patterns of unexplained VF defects on repeated testing despite FN <33%, FP <33%, and FL <20% (n = 12); (3) Glaucoma Hemifield Test[15] outside normal limits (n = 76) or showing a general reduction of sensitivity (n = 4) despite FN, FP, and FL all being <5%. The reason for the exclusion of group 3 is that the objective of this study was to evaluate the effect of reliability indices on normal eyes that would otherwise have normal VFs; thus, the inclusion of VFs that were significantly abnormal despite being reliable would not be suited for this purpose. In addition, a single outlier VF was excluded (FN = 0%, FP = 34%, FL = 43.8%, MD = −25.79 decibels [dB], and PSD = 11.14 dB) because the FN rate of 0% was judged to be erroneously and atypically low secondary to an extremely poor response rate. Therefore, 147 VF results (7.4%) were excluded in total.

Visual Field Parameters

The FN responses are designed to reflect patient inattention to stimuli during the test. In the SITA threshold program, FN rates are calculated by catch trials where the catch trial intensity is at least 9 dB brighter than the estimated threshold at normal areas, and even brighter in areas with relative VF defects.[16] The FP responses are designed to reflect the occurrence of the patient pressing the response button despite no stimulus being seen at the time of response. In SITA, any response occurring too early to be physiologic or too late based on the patient's average response time is deemed to be outside the response time window and is thus labeled as an FP response.[17] The FL responses are detected using the method described by Heijl and Krakau[18] based on the presentation of a suprathreshold stimulus in the blind spot of the tested eye. If the patient is maintaining fixation, he or she should register no response; conversely, if a response is registered, it is recorded as an FL. In SITA, after the actual VF test procedure, the rates of false responses are calculated in postprocessing; these false response rates are included in the calculation of frequency-of-seeing (FOS) curves, and consequently the probability curves used for threshold estimation. The values of the thresholds at each tested location are in turn used to calculate the MD and PSD.[16]

The MD is a summary value that represents the overall height of the island of vision. It is the weighted mean of the 52 total