AMERICAN ACADEMY OF OPHTHALMOLOGY®

Reversal of Glaucoma Hemifield Test Results and Visual Field Features in Glaucoma

Mengyu Wang, PhD,[1] Louis R. Pasquale, MD,[2,3] Lucy Q. Shen, MD,[2] Michael V. Boland, MD, PhD,[4] Sarah R. Wellik, MD,[5] Carlos Gustavo De Moraes, MD,[6] Jonathan S. Myers, MD,[7] Hui Wang, PhD,[1,8] Neda Baniasadi, MD, PhD,[1] Dian Li, MS,[1] Rafaella Nascimento E. Silva, MD,[2] Peter J. Bex, PhD,[9] Tobias Elze, PhD[1,10]

Purpose: To develop a visual field (VF) feature model to predict the reversal of glaucoma hemifield test (GHT) results to within normal limits (WNL) after 2 consecutive outside normal limits (ONL) results.

Design: Retrospective cohort study.

Participants: Visual fields of 44 503 eyes from 26 130 participants.

Methods: Eyes with 3 or more consecutive reliable VFs measured with the Humphrey Field Analyzer (Swedish interactive threshold algorithm standard 24-2) were included. Eyes with ONL GHT results for the 2 baseline VFs were selected. We extracted 3 categories of VF features from the baseline tests: (1) VF global indices (mean deviation [MD] and pattern standard deviation), (2) mismatch between baseline VFs, and (3) VF loss patterns (archetypes). Logistic regression was applied to predict the GHT results reversal. Cross-validation was applied to evaluate the model on testing data by the area under the receiver operating characteristic curve (AUC). We ascertained clinical glaucoma status on a patient subset (n = 97) to determine the usefulness of our model.

Main Outcome Measures: Predictive models for GHT results reversal using VF features.

Results: For the 16 604 eyes with 2 initial ONL results, the prevalence of a subsequent WNL result increased from 0.1% for MD < -12 dB to 13.8% for MD ≥ -3 dB. Compared with models with VF global indices, the AUC of predictive models increased from 0.669 (MD ≥ -3 dB) and 0.697 (-6 dB $\leq$ MD < -3 dB) to 0.770 and 0.820, respectively, by adding VF mismatch features and computationally derived VF archetypes ($P < 0.001$ for both). The GHT results reversal was associated with a large mismatch between baseline VFs. Moreover, the GHT results reversal was associated more with VF archetypes of nonglaucomatous loss, severe widespread loss, and lens rim artifacts. For a subset of 97 eyes, using our model to predict absence of glaucoma based on clinical evidence after 2 ONL results yielded significantly better prediction accuracy (87.7%; $P < 0.001$) than predicting GHT results reversal (68.8%) with a prescribed specificity 67.7%.

Conclusions: Using VF features may predict the GHT results reversal to WNL after 2 consecutive ONL results. *Ophthalmology 2018;125:352-360 © 2017 by the American Academy of Ophthalmology*

Supplemental material available at www.aaojournal.org.

The diagnosis of glaucoma relies heavily on the use of standard automated perimetry to measure visual field (VF) loss. The glaucoma hemifield test (GHT) is an important measure in standard automated perimetry to assist in the interpretation of VFs measured with the Humphrey Field Analyzer (Carl Zeiss Meditec, Dublin, CA).[1,2] The GHT is partially inspired by retinal nerve fiber anatomic characteristics and compares symmetric VF sectors between the upper and lower hemifields.[1] The GHT has 6 possible outcomes: within normal limits (WNL), borderline, outside normal limits (ONL), general reduction of sensitivity, abnormally high sensitivity, and borderline or general reduction of sensitivity. Outside normal limits appears when the differences between a matched pair of mirrored zones exceeds the differences of 99% of individuals in a normal population or both members of 2 paired zones are more abnormal than 99.5% of individuals in a normal population. Borderline denotes the case where 2 paired zones are more abnormal than 97% of the individuals, whereas the abnormality of the paired zones do not meet criteria for ONL. General reduction of sensitivity appears when both conditions for ONL are not met and the best region of the VF is more abnormal than 99.5% of the individuals in a normal population. Abnormally high sensitivity denotes that the best region of the VF has higher sensitivity than 99.5% of the individuals in a normal population, which may indicate low reliability of the VF test. Within normal limits is assigned to the VF when none of those aforementioned conditions are met.

To reduce false discovery, 2 consecutive GHT ONL results are recommended before considering a diagnosis of glaucomatous VF loss.[2] In addition, it has been shown that the sensitivity of GHT for early glaucomatous VF loss is

https://doi.org/10.1016/j.ophtha.2017.09.021
ISSN 0161-6420/17

limited,[3] whereas the sensitivity of the GHT for the full range of glaucomatous VF loss is high.[4] Assuming that glaucomatous VF loss is irreversible, a conversion from 2 consecutive GHT ONL results to WNL results represented a GHT results reversal in this study. For the purpose of this work, a borderline GHT result on a third test did not constitute a GHT results reversal.

In this study, we aimed to predict the occurrence of GHT results reversal to WNL using VF features. The VF features include the VF global indices, VF mismatch measures between baseline VFs, and previously described computationally derived representative VF loss patterns (archetypes).[5] The VF mismatch measures capture the variation and similarity between the 2 baseline VFs, and the archetype decompositions quantify the spatial patterns of VF loss. Our model aims to support clinicians quantitatively in the decision of whether 2 consecutive ONL GHT results will revert to WNL results.

Methods

The VF results used for this study were obtained by the Glaucoma Research Network, a consortium including the following glaucoma centers: Massachusetts Eye and Ear (MEE), Wilmer Eye Institute, New York Eye and Ear Infirmary, Bascom Palmer Eye Institute, and Wills Eye Hospital. The institutional review boards of each institution approved this retrospective study. This study adhered to the tenets of the Declaration of Helsinki and all federal and state laws, including the Health Insurance Portability and Accountability Act of 1996.

Participants and Data

From our large dataset of Swedish interactive thresholding algorithm standard 24-2 VFs measured with the Humphrey Field Analyzer between June 1999 and Oct 2014, all eyes with at least 3 reliable consecutively measured VFs were selected. The reliability criteria for VF selection were fixation loss of 33% or less, false-negative rates of 20% or less, and false-positive rates of 20% or less.[6,7] The cutoffs for fixation loss and false-positive rate are based on published recommendations.[8,9] The cutoff for false-negative rate is consistent with criteria used to develop archetype analysis[5] and have been adopted in the identification of glaucoma in population-based studies.[10,11] Subsequently, a subset of eyes was selected additionally such that: the GHT results for the first 2 VFs were ONL and the GHT results of the third VF were any of WNL, borderline, or ONL. The total deviation (TD) values from each of the 52 locations tested in the 24-2 pattern were extracted and used to derive the VF mismatch features and the VF loss patterns.

Statistical Analyses

Initially, the proportions of eyes with GHT results reversal from ONL at baseline to WNL on the second test for all VF loss severities were calculated. For the subset with 2 consecutive ONL results, the proportions of eyes with GHT results reversal on the third measurement to WNL for all VF loss severities also were evaluated. All statistical analyses were performed using R software (Version 3.3.1, R Foundation, Vienna, Austria).[12]

Feature Extraction

For the subset of eyes with 2 consecutive ONL results, we extracted 3 groups of features from baseline VFs: the average VF global indices, VF mismatch measures between baseline VFs, and the archetype decompositions of the mean baseline VFs. The global indices extracted included the mean deviation (MD), the pattern standard deviation (PSD), and the MD and PSD differences between the second and first VFs. The VF mismatch measures calculated include the standard deviation of the TD difference in all 52 locations between baseline VFs and the similarity index of the TDs between baseline VFs measured by the cosine similarity, a standard similarity measure between 2 vectors that measures the cosine of the angle between them.[13,14] For the archetype decomposition to quantify the VF spatial patterns, the average VFs (i.e., average TD values at all 52 locations) of the first 2 VFs were decomposed into 16 VF patterns (archetypes) computationally derived as described previously (Fig 1A).[5] The VF loss patterns then were represented by the decomposition coefficients, which sum up to 100% (Fig 1B). In short, the 16 VF archetypes were identified by an unsupervised machine learning method (archetypal analysis) based on more than 13 000 reliable VFs. Nine of those archetypes represent clinically recognizable glaucomatous patterns[5] with similarity to previously described patterns determined by manual inspection of VF data in the Ocular Hypertension Treatment Study[15] and confirmed by a clinical correlation study[16]: archetypes 8 and 13 (altitudinal VF loss); archetypes 9, 10, and 16 (partial arcuate defects); archetypes 3 and 5 (nasal step); and archetypes 14 and 16 (paracentral). Archetype 2 was associated with both glaucomatous VF loss and a higher occurrence of ptosis.[16] Archetype 1 represents the normal VF. All other archetypes represent clinical conditions different from glaucoma, such as hemianopia (archetypes 12 and 15).

Statistical Modelling

Logistic regression was applied to predict GHT results reversal to WNL after 2 consecutive GHT ONL results using the VF features as independent variables.[17] The technique of weighted error penalization was used to mitigate the underestimation of GHT results reversals because of an imbalanced dataset.[18,19] Stepwise regression was performed to select the optimal feature combination that predicts the GHT results reversal based on Bayesian information criterion.[20] The regression analyses were implemented for eyes with MD of −3 dB or more and MD of −6 dB or more and less than −3 dB, respectively. Ten-fold cross-validation[21] was applied to evaluate the predictive model performance by the area under the receiver operating characteristic curve (AUC).[22] The AUCs of our optimal models to predict GHT results reversal were compared with the AUC performance of models that included only VF global indices and models that also included the VF global indices plus VF mismatch measures. We used cross-validation to test the performance of our model on the data that are not used in model training.[21,23] In short, the dataset in this study was partitioned into 10 parts, and each of the 10 subsets was used once as testing partitions, whereas the model was trained on the 9 remaining partitions. Thus, we ensured that the AUCs for model evaluation were calculated on different data subsets than those used for generating the models.

Because clinical data were available only in the MEE dataset, we excluded it from the training dataset and used its clinical data to test the robustness of our model. The AUC performance of the model was evaluated. The jackknife resampling method was used to compute the AUC confidence interval (CI).[24] For a subset of the MEE data, an assessment of glaucoma status at the time of the third