Published in final edited form as:

Ophthalmology. 2018 March ; 125(3): 352–360. doi:10.1016/j.ophtha.2017.09.021.

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

Mengyu Wang, PhD¹, Louis R. Pasquale, MD^{2,3}, Lucy Q. Shen, MD², Michael V. Boland, MD, PhD⁴, Sarah R. Wellik, MD⁵, Carlos Gustavo De Moraes, MD⁶, Jonathan S. Myers, MD⁷, Hui Wang, PhD^{1,8}, Neda Baniasadi, MD, PhD¹, Dian Li, MS¹, Rafaella Nascimento E. Silva, MD², Peter J. Bex, PhD⁹, Tobias Elze, PhD^{1,10}

¹Schepens Eye Research Institute, Harvard Medical School, Boston, Massachusetts.

²Massachusetts Eye and Ear, Harvard Medical School, Boston, Massachusetts. ³Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts. ⁴Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland. ⁵Bascom Palmer Eye Institute, University of Miami School of Medicine, Miami, Florida.

⁶Edward S. Harkness Eye Institute, Columbia University Medical Center, New York, New York.

⁷Wills Eye Hospital, Thomas Jefferson University, Philadelphia, Pennsylvania. ⁸Institute for Psychology and Behavior, Jilin University of Finance and Economics, Changchun, China.

⁹Department of Psychology, Northeastern University, Boston, Massachusetts. ¹⁰Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany.

Abstract

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

Correspondence: Tobias Elze, PhD, Schepens Eye Research Institute, Harvard Medical School, 20 Staniford Street, Boston, MA 02114. tobias-elze@tobias-elze.de.

Author Contributions:

Conception and design: M.Wang, Elze

Analysis and interpretation: M.Wang, Elze

Data collection: M.Wang, Pasquale, Shen, Boland, Wellik, De Moraes, Myers, Li, Silva

Obtained funding: none

Overall responsibility: M.Wang, Pasquale, Shen, Boland, Wellik, De Moraes, Myers, H.Wang, Baniasadi, Bex, Elze

HUMAN SUBJECTS: Human subjects were included in this study. The institutional review boards of Massachusetts Eye and Ear, Wilmer Eye Institute, New York Eye and Ear Infirmary, Bascom Palmer Eye Institute, and Wills Eye Hospital 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.

Footnotes and Financial Disclosures

Supplemental material available at www.aaojournal.org.

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 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.

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 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 arche-type analysis⁵ 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 patterns⁵ with similarity to previously described patterns determined by manual inspection of VF data in the Ocular Hypertension Treatment Study¹⁵ and confirmed by a clinical correlation study¹⁶: 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-validation²¹ 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 VF test was made based on the consensus of 2

glaucoma specialists (L.Q.S. and L.R.P.) masked to study results by reviewing the fundus photography for glaucomatous optic disc changes and OCT images for characteristic nerve fiber layer thinning closest to the test date of the third VF. When structural data were equivocal, a reliable VF that postdated the third test was used to confirm the presence or absence of glaucoma. In addition, eye surgical history was extracted from medical records for the subset. Thus, instead of defining GHT results reversal as 2 ONL test results followed by WNL results on the third test, we alternatively defined GHT results reversal as 2 ONL results and absence of clinical glaucoma. We evaluated the AUC performance and prediction accuracy using our model that was trained to predict GHT results reversal defined as 2 GHT ONL results followed by WNL results.

Results

We selected 44 503 eyes of 26 130 patients (mean age, 63.8 ± 14.3 years) with at least 3 reliable annually measured VFs. Of these, 16 604 eyes of 12 688 patients (mean age, 66.4 ± 13.5 years) demonstrated GHT ONL results on the first 2 visits. The GHT results reversal prevalence to WNL results after 2 consecutive ONL determinations increased from 0.1% for MD less than -12 dB to 13.8% for MD of -3 dB or more (see more details in Table S1, available at www.aaojournal.org). The GHT results are most relevant for the diagnosis of glaucomatous VF loss at mild stage (MD, -6 dB).^{1,3,25} Therefore, we specifically analyzed all eyes with mild VF loss and the first 2 VFs with GHT ONL results. For the resulting 6481 eyes, 9.2% reversed to WNL results at the third visit.

We then selected for MD of -3 dB or more with the first 2 VFs showing GHT ONL results. The dataset yielded 2199 eyes of 2077 patients (mean age, 64.3 ± 12.0 years) with 13.8% showing GHT results reversals. Figure 2 shows the best predictive model for MD of -3 dB or more selected by stepwise regression to predict the GHT results reversals with the optimal parameter combination. Mean deviation was associated positively and PSD was associated negatively with GHT results reversals. The standard deviation of the TD difference was associated positively and the similarity index of TDs was associated negatively with GHT results reversals. Based on stepwise regression, 8 of the original 16 archetypes were selected. Seven archetypes (archetypes 2, 4, 5, 7, 9, 11, and 12) were associated positively and 1 archetype (archetype 16) was associated negatively with GHT results reversals. The MD and PSD difference between baseline VFs did not remain in the optimal feature combination. Compared with the model with VF global indices, the AUC performance of cross-validation to predict GHT results reversals increased significantly from 0.669 (95% CI, 0.668–0.671) to 0.745 (95% CI, 0.744–0.746; $P < 0.001$) by adding VF mismatch features. Furthermore, the AUC performance of the models increased significantly to 0.770 (95% CI, 0.769–0.772; $P < 0.001$) by adding the archetype features. If we chose a false-positive rate for our model of 33% (denoted by the blue cross in Fig 2B) for instance, 74.5% of the GHT results reversals could be predicted correctly. The corresponding probability threshold was 0.51. For eyes with MD of -6 dB or more and less than -3 dB, the AUC for the best predictive model was 0.820 (95% CI, 0.819–0.820; Figure S1, available at www.aaojournal.org). The detailed coefficients of the logistic regression models to predict the GHT results reversal can be found in Table S3 (available at www.aaojournal.org) (MD -3 dB) and Table S4 (available at www.aaojournal.org) (-6 dB $\leq$ MD < -3 dB).

Figure 3 shows a patient with 3 consecutive GHT ONL results (Fig 3A) and a patient with GHT results reversal for MD of -3 dB or more (Fig 3B). Consistent with the relationship between the VF features and the occurrence of GHT results reversals shown in Figure 2A, the standard deviation of the TD difference is higher and the similarity index of TDs between the baseline VFs of the patient with GHT results reversal is lower than those of the patient with 3 consecutive GHT ONL results. Furthermore, the substantial presence of archetype 11 (9.7%) suggests that lens rim artifacts also contribute to the GHT results reversal as shown in Figure 3B. Likewise, the presence of archetypes only related to glaucomatous loss, namely, archetype 2 (21.5%), archetype 4 (4.7%), and archetype 5 (15.8%), provides evidence that the third GHT is likely to reveal ONL results after 2 GHT ONL results (Fig 3A). In contrast, the average MD and PSD of the patient with GHT results reversal for the baseline VFs are not higher and lower than the patient with 3 GHT ONL results as expected, respectively. For eyes with MD of -6 dB or more and less than -3 dB, Figure S2 shows a patient with 3 consecutive GHT ONL results (Fig 2SA; www.aaojournal.org) and a patient with GHT results (Fig 2SB; www.aaojournal.org).

Six hundred forty-four eyes of 576 patients (mean age, 64.2 ± 14.3 years) with MD of -6 dB or more from MEE were used to test our model performance trained with the data from the other Glaucoma Research Network sites. Fifty of 644 eyes (7.8%) showed GHT results reversals from 2 ONL results to WNL results. The AUC performance to predict the GHT results reversals was 0.870 (95% CI, 0.870–0.870). As shown in Figure 4, 92.0% (95% CI, 92.0%–92.0%) of the GHT results reversals were predicted correctly by our model, with the tradeoff of 33.3% of the patients with 3 consecutive GHT ONL results to be misclassified. The corresponding probability threshold was 0.44. At this specificity level, any VFs with predicted probability larger than 0.44 were classified to be GHT results reversals.

We additionally selected 97 eyes that included 48 eyes with GHT results reversals and 49 eyes without GHT results reversals. Of the 40 eyes diagnosed with glaucoma based on clinical data, 20.0% showed GHT results reversal. Of the 57 eyes without glaucoma, 70.2% showed GHT results reversal. The impact of eye surgeries (12.4% of eyes) on GHT results reversal was assessed further in the Appendix (“The Impact of Eye Surgical History,” available at www.aaojournal.org). The AUC for predicting the GHT results reversals for the subset of 97 eyes was 0.774 (95% CI, 0.773–0.775). Our model correctly predicted 68.8% (95% CI, 68.6%–68.9%) of the GHT results reversals, with the tradeoff of 33.3% of misclassification for the patients with 3 consecutive GHT ONL results.

The AUC for predicting the GHT results reversals defined by 2 ONLs with absence of glaucoma was 0.773 (95% CI, 0.772–0.774). Our model correctly predicted 87.7% (95% CI, 87.6%–87.8%) of the GHT results reversals, with the tradeoff of 33.3% of misclassification. Although the AUC performance to predict 2 ONL results with absence of glaucoma is not significantly different ($P = 0.18$) from the AUC performance to predict the GHT results reversals, the prediction accuracy for 2 ONL results with absence of glaucoma was significantly higher ($P < 0.001$) than that of predicting the GHT results reversals.

Discussion

The GHT is a standard parameter included in the Humphrey Field Analyzer that aims to aid clinicians in the diagnosis of glaucomatous VF loss. Our results demonstrated that in VFs with mild severity, GHT can revert from 2 consecutive ONL results to WNL results in a significant portion of eyes (13.8% for MD -3 dB). Our results suggest that the inclusion of computationally derived VF mismatch and archetype features significantly improves the prediction of whether 2 consecutive GHT ONL results will revert to WNL results compared with models with VF global indices alone. As expected, GHT results reversals are related positively to MD and negatively to PSD. The standard deviation of the TD difference and the similarity index of TDs that characterize the consistency of the baseline VFs are related positively and negatively to GHT results reversals, respectively.

For eyes with MD of -3 dB or more, the occurrence of GHT results reversals is more likely to be associated with archetypes related to nonglaucomatous, severe widespread VF loss and lens rim artifacts, and less likely to be associated with archetypes related to typical early glaucomatous VF loss (Fig 2). For example, the 3 archetypes with larger positive coefficients are archetypes 7, 11, and 12. Archetype 7 denotes central VF defects that are more typical for macular disorders. Archetype 11 typically is associated with VF measurement rim artifacts related to the use of high hyperopic correcting lenses. Archetype 12 is representative of hemianopia, which typically is caused by stroke. The high and positive coefficients of the nonglaucomatous archetypes 7, 11, and 12 in the predictive model of GHT results reversals therefore are explained. The 4 archetypes with smaller positive coefficients are archetypes 2, 4, 5, and 9, and the only archetype with negative coefficient is archetype 16. Archetypes 2, 4, 5, 9, and 16 are all related to early glaucomatous VF loss.¹⁶

Our model may provide additional aid to clinicians when interpreting GHT ONL results. Our model generates probabilities for GHT results reversals from VF features calculated from baseline VFs. For a given false-positive rate of 33.3%, the decision probability thresholds are 0.48 (MD -3 dB) and 0.51 (-6 dB $\leq$ MD < -3 dB), respectively. A value greater than the respective threshold would falsely predict a GHT results reversal in 33.3% of the cases with 3 consecutive GHT ONL results, but correctly predict it in 74.5% for MD of -3 dB or more and in 83.9% for MD of -6 dB or more and less than -3 dB of the GHT results reversals, respectively. Because the mathematical decomposition of a VF measurement into VF loss archetypes is publicly available⁵ and the VF mismatch features also are relatively easy to calculate, our models can be implemented widely.

The AUC and accuracy for predicting GHT results reversals in the MEE validation data were 0.870 (95% CI, 0.870–0.870) and 92.0% (95% CI, 92.0%–92.0%), respectively, with a specificity of 66.7% prescribed, which were consistent with our cross-validation results over all Glaucoma Research Network data, and thus demonstrated that our model performance was robust. The AUC performance of our model on the MEE validation data was better than the AUC performance of cross-validation on all institution data. The reason for the better predicted results may speak to the robustness of the model or merely may represent chance occurrence of a prediction that was more than the upper bound of the 95% CI.

For the additional subset of 97 eyes with clinical data, the model performance (0.774; 95% CI, 0.773–0.775) was significantly lower ($P < 0.001$) than the model performance with all MEE data. The lower model performance was expected, because the GHT results reversal frequency was set to be 50% in this subset and is significantly higher than the GHT results reversal frequency of the overall MEE data (7.8%).

Instead of predicting GHT results reversals defined solely based on VFs, we alternatively predicted the GHT results reversals defined by 2 ONL results with negative glaucoma diagnosis. The accuracy to predict the GHT results reversals defined based on glaucoma diagnosis significantly outperformed ($P < 0.001$) the predicting accuracy for GHT results reversals. Interestingly, our model trained based on fitting the GHT results reversals defined solely by VFs was better at predicting patients with 2 ONL results and no glaucoma diagnosis. The favorable results from our model are encouraging and suggest that a model purely based on parameters from 2 previous VFs can predict the glaucoma diagnosis. Although clinicians naturally rely on other clinical data to make glaucoma management decisions, especially information about optic nerve integrity, our model based purely on VF features may augment the clinical decision-making process.

Studies have shown that eye surgeries, including intraocular pressure-lowering surgeries,^{26–29} cataract extraction,^{30,31} and ranibizumab treatment,^{32,33} can lead to enhanced VF sensitivity, which may trigger GHT results reversals. For our 97 eyes, we did not find any significant effects of eye surgeries on GHT results reversals (Table S2, available at www.aaajournal.org). However, the sample size may limit our ability to assess fully the effect of eye surgeries on GHT results reversal.

This study was limited in several respects. First, we do not have structural data for our multicenter dataset, which may serve as a guide to assess the significance of GHT results. We also do not have data on the prior experience with VF testing for patients in this study. Thus, the degree that any learning curve may be affecting GHT results reversals cannot be ascertained from this dataset. Also, although we benefit from a large de-identified clinical dataset, we manually extracted the clinical diagnosis (e.g., eye surgical history) and assessed the status of glaucoma diagnosis for only a small proportion of patients ($n = 97$). Finally, these data apply only to VFs with reliability parameters set in this study. Although there is strong precedent for the fixation loss rate and false-positive rate we set, allowing VFs with higher false-negative rates, which can be a sign of early glaucoma,³⁴ may yield different results. We also did not study predicting the opposite scenario of 2 GHTs with WNL findings with positive glaucoma diagnosis at the time of the second VF test. It is definitely worthwhile to study this scenario, which needs substantial effort involving manual assessment rather than using our large retrospective clinical dataset constructed by automatic data processing. The sparsity of GHT results reversals in our dataset may impair our model performance.³⁵ In this study, we used a weighted error penalization approach to mitigate the underestimation of GHT results reversals.^{18,19} In future work, we will address the sparsity of GHT results reversals by using advanced machine learning techniques such as data over-sampling or undersampling and synthetic samples generations.^{36–38}

The prediction accuracy of GHT results reversal of 74.5% (for MD ≥ -3 dB) and 83.9% (for -6 dB $\leq$ MD < -3 dB) at a cost of 33.3% of the eyes with 3 consecutive GHT ONL results to be misclassified may not be considered to be highly predictive. Although using 3 VFs to confirm glaucomatous VF loss can be better than relying on 2 test results to diagnosis glaucoma (13.8% GHT results reversal from 2 ONL results to WNL results for MD ≥ -3 dB), simply using 3 VFs without our predictive model would allow 86.2% of the eyes with 3 ONL results to be deferred for treatment. In comparison, with our model, we can predict correctly 74.5% (for MD ≥ -3 dB), 83.9% (for -6 dB $\leq$ MD < -3 dB), and 92.0% (for the MEE data) of the GHT results reversals at a cost of 33.3% of the eyes with 3 ONL results to be misclassified. Our model may help clinicians to strike a balance between the cost savings associated with deferral of treatment despite 2 consecutive GHT ONL results versus the sight preservation associated with treatment after a second abnormal GHT result is recorded.

To summarize, our study demonstrated that the occurrence of GHT results reversals can be predicted by assessing VF mismatch features and quantifying the VF loss patterns in addition to the VF global indices. Our results may assist clinicians with determining whether GHT ONL results represent true glaucomatous VF loss.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by the Lions Foundation (M.W., N.B., T.E.); the Grimshaw-Gudewicz Foundation (M.W., N.B., T.E.); Research to Prevent Blindness, Inc., New York, New York (M.W., N.B., T.E.; and departmental grant to C.G.D.M.); the Harvard Glaucoma Center of Excellence (M.W., T.E., L.R.P., L.Q.S.); the China Scholarship Council (H.W.); the Eleanor and Miles Shore Fellowship (L.Q.S.); BrightFocus Foundation (M.W., T.E.); NEI Core Grant P30EYE003790 (M.W., N.B., T.E.).

Financial Disclosure(s):

The author(s) have made the following disclosure(s): M.V.B.: Financial support – Alcon

J.S.M.: Financial support – Haag Streit

T.E.: Patent – United States PCT/US2014/052414

Abbreviations and Acronyms:

AUC	area under the receiver operating characteristic curve
CI	confidence interval
GHT	glaucoma hemifield test
MD	mean deviation
MEE	Massachusetts Eye and Ear
ONL	outside normal limits
PSD	pattern standard deviation

TD	total deviation
VF	visual field
WNL	within normal limits

References

1. Åsman P, Heijl A. Glaucoma hemifield test: automated visual field evaluation. *Arch Ophthalmol*. 1992;110:812–819. [PubMed: 1596230]
2. Katz J, Quigley HA, Sommer A. Detection of incident field loss using the glaucoma hemifield test. *Ophthalmology*. 1996;103:657–663. [PubMed: 8618767]
3. Zalta AH. Limitations of the glaucoma hemifield test in identifying early glaucomatous field loss. *Ann Ophthalmol*. 2000;32:33–45.
4. Sekhar GC, Naduvilath TJ, Lakkai M, et al. Sensitivity of swedish interactive threshold algorithm compared with standard full threshold algorithm in humphrey visual field testing. *Ophthalmology*. 2000;107:1303–1308. [PubMed: 10889102]
5. Elze T, Pasquale LR, Shen LQ, et al. Patterns of functional vision loss in glaucoma determined with archetypal analysis. *J Royal Soc Interface*. 2015;12(103).
6. Leung CKS, Liu S, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in glaucoma: a prospective analysis with neuroretinal rim and visual field progression. *Ophthalmology*. 2011;118:1551–1557. [PubMed: 21529958]
7. Leung CKS, Yu M, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: patterns of retinal nerve fiber layer progression. *Ophthalmology*. 2012;119:1858–1866. [PubMed: 22677426]
8. Birt CM, Shin DH, Samudrala V, et al. Analysis of reliability indices from Humphrey visual field tests in an urban glaucoma population. *Ophthalmology*. 1997;104:1126–1130. [PubMed: 9224465]
9. Newkirk MR, Gardiner SK, Demirel S, et al. Assessment of false positives with the Humphrey Field Analyzer II perimeter with the SITA algorithm. *Invest Ophthalmol Vis Sci*. 2006;47: 4632–4637. [PubMed: 17003461]
10. Pasquale LR, Kang JH, Manson JE, et al. Prospective study of type 2 diabetes mellitus and risk of primary open-angle glaucoma in women. *Ophthalmology*. 2006;113:1081–1086. [PubMed: 16757028]
11. Pasquale LR, Willett WC, Rosner BA, et al. Anthropometric measures and their relation to incident primary open-angle glaucoma. *Ophthalmology*. 2010;117:1521–1529. [PubMed: 20382429]
12. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Core Team <http://www.R-project.org>; 2015 Accessed September 23, 2016.
13. Nguyen HV, Bai L. Cosine similarity metric learning for face verification. In: Kimmel R, Klette R, Sugimoto A, eds. *Asian Conference on Computer Vision Berlin/Heidelberg, Germany: Springer*; 2010:709–720.
14. Steinbach M, Karypis G, Kumar V, et al. A comparison of document clustering techniques. In: *KDD Workshop on Text Mining Vol 400*. Boston: The Association for Computing Machinery's (ACM) Special Interest Group (SIG) on Knowledge Discovery and Data Mining; 2000:525–526.
15. Keltner JL, Johnson CA, Cello KE, et al. Classification of visual field abnormalities in the Ocular Hypertension Treatment Study. *Arch Ophthalmol*. 2003;121:643–650. [PubMed: 12742841]
16. Cai S, Elze T, Bex PJ, et al. Clinical correlates of computationally derived visual field defect archetypes in patients from a glaucoma clinic. *Curr Eye Res*. 2017;42(4):568–574. [PubMed: 27494512]
17. Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. Vol 398 Hoboken, NJ: John Wiley & Sons; 2013.
18. Agterberg F, Bonham-Carter G, Cheng QM, et al. Weights of evidence modeling and weighted logistic regression for mineral potential mapping. *Comput Geol*. 1993;25:13–32.
19. King G, Zeng L. Logistic regression in rare events data. *Political Analysis*. 2001;9:137–163.

20. Jennrich RI. Stepwise regression. *Statistical Methods for Digital Computers*. 1977;3:58–75.
21. Kohavi R, Mellish CS. A study of cross-validation and bootstrap for accuracy estimation and model selection; *International Joint Conference on Artificial Intelligence*; Stanford, CA. 1995. 1137–1145.
22. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29–36. [PubMed: 7063747]
23. Efron B, Gong G. A leisurely look at the bootstrap, the jack-knife, and cross-validation. *Am Statist*. 1983;37:36–48.
24. Efron B The jackknife, the bootstrap and other resampling plans. *Society of Industrial and Applied Mathematics*; Philadelphia, PA; 1982.
25. Åsman P, Heijl A. Evaluation of methods for automated hemifield analysis in perimetry. *Arch Ophthalmol*. 1992;110: 820–826. [PubMed: 1596231]
26. Wright TM, Goharian I, Gardiner SK, et al. Short-term enhancement of visual field sensitivity in glaucomatous eyes following surgical intraocular pressure reduction. *Am J Ophthalmol*. 2015;159:378–385. [PubMed: 25447113]
27. Waisbourd M, Ahmed OM, Molineaux J, et al. Reversible structural and functional changes after intraocular pressure reduction in patients with glaucoma. *Graefes Arch Clin Exp Ophthalmol*. 2016;254:1159–1166. [PubMed: 26995555]
28. Caprioli J, de Leon JM, Azarbod P, et al. Trabeculectomy can improve long-term visual function in glaucoma. *Ophthalmology*. 2016;123:117–128. [PubMed: 26602970]
29. Cohen SL, Rosen AI, Tan X, et al. Improvement of the visual field index in clinical glaucoma care. *Canadian J Ophthalmol*. 2016;51:445–451.
30. Kouchehi B, Nouri-Mahdavi K, Patel G, et al. Visual field changes after cataract extraction: the AGIS experience. *Am J Ophthalmol*. 2004;138:1022–1028. [PubMed: 15629295]
31. Siddiqui MR, Khairy HA, Azuara-Blanco A. Effect of cataract extraction on sita perimetry in patients with glaucoma. *J Glaucoma*. 2007;16:205–208. [PubMed: 17473731]
32. Frennesson C, Nilsson UL, Peebo BB, et al. Significant improvements in near vision, reading speed, central visual field and related quality of life after ranibizumab treatment of wet age-related macular degeneration. *Acta Ophthalmol*. 2010;88: 420–425. [PubMed: 19678811]
33. Squirrell DM, Mawer NP, Mody CH, et al. Visual outcome after intravitreal ranibizumab for wet age-related macular degeneration: a comparison between best-corrected visual acuity and microperimetry. *Retina*. 2010;30:436–442. [PubMed: 19996834]
34. 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. [PubMed: 10892863]
35. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowledge Data Eng*. 2009;21:1263–1284.
36. He H, Bai Y, Garcia EA, et al. Adasyn: adaptive synthetic sampling approach for imbalanced learning. In: *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE Int Joint Conf on IEEE 2008*;1322–1328.
37. Han H, Wang WY, Mao BH. Borderline-smote: a new over-sampling method in imbalanced data sets learning. In: *Ada-syn: adaptive synthetic sampling approach for imbalanced learning. 2008 IEEE International Joint Conference on Neural Networks (International Conference on Intelligent Computing) Berlin/Heidelberg, Germany: Springer; 2005*:878–887.
38. Maalouf M, Trafalis TB. Robust weighted kernel logistic regression in imbalanced and rare events data. *Comput Stat Data Anal*. 2011;55:168–183.

Figure 1.

Illustration of quantifying visual field (VF) loss patterns with archetypes (ATs): **(A)** the 16 computationally derived archetypes and **(B)** an example of the VF decomposition to the VF archetypes

Figure 2.

The best predictive model selected by stepwise regression to predict glaucoma hemfield test results reversals for mean deviation (MD) of -3 dB or more: **(A)** parameter coefficients of the logistic regression model and **(B)** receiver operating characteristic curve. The *blue cross* illustrates the decision threshold for a fixed false-positive rate of one third, as described in the text. AT = archetype; AUC = area under the receiver operating characteristic curve; CI = confidence interval; PSD = pattern standard deviation; SD-TDD = standard deviation of the total deviation difference between the 2 baseline visual fields; SI-TD = similarity index of the total deviations between the 2 baseline visual fields.

Figure 3.

Examples of visual fields (VFs) (A) with 3 consecutive glaucoma hemifield test (GHT) results outside normal limits (ONL) and (B) with GHT results reversal from 2 consecutive GHT ONL results. Visual fields are decomposed into the combination of archetypes. Nonconsidered archetypes (ATs) are those archetypes that are not in the parameter set of the best predictive model. Avg. = average; MD = mean deviation; PSD = pattern standard deviation; SD-TDD = standard deviation of the total deviation difference between the 2

baseline visual fields; SI-TD = similarity index of the total deviations between the 2 baseline visual fields; WNL = within normal limits.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4.

Receiver operating characteristic curve that illustrates our model performance on the validation subset of the Massachusetts Eye and Ear data including 644 eyes with mean deviation of -6 dB or more. The *blue cross* illustrates the decision threshold for a fixed false-positive rate of one third, as described in the text. AUC = area under the receiver operating characteristic curve; CI = confidence interval.