

Predicting Progressive Glaucomatous Optic Neuropathy Using Baseline Standard Automated Perimetry Data

Shaban Demirel,¹ Brad Fortune,¹ Juanjuan Fan,² Richard A. Levine,² Rodrigo Torres,¹ Hau Nguyen,¹ Steven L. Mansberger,¹ Stuart K. Gardiner,¹ George A. Cioffi,¹ and Chris A. Johnson³

PURPOSE. To test the hypothesis that specific locations and patterns of threshold findings within the visual field have predictive value for progressive glaucomatous optic neuropathy (pGON).

METHODS. Age-adjusted standard automated perimetry thresholds, along with other clinical variables gathered at the initial examination of 168 individuals with high-risk ocular hypertension or early glaucoma, were used as predictors in a classification tree model. The classification variable was a determination of pGON, based on longitudinally gathered stereo optic nerve head photographs. Only data for the worse eye of each individual were included. Data from 100 normal subjects were used to test the specificity of the models.

RESULTS. Classification tree models suggest that patterns of baseline visual field findings are predictive of pGON with sensitivity 65% and specificity 87% on average. Average specificity when data from normal subjects were run on the models was 69%.

CONCLUSIONS. Classification trees can be used to determine which visual field locations are most predictive of poorer prognosis for pGON. Spatial patterns within the visual field convey useable predictive information, in most cases when thresholds are still well within the classically defined normal range. (*Invest Ophthalmol Vis Sci.* 2009;50:674–680) DOI: 10.1167/iovs.08-1767

Baseline results from standard automated perimetry (SAP) have been used to quantify the risk of conversion to glaucoma in individuals with ocular hypertension (OH) and the risk of progression in patients with early glaucoma (EG).^{1,2} In previous investigations of this predictive capacity, visual field data in terms of global indices were used, such as mean deviation (MD), pattern or corrected pattern SD (PSD/CPSD) or the glaucoma hemifield test (GHT).³ Other studies have used scoring systems that also condense visual field data into a single value, such as the AGIS (Advanced Glaucoma Intervention

Study) and CIGTS (Collaborative Initial Glaucoma Treatment Study).^{4–6}

This report describes the application of a statistical machine learning technique, called classification and regression tree (CART) analysis,⁷ to a longitudinal dataset collected from patients with high-risk OH or EG. The principal goal was to use data from a baseline SAP examination, together with common clinical and demographic variables, to predict which patients would go on to exhibit progressive glaucomatous optic neuropathy (pGON). In addition, we addressed the hypothesis that certain locations and patterns of threshold findings within the visual field convey greater predictive information for pGON than do other locations. We used all visual field test locations individually and avoided using indices that condense the data contained in a visual field to a single number.

Breiman et al.⁷ first described CART as a flexible, nonparametric, data-mining tool that, unlike traditional statistical models, makes few assumptions about the distribution of the underlying data. For example, it deals well with independent variables of mixed type that are correlated, high-dimensional, and inhomogeneous. CART analysis produces decision rules arranged in a tree structure that are relatively simple to interpret for nonstatisticians. CART also performs well with some missing data by exploiting correlation between independent variables, so long as the data are missing at random.

CART relies on recursive binary partitioning of a dataset based on several independent variables, or inputs, as they are called in machine learning. During the process, the dataset is successively split into smaller and smaller subsets. The purpose is to produce a set of decision rules applied to the inputs that can divide the data with respect to the classification variable (i.e., separate those who display pGON from those who do not). If the inputs contain data that were collected at an earlier time point than the classification variable, then a truly predictive model is produced. In an early example of this technique in the literature, vital signs and laboratory results at hospital admittance were used to predict the mortality risk of patients with acutely decompensated heart failure.⁸ Forms of CART analysis have already been applied to ocular findings, including data from patients with glaucoma. For example, CART has been applied to results from confocal scanning laser ophthalmoscopy to assist in classifying patients as normal or glaucomatous.^{9–13}

Conceptually, the process can be thought of as growing a tree where the trunk (entire cohort) is repeatedly split into two branches. Each new case (eye) follows a path along the decision tree, with the direction taken at each branch determined by questions applied to the inputs. An optimal set of inputs, with appropriate cutoffs applied to them, is selected to minimize the number of misclassified cases. The operator has the option of weighting the importance of different misclassification events, for example placing more importance on misclassifying progressing cases versus misclassifying nonprogressing cases, which has the effect of pushing the decision tree toward higher sensitivity or higher specificity.

From the ¹Devers Eye Institute, Legacy Health System, Portland, Oregon; the ²Department of Mathematics and Statistics, San Diego State University, San Diego, California; and the ³Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, Iowa.

Supported by National Institutes of Health Grant EY 03424 (CAJ). Submitted for publication January 20, 2008; revised April 23 and July 13, 2008; accepted December 3, 2008.

Disclosure: **S. Demirel**, None; **B. Fortune**, None; **J. Fan**, None; **R.A. Levine**, None; **R. Torres**, None; **H. Nguyen**, None; **S.L. Mansberger**, None; **S.K. Gardiner**, None; **G.A. Cioffi**, None; **C.A. Johnson**, None

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Shaban Demirel, Devers Eye Institute, Legacy Health System, Legacy Clinical Research and Technology Center, 1225 NE 2nd Avenue, Portland, OR, 97232; sdemirel@deverseye.org.

The purpose of the present study is to test the hypothesis that specific locations and patterns of threshold findings within the visual field have predictive value for progressive glaucomatous optic neuropathy pGON. This approach may allow useful information to be extracted that is normally ignored.

METHODS

The protocols and procedures used in this study complied with the tenets of the Declaration of Helsinki. All protocols were approved by the Legacy Health System Institutional Review Board. All subjects confirmed their willingness to participate in this longitudinal study after the risks and benefits of participation were explained.

Subjects with High-Risk OH or EG

Data from both eyes of 168 individuals with high-risk OH or EG were available for this analysis. Most participants were recruited from the Devers Eye Institute glaucoma clinic, whereas community eye care providers referred the remainder to the study. Inclusion criteria for the study are described elsewhere.¹⁴ Briefly, they include (1) a previous diagnosis of glaucomatous optic neuropathy (GON) or suspicious optic nerve head (ONH) appearance (vertical cup-to-disc ratio ≥ 0.6 , cup-to-disc ratio asymmetry between eyes > 0.2 with no disc size asymmetry, potential neuroretinal rim notching and narrowing, and disc hemorrhage), and/or (2) OH defined as untreated intraocular pressure (IOP) ≥ 22 mm Hg with at least one additional risk factor (family history of glaucoma, history of migraine or Raynaud's syndrome, African-American race, age ≥ 70 years, history of systemic hypertension, or diet-controlled diabetes). Exclusion criteria consisted of any other serious ocular disease, previous ocular surgery (except uncomplicated cataract surgery), visual acuity $< 20/40$ in either eye, spectacle refraction $> \pm 5.00$ D sphere and $> \pm 2.00$ D cylinder, any media opacity greater than mild age-related lens changes, diabetes requiring medication, or full-threshold 24-2 SAP MD worse than -6 dB before recruitment. Other than ensuring subjects had no worse than mild perimetric defects, visual fields played no other part in study recruitment. All subjects had performed several automated visual field examinations before this study's baseline examination, so learning effects should have been minimal. Only subjects with reliable visual field results (false positives and negatives < 0.33) were included in the analysis.

Yearly visits were scheduled at which participants were examined with SAP (HFA II; Carl Zeiss Meditec, Dublin, CA), tonometry (Goldmann applanation) and simultaneous stereo nerve head photography (3-Dx; Nidek Co., Ltd., Gamagori, Japan) after maximum pupil dilation. Central corneal thickness (CCT) was measured once during the follow-up period using an ultrasonic pachymeter (DGH Technology, Exton, PA), but this was not at the baseline visit for most subjects. Only baseline findings, along with the single measure of CCT, were used as CART inputs, but the maximum follow up interval was used to establish pGON for each individual. Subjects were being treated at the discretion of their managing eye care specialists who were sent a copy of study-related test results yearly. Findings from the OH/EG subjects are given in Table 1.

We attempted to quantify the OH/EG cohort by estimating the number of participants that could be considered to have glaucoma using the following set of criteria. (1) A subject is considered to have glaucoma if either eye is considered glaucomatous. (2) An eye is considered glaucomatous if either the ONH is glaucomatous or the visual field is glaucomatous. (3) The ONH is considered glaucomatous if pGON is observed (as also used by e.g., Medeiros et al.¹⁵) or if the ONH is considered GON at both baseline and follow-up, regardless of whether progression has taken place. (4) A visual field is considered glaucomatous if the GHT is outside normal limits (ONL), the PSD is abnormal at the $P < 0.05$ level, or a cluster of at least three abnormal locations exist ($P < 0.05$) that is confined to a hemifield and contains at least one location at the $P < 0.01$ level (i.e., a Hodapp-Parish-Anderson mild defect¹⁶). According to this classification scheme, 88%

TABLE 1. Findings for the OH/EG and Normal Subjects

Variable	OH/EG	Normals
Sex, female/male	96/72	63/37
Age, y (mean \pm SD)	58.2 \pm 11.1	48.8 \pm 13.6
IOP, mm Hg (mean \pm SD)*	19.7 \pm 3.9	14.6 \pm 2.7
CCT, μ m (mean \pm SD)†	557 \pm 38	NA
MD, dB (mean \pm SD)	-0.18 \pm 2.3	+0.85 \pm 1.1
PSD, dB (mean \pm SD)	2.51 \pm 1.7	1.55 \pm 0.4
PSD ONL, (% eyes)	16.7	0.5
GHT ONL, (% eyes)	27.4	2.5
Cluster, (% eyes)‡	23.5	1.0
VF ONL, (% eyes)	34.8	3.0
bGON, (% eyes)	57.1	0.0
pGON, (% eyes)	27.7	NA
Glaucoma, (% subjects)	87.5	6.0

ONL, outside normal limits; bGON, baseline glaucomatous optic neuropathy; pGON, progressive glaucomatous optic neuropathy.

* Measurement missing for both eyes of three OH/EG individuals and three normals.

† Measure missing for both eyes of 20 OH/EG individuals and for all normals.

‡ Clusters were confined to a single hemifield and consisted of at least three locations at or below $P < 0.05$, one of which was at or below $P < 0.01$ level.

of subjects were considered to have glaucoma (Table 1). It should be noted that of all the variables listed in Table 1, only age, IOP, and CCT were used to generate the tree models.

Normal Subjects

Data from both eyes of 100 normal subjects were available to test the specificity of the CART models. These subjects were employees of Legacy Health System, their families, and the friends and spouses of the OH/EG subjects. Normal subjects were required to be within normal limits in all findings of a comprehensive eye examination that included visual acuity ($\geq 20/40$), slit lamp biomicroscopy, IOP (< 21 mm Hg), and dilated fundus examination. When the eye examination result suggested that subjects had normal eyes, they were included unless their visual field result was unreliable or suggestive of disease. Consequently, a small number of normal eyes had visual field results that were outside classically defined normal limits (i.e., $P < 0.05$ on PSD) as shown in Table 1. Apart from CCT, all information that was available from the baseline examination of the OH/EG subjects was also available for the normal subjects.

Determination of pGON

Either the baseline nerve head photograph or the most recent follow-up photograph was randomly labeled slide A; the other photograph was labeled slide B. Masked to all other subject information, two fellowship-trained glaucoma specialists (HN and RT) independently graded each stereo pair (Stereo Viewer II; Asahi-Pentax, Tokyo, Japan) as either normal or GON, based on the following characteristics: adequate clarity and stereopsis, neuroretinal rim thinning (generalized or localized), excavation, retinal nerve fiber layer defect, violation of the normal pattern of rim thickness (also known as the ISNT rule),¹⁷ and cup-to-disc ratio by contour.¹⁸ The graders then determined whether there had been any change between the two photographs, and if so, which photograph was worse. Graders based their determination of change on decreasing rim thickness (if ≥ 2 clock hours), new neuroretinal rim notch (if ≤ 1 clock hour), increased excavation (undermining of the disc margin), and new or enlarged nerve fiber layer defect(s). Changes in rim color, presence of a new disc hemorrhage or progressive peripapillary atrophy were not sufficient for a determination of change to be made. Furthermore, pGON was deemed to have occurred only if the photograph that was called worse was from the follow-up visit. Initial agreement between the two primary graders was

71%, which is comparable to published agreement rates.^{15,19} Disagreements were initially addressed by asking graders to reach a consensus. If a consensus could not be reached, then one additional masked grader (GAC or SLM) made a final adjudication. The mean interval between baseline and most recent follow-up ONH photograph was 5.5 ± 1.7 (SD) years (range, 2.0–7.9) with a median of 6.1 years.

During the follow-up period, pGON was observed in 67 individuals. For 41 individuals, pGON was observed in one eye, whereas for 26 individuals, pGON was observed in both eyes (see Table 1). It is worth noting that longitudinal follow-up was not performed for the normal subjects, and pGON was assumed to be zero for these individuals, but it was not assessed.

Age Correction of Visual Field Data

We used thresholds from individual visual field locations, along with other clinical variables, to predict pGON. Consequently, it was necessary to correct for the normal decline of perimetry thresholds with age. Initially, all visual fields from left eyes were made right eye equivalent by reflecting about the vertical midline. Slope parameters from linear regressions of threshold on age for a group of 348 normal subjects for all 24-2 visual field locations were obtained from the investigators in a previous study.²⁰ None of the 100 normal subjects used in the present study were part of the group of 348 used to generate the age-correction parameters.

The mean age-related rate of decline for the 53 nonblindspot locations was -0.06 dB/year. These regression slopes were used to adjust all SAP data in the present study to 48.5 years, as that was the mean age of the 348 normal subjects. If age adjustment resulted in a threshold that was less than 0 dB, then the age adjusted threshold was set to 0 dB. The net effect of age adjustment was to generate the best estimate of SAP thresholds, if all participants had been 48.5 years of age, and to effectively remove the influence of age on SAP thresholds. This allowed us to use age in the tree models as an independent predictor of pGON.

We could have used total deviation (TD) values instead of age-adjusted thresholds in this analysis, with similar results. Calculating age-adjusted thresholds was simpler for us, as we are able to digitally extract threshold data from saved Humphrey visual fields. To use TD values would have required hand entry of data from 536 visual fields (two eyes each for 168 OH/EG and 100 normal subjects), which would have been inefficient and error prone. In addition, TD values are integers, whereas our age-adjusted thresholds maintain high precision.

Building CART Models

All analyses were performed in the R language and environment for statistical computing,²¹ in combination with the package rpart,²² which was used for CART analyses. Package randomForest²³ was used to compute the importance of the inputs. In the present analysis, the classification variable was pGON, which was predicted using the inputs listed in Table 2.

In CART analyses, it is customary to initially allow tree growth to continue until it cannot continue any further. It is likely that such an exhaustively grown tree will perform poorly when applied to an independent dataset, because idiosyncrasies and noise in the data are being incorporated into the model (over fitting). It is therefore necessary to prune the large initial tree and produce a family of smaller trees in an attempt to capture only robust effects. It is possible for the user

to limit initial tree growth by handicapping large complex trees in favor of smaller, simpler ones under the assumption that much of the large tree will have to be pruned later. A common method of selecting the optimum tree from the family of pruned trees is to perform 10-fold cross-validation (CV),⁷ which was used in this analysis. During 10-fold CV, the dataset is randomly divided into 10 equal-sized partitions. Each partition is held aside, while a tree is grown with the other 90% of the data. The held-aside cases are then used to validate the performance of the trees constructed. For those familiar with CART, the control parameters used in tree construction are shown in Appendix 1, online at <http://www.iovs.org/cgi/content/full/50/2/674/DC1>.

Unlike some other statistical methods,^{24,25} the available implementations of CART cannot account for the correlated nature of data from the two eyes of an individual,²⁶ and only data from one eye per subject can be used in construction of a tree model. Consequently, analysis was performed on the worse eye of each subject. For OH/EG participants who only had one eye exhibit pGON (41/168), the progressing eye was considered the worse eye. For OH/EG participants who had both eyes (26/168) or neither eye (101/168) exhibit pGON (127/168 total), one eye was randomly chosen to be the worse eye. Using the worse eye allowed us to maximize the number of pGON cases available for tree construction. In an effort to explore the effect of this selection process the random choice of worse eye was repeated 10 times, and a tree model was generated for each of the 10 samples.

Testing Specificity of CART Models

One eye was randomly chosen for each of the normal subjects, and this selection process was repeated 10 times. Each one of the 10 tree models was tested using a different random eye selection from the normal subjects. The decision rules generated from the OH/EG subjects were applied to the data from the normal subjects and a prediction (stable or pGON) was made.

RESULTS

A classification tree generated from one of the 10 worse-eye selection samples is shown in Figure 1. This tree is displayed because its performance was near average for all 10 trees. The nine additional trees from the other worse eye selection samples are shown in Appendix 2, online at <http://www.iovs.org/cgi/content/full/50/2/674/DC1>. The decision tree is entered at the top, and the first question is evaluated to produce a yes or no answer. If the question evaluates to yes (true), the left branch is followed; otherwise, the right branch is followed. Successive questions are evaluated for each case until it arrives at one of the terminal regions (called nodes) where no further splitting takes place. In the figure, terminal nodes are accompanied by the number of stable and pGON eyes that were assigned to that node. The label that appears first (stable or pGON) determines the classification given to all cases within that node. Visual field test locations are designated as TP#, where # is an identification number. To place the split values in perspective, the normal percentile associated with the age-adjusted threshold values shown in the tree are displayed in parentheses to the right of each decision rule. Figure 2 shows the test locations identified within the Humphrey 24-2 pattern for a right eye.

Figure 1 suggests that an eye with a baseline age-adjusted SAP threshold of 29 dB at test point 45 (TP45) would result in the first question evaluating to yes (29 is greater than 28.8) and the case would follow the left branch to arrive at a terminal node. Forty-four eyes ended up in this terminal node within which all eyes were predicted to have a stable ONH appearance. Eight of the 44 (18%) eyes exhibited pGON and were therefore misclassified as stable. We can compare this terminal node to the lower rightmost terminal node. To arrive in the lower rightmost node, an eye was required to have baseline age-adjusted SAP thresholds that met all the following criteria:

TABLE 2. Inputs Used during Tree Construction to Predict pGON

Visual Field	Clinical	Demographic
53 Nonblindspot test point thresholds (age-adjusted dB)	Baseline IOP (mm Hg) CCT (μ m)	Sex (male or female) Age at baseline (y)

Units/factor levels are given in parentheses. Columns show type of predictor.

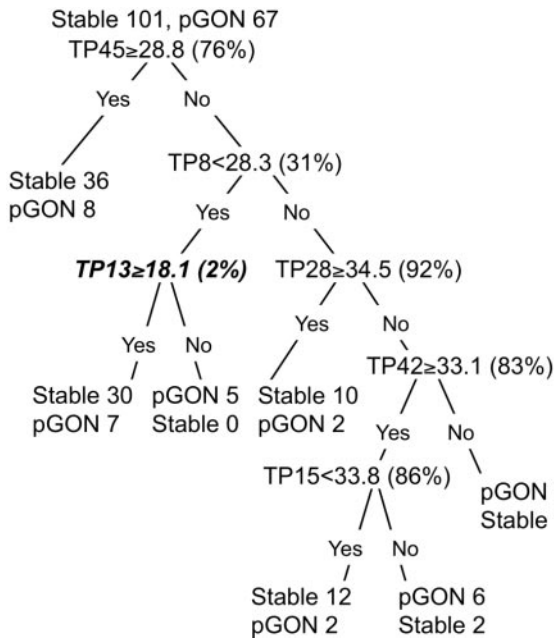


FIGURE 1. The decision tree closest to average performance for all 10 trees generated. Normal percentiles associated with split value are shown in parentheses next to the decision rule. Decisions shown in *bold italic* identify split values that are at or below the normal lower 5th percentile. TP stands for test point, the locations of which are shown in Figure 2.

TP45 < 28.8 dB, TP8 ≥ 28.3 dB, TP28 < 34.5 dB, and TP42 < 33.1 dB. Forty-eight eyes met these criteria and were predicted to exhibit pGON. Thirty-seven (77%) of these eyes exhibited pGON and were correctly classified. However, 11 eyes were misclassified, as they had stable ONH appearance.

The rate of pGON in our dataset was 40% (67/168 eyes for each of the 10 worse-eye random samples). Figure 1 demonstrates that this CART model was able to split the cohort into those at high risk for pGON and those at low risk for pGON, by using only the age-adjusted baseline SAP thresholds at six locations. This ability can be observed by examining the lower right most terminal node where the predicted rate of pGON is 1.9 times the average rate and comparing that to the upper leftmost terminal node where the predicted rate of pGON is

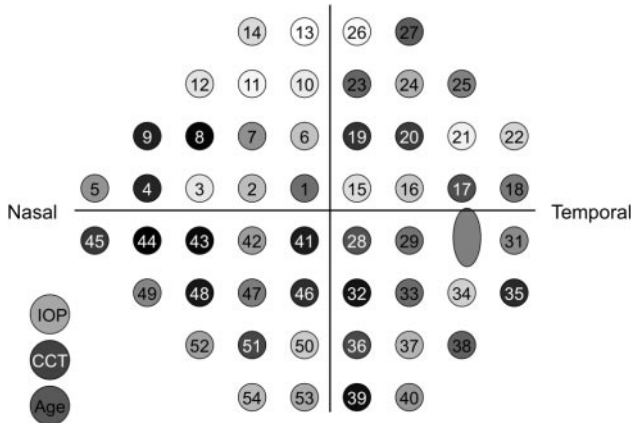


FIGURE 2. Most important locations arranged as a 24-2 test pattern for a right eye with baseline IOP, CCT, and baseline age included (*left*). The darker the *gray shading* used to display a test location or variable the more important it is in the tree models.

0.46 times the average rate. This result represents a 4.2 times difference in the predicted rate of pGON.

Visual field locations within Figure 2, along with IOP, CCT, and baseline age, have been shaded according to the ranked variable importance generated from randomForest using the 10 random samples of worse eye. The variable with the greatest importance is black (TP44), and shading has been made progressively lighter (equal gray steps) with decreasing rank of variable importance, with the least important variable (TP13) being white.

In an attempt to quantify the ability of the 10 trees to separate stable from pGON eyes, we calculated the discriminability index (d'), borrowed from signal-detection theory.²⁷ The discriminability index is based on the true- and false-positive rates. The average d' for the 10 trees was 1.64 (95% confidence interval [CI] 1.45-1.83; range 1.08-1.97), and this value was used to plot the solid curve in Figure 3.

The short-dash curves on either side of the solid curve represent the 95% CI of the average d' value. A decision process, in this case trying to predict whether an eye will exhibit pGON or not, is often defined as being at its threshold if $d' = 1$. A line that depicts $d' = 1$ (dash-dot curve) is shown in Figure 3 along with the chance-performance line (long-dash curve), which corresponds to $d' = 0$.

We also examined the ability of the baseline summary indices MD, PSD, and GHT (borderline grouped with ONL) and baseline GON (bGON) to predict pGON using univariate logistic regression. Data from both eyes of each participant were used with results adjusted for the correlated nature of findings from the two eyes (generalized estimating equations [GEE] with logit link). Neither baseline MD nor PSD was significantly related to pGON ($P > 0.05$ in both cases), suggesting that predictive information for pGON is lost when summary indices that are based on all visual field locations are calculated. Baseline GHT was significantly related to pGON (GEE: Wald = 7.4, $P = 0.006$) with greater risk for pGON if baseline GHT was borderline or ONL. Having bGON was also highly predictive of pGON (GEE: Wald = 19.7, $P < 0.0001$).

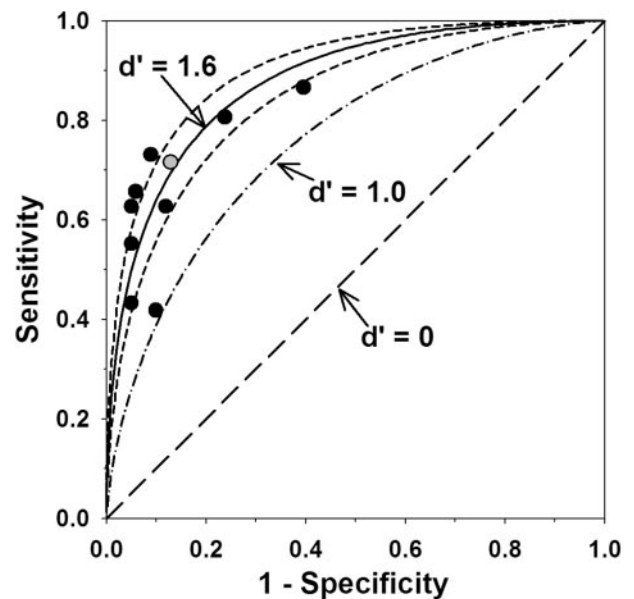


FIGURE 3. Sensitivity (true-positive rate) plotted against 1 - specificity (false-positive rate) for the 10 trees generated (symbols). *Gray circle*: tree shown in Figure 1. Also plotted are the average discriminability index (d' , *solid line*), the 95% CI (*short-dash line*), the $d' = 1$ line (*dash-dot line*) and the $d' = 0$ line (*long-dash line*).

If MD, PSD, GHT, and bGON were added to the individual test point thresholds and used as inputs for the tree models, MD and GHT were near bottom in ranked variable importance, PSD was ranked 16th of 62 and bGON was ranked 1st. However, in terms of sensitivity, specificity, and overall misclassification rate, performance was essentially unchanged when these inputs were included. The ranked importance of visual field locations also changed only minimally going from models that excluded to models that included MD, PSD, GHT, and bGON. The correlation between the ranked importances of test location from the two sets of models was 0.93.

When applied to the data from the 100 normal subjects, the average specificity of the 10 tree models was 69%. Thirty-one percent of eyes, on average, were predicted to exhibit pGON.

DISCUSSION

Classification tree models are data-mining tools that enhance depiction of interactions between variables, especially when complex or unexpected. Complex interactions may be difficult or impossible to tease apart using more traditional statistical techniques.⁷ CART allows the structure within the data to drive the model determined, rather than requiring a model be formalized in advance and then tested against the data. Safeguards, such as 10-fold CV, protect against idiosyncrasies in the learning dataset being incorporated into the model and attempt to forecast performance in an independent dataset. What this study shows is that baseline SAP fields contain information that can be used to predict pGON and that certain visual field locations tend to be more important for this task than other locations, at least in this cohort of subjects with OH or EG.

It can be observed that only one of the split values in Figure 1 is below the normal lower 5th percentile level which is traditionally used to define statistical significance. It is also worth pointing out that an eye could reach the lower right terminal node in Figure 1, which contains eyes at high risk for pGON, without a single test location having $P \leq 0.05$ on a traditional TD probability plot. In all 10 trees combined, there were a total of 55 splits made, with only 9 (16%) of these based on an age-adjusted threshold value that is abnormal at the $P \leq 0.05$ level.

In 7 of the 10 tree models generated (Fig. 1 and Appendix 2, <http://www.iovs.org/cgi/content/full/50/2/674/DC1>), the initial decision was based on test point 44 with the criterion value being near the 75th normal percentile. For the remaining trees, the initial split was based on test point 29, 41, or 45. Of note, all four of these test locations (TP29, TP41, TP44, and TP45) lie along the inferior horizontal meridian, with three of them in the nasal step area.

If the ranked importance of test points is evenly divided into low, medium, and high ranges, the inferior visual field appears to have greater importance. Eleven (42%) of 26 locations in the inferior field have high importance with only 4 (15%) of 26 having low importance. By contrast, 6 (22%) of 27 locations in the superior field have high importance with 14 (52%) of 27 having low importance.

Henson and Chauhan²⁸ report that visual field locations in the superior arcuate region and in the inferior nasal quadrant provide the maximum amount of information for diagnosis of glaucoma. They also find that the extreme superior periphery carries little diagnostic information. These statements are in general agreement with Figure 2, in that the inferior nasal quadrant contains many high-importance locations, high-importance locations in the superior field are almost exclusively in the arcuate region and the extreme superior periphery is devoid of high-importance locations. Our findings differ on the

importance of the inferior temporal quadrant and the area adjacent to the physiological blind spot, as Henson and Chauhan suggest that these areas provide the least amount of information, whereas we find that the inferior temporal quadrant contains quite a number of high-importance locations. Only two of eight locations bordering the blind spot have low importance in our analysis.

Heijl and Lundqvist²⁹ examined eyes from OH with or without established glaucoma in the fellow eye. They identified defective locations evident in the first glaucomatous field after repeatedly normal fields. The most common defective locations, especially those with absolute defects, were predominantly in the superior field and near the physiological blind spot, in contrast to the present study, in which the inferior visual field appeared most important but agrees that the area near the blind spot may be important.

It should be remembered that in the present study the importance of visual field locations pertains to prediction of pGON and not making a diagnosis of glaucoma. In that regard, the question being asked is different in this study compared with both studies.^{28,29} Different visual field locations may be most important for diagnosis of glaucoma and for predicting pGON.

It is also critical to recognize that being an important visual field location in this analysis is not equivalent to suggesting that threshold must be depressed at that location. One must resist the temptation to interpret the importance map in the same way that one interprets a visual field printout. High importance suggests only that a location provides information for predicting pGON. Some important locations may be acting as anchors for normalcy, and it is only in combination with a low threshold at another location that predictive information is manifest. For that reason, interpreting important locations in terms of anatomy and physiology of the ganglion cells and RNFL may be questionable.

Making 10 random samples of worse eye allowed us to estimate the influence of the worse eye selection process. Examination of Figure 3 shows that some of the decision trees had high sensitivity but generally at the cost of poorer specificity and vice versa. It appears that the 10 tree models reflect the same underlying decision process as they show similar d' values. The worse-eye selection process resulted in decision trees that were slightly more sensitive or slightly more specific but did not substantially affect their ability to discriminate between eyes likely to have stable versus progressing ONH appearance.

The location importance shown in Figure 2 suggests that it may be possible to test fewer visual field locations while monitoring glaucoma patients for progression. Testing at fewer locations would reduce test duration and patient fatigue and perhaps would improve reliability. Alternatively, with the same test duration as current tests, it may be possible to measure threshold twice at a reduced set of locations and average the two determinations, reducing test-retest variability. Others have examined the possibility of producing optimized sets of test locations³⁰⁻³³ but the concept has not found traction in visual field testing for glaucoma. In particular, Weber and Diestelhorst³⁴ have even examined the utility of reduced sets of test points to detect visual field progression in glaucoma. The purpose of the present study was to predict pGON, so the reduced sets of test locations may be different for the two purposes. We have not examined the usefulness of reduced sets of test locations in this article, and the suggestion must therefore be considered speculative.

This application of CART is limited in five aspects. First, only one eye could be used per subject. We chose a worse eye before performing the analyses. Choosing a worse eye instead of randomly choosing an eye may have allowed a slightly

greater chance of bias. Eye selection was essentially random, however, for 227 of 268 subjects. Although it is not uncommon in ophthalmic statistics to have to randomly select one eye from each individual or use an average from both eyes, this method sacrifices information. An allied point is that subjects who had both eyes display pGON should perhaps have been given greater weight in the tree models. Currently, we have no data to suggest what this greater weighting should be and weighted all cases equally. CART methods are under development that can account for correlation between cases. These methods may make eye choice and weighting moot topics. Second, our classification variable was whether pGON was observed between baseline and the most recent follow-up visit. We have not determined time-to-pGON and therefore do not have survival time information for the classification variable. CART methods that take advantage of survival data have been developed³⁵ and information regarding time-to-pGON may have improved performance of the tree models or altered outcomes. Third, our determination of pGON was predicated on one baseline and one follow-up stereo nerve head photograph. Confirming progression with a second follow-up photograph would have been preferable and not seeking confirmation may have allowed a small number of false pGON determinations to be made, affecting results. Fourth, the number of subjects used in this study (168 OH/EG and 100 normal subjects) is limited, and validation in a larger, independent cohort is needed before these findings can be considered generalizable. Finally, even though the age-correction process used location specific rates of change, this change was assumed to be linear. Other studies suggest the relationship between age and perimetric sensitivity,³⁶ or age and test-retest variability,³⁷ may be nonlinear (but see also Ref. 38). If this relationship is nonlinear then our age-adjusted data would be underadjusted for older individuals and this may have impacted results.

We have attempted to validate the specificity of our decision trees by applying them to data collected from 100 normal subjects. However, this is not really a fair comparison dataset for validating the specificity of the decision trees as they have been trained to predict which OH/EG subjects will display pGON and which will display stable ONH appearance. The ideal dataset for validating the tree models would come from OH/EG subjects that have displayed longitudinally stable ONH appearance. The average finding of 69% specificity when the data from normal subjects was run through the tree models is a little troubling, though. We had expected the decision trees to have better specificity when data from normal subjects was used, but it is difficult to know exactly what features in the data are being exploited by the tree models to allow prediction of pGON. The lower than expected specificity of the tree models when data from normal subjects was used is a further argument for validation of these tree models in larger, independent datasets before they can be considered generalizable.

In summary, the current analyses used decision trees to allow prediction of pGON from baseline SAP examination coupled with CCT, baseline IOP and baseline age. The decision tree with average performance in this study had sensitivity and specificity of 65% and 87%, respectively. When visual field locations are ranked in terms of importance for predicting pGON, the inferior visual field seems more important for this task, particularly along the nasal horizontal meridian. Subtle visual field features—for example, being in the normal lower quartile at certain visual field locations while being in the normal upper quartile at other locations—conveyed information that was useful for predicting which eyes would exhibit pGON. In only a few instances did the decision process rely on a threshold value that would be considered abnormal in a more traditional statistical sense (i.e., $P < 0.05$). Using information

regarding the exact percentile associated with threshold values and not just whether they are below the normal lower 5th percentile, may assist in assessing the functional status of glaucoma patients and their risk for progressive change at the ONH.

Acknowledgments

The authors thank Cindy Blachly, Thie Smith, and Judith Thompson for assistance in data collection and ensuring ongoing subject participation.

References

- Gordon MO, Beiser JA, Brandt JD, et al. The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma. *Arch Ophthalmol*. 2002;120(6):714–720.
- Heijl A, Leske MC, Bengtsson B, Hyman L, Bengtsson B, Hussein M. Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial. *Arch Ophthalmol*. 2002;120(10):1268–1279.
- Åsman P, Heijl A. Glaucoma hemifield test: automated visual field evaluation. *Arch Ophthalmol*. 1992;110:812–819.
- The Advanced Glaucoma Intervention Study Group. The Advanced Glaucoma Intervention Study (AGIS). 2. Visual field test scoring and reliability. *Ophthalmology*. 1994;101(8):1445–1455.
- Musch DC, Lichter PR, Guire KE, Standardi CL. The Collaborative Initial Glaucoma Treatment Study: study design, methods, and baseline characteristics of enrolled patients. *Ophthalmology*. 1999;106(4):653–662.
- Gillespie BW, Musch DC, Guire KE, et al. The collaborative initial glaucoma treatment study: baseline visual field and test-retest variability. *Invest Ophthalmol Vis Sci*. 2003;44(6):2613–2620.
- Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Pacific Grove, CA: Wadsworth; 1984.
- Fonarow GC, Adams KF Jr, Abraham WT, Yancy CW, Boscardin WJ. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. *JAMA*. 2005;293(5):572–580.
- Mardin CY, Hothorn T, Peters A, Junemann AG, Nguyen NX, Lausen B. New glaucoma classification method based on standard Heidelberg Retina Tomograph parameters by bagging classification trees. *J Glaucoma*. 2003;12(4):340–346.
- Hothorn T, Lausen B. Bagging tree classifiers for laser scanning images: a data- and simulation-based strategy. *Artif Intell Med*. 2003;27(1):65–79.
- Adler W, Hothorn T, Lausen B. Simulation based analysis of automated classification of medical images. *Methods Inf Med*. 2004; 43(2):150–155.
- Manassakorn A, Nouri-Mahdavi K, Caprioli J. Comparison of retinal nerve fiber layer thickness and optic disk algorithms with optical coherence tomography to detect glaucoma. *Am J Ophthalmol*. 2006;141(1):105–115.
- Naithani P, Sihota R, Sony P, et al. Evaluation of optical coherence tomography and Heidelberg retinal tomography parameters in detecting early and moderate glaucoma. *Invest Ophthalmol Vis Sci*. 2007;48(7):3138–3145.
- Spry PG, Johnson CA, Mansberger SL, Gioffi GA. Psychophysical investigation of ganglion cell loss in early glaucoma. *J Glaucoma*. 2005;14(1):11–19.
- Medeiros FA, Zangwill LM, Bowd C, Sample PA, Weinreb RN. Use of progressive glaucomatous optic disk change as the reference standard for evaluation of diagnostic tests in glaucoma. *Am J Ophthalmol*. 2005;139(6):1010–1018.
- Hodapp E, Parrish RK, Anderson DR. *Clinical Decisions in Glaucoma*. St. Louis: CV Mosby; 1993:204.
- Jonas JB, Budde WM, Panda-Jonas S. Ophthalmoscopic evaluation of the optic nerve head. *Surv Ophthalmol*. 1999;43(4):293–320.
- Zangwill L, Shakiba S, Caprioli J, Weinreb RN. Agreement between clinicians and a confocal scanning laser ophthalmoscope in estimating cup/disk ratios. *Am J Ophthalmol*. 1995;119:415–421.
- Coleman AL, Sommer A, Enger C, Knopf HL, Stamper RL, Minckler DS. Interobserver and intraobserver variability in the detection of

- glaucomatous progression of the optic disc. *J Glaucoma*. 1996;5(6):384-389.
20. Johnson CA, Sample PA, Cioffi GA, Liebmann JR, Weinreb RN. Structure and function evaluation (SAFE): I. criteria for glaucomatous visual field loss using standard automated perimetry (SAP) and short wavelength automated perimetry (SWAP). *Am J Ophthalmol*. 2002;134(2):177-185.
 21. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Ver. 2.6.1. Vienna, Austria: R Foundation for Statistical Computing; 2007. Available at <http://www.R-project.org>. Accessed December 28, 2007.
 22. Therneau TM, Atkinson B. *rpart: Recursive Partitioning*. Ver. 3.1-38. Porting to R by Brian Ripley. Rochester, MN: Mayo Clinic; 2007. Available at <http://mayoresearch.mayo.edu/mayo/research/biostat/splusfunctions.cfm>. Accessed December 28, 2007.
 23. Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002;2(3):18-22.
 24. Katz J, Zeger S, Liang KY. Appropriate statistical methods to account for similarities in binary outcomes between fellow eyes. *Invest Ophthalmol Vis Sci*. 1994;35(5):2461-2465.
 25. Levine RA, Demirel S, Fan J, Keltner JL, Johnson CA, Kass MA. Asymmetries and visual field summaries as predictors of glaucoma in the ocular hypertension treatment study. *Invest Ophthalmol Vis Sci*. 2006;47(9):3896-3903.
 26. Katz J. Two eyes or one? The data analyst's dilemma. *Ophthalmic Surg*. 1988;19(8):585-589.
 27. Green DM, Swets JA. *Signal Detection Theory and Psychophysics*. New York: Wiley; 1966.
 28. Henson DB, Chauhan BC. Informational content of visual field location in glaucoma. *Doc Ophthalmol*. 1985;59(4):341-352.
 29. Heijl A, Lundqvist L. The frequency distribution of earliest glaucomatous visual field defects documented by automatic perimetry. *Acta Ophthalmol Scand*. 1984;62(4):658-664.
 30. Henson DB, Chauhan BC, Hobbey A. Screening for glaucomatous visual field defects: the relationship between sensitivity, specificity and the number of test locations. *Ophthalmic Physiol Opt*. 1988;8(2):123-127.
 31. Krakau CET. Visual field testing with reduced sets of test points a computerized analysis. *Doc Ophthalmol*. 1989;73:71-80.
 32. Zeyen TG, Zulauf M, Caprioli J. Priority of test locations for automated perimetry in glaucoma. *Ophthalmology*. 1993;100:518-523.
 33. Chauhan BC, Johnson CA. Evaluating and optimizing test strategies in automated perimetry. *J Glaucoma*. 1994;3(suppl. 1):S73-S81.
 34. Weber J, Diestelhorst M. Perimetric follow-up in glaucoma with a reduced set of test points. *Ger J Ophthalmol*. 1992;1(6):409-414.
 35. LeBlanc M, Crowley J. Relative risk trees for censored survival data. *Biometrics*. 1992;48(2):411-425.
 36. Spry PG, Johnson CA. Senescent changes of the normal visual field: an age-old problem. *Optom Vis Sci*. 2001;78(6):436-441.
 37. Katz J, Sommer A. A longitudinal study of the age-adjusted variability of automated visual fields. *Arch Ophthalmol*. 1987;105:1083-1086.
 38. Heijl A, Lindgren G, Olsson J. Perimetric threshold variability and age. *Arch Ophthalmol*. 1988;106(4):450-452.