Article

A Statistical Model to Analyze Clinician Expert Consensus on Glaucoma Progression using Spatially Correlated Visual Field Data

Joshua L. Warren[1], Jean-Claude Mwanza[2], Angelo P. Tanna[3], and Donald L. Budenz[2]

[1] Department of Biostatistics, Yale University, New Haven, Connecticut, USA
[2] Department of Ophthalmology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[3] Department of Ophthalmology, Northwestern University, Chicago, Illinois, USA

Correspondence: Joshua L. Warren, Department of Biostatistics, Yale University, New Haven, CT, USA. e-mail: joshua.warren@yale.edu

Purpose: We developed a statistical model to improve the detection of glaucomatous visual field (VF) progression as defined by the consensus of expert clinicians.

Methods: We developed new methodology in the Bayesian setting to properly model the progression status of a patient (as determined by a group of expert clinicians) as a function of changes in spatially correlated sensitivities at each VF location jointly. We used a spatial probit regression model that jointly incorporates all highly correlated VF changes in a single framework while accounting for structural similarities between neighboring VF regions.

Results: Our method had improved model fit and predictive ability compared to competing models as indicated by the deviance information criterion (198.15 vs. 201.29–213.38), a posterior predictive model selection metric (130.08 vs. 142.08–155.59), area under the receiver operating characteristic curve (0.80 vs. 0.59–0.72; all P values < 0.018), and optimal sensitivity (0.92 vs. 0.28–0.82). Simulation study results suggest that estimation (reduction of mean squared errors) and inference (correct coverage of 95% credible intervals) for the model parameters are improved when spatial modeling is incorporated.

Conclusions: We developed a statistical model for the detection of VF progression defined by clinician expert consensus that accounts for spatially correlated changes in visual sensitivity over time, and showed that it outperformed competing models in a number of areas.

Translational Relevance: This model may easily be incorporated into routine clinical practice and be useful for detecting glaucomatous VF progression defined by clinician expert consensus.

Introduction

Glaucomatous vision loss is irreversible, making early detection of the disease and, once detected, prevention of its progression essential. The timely detection of visual field (VF) progression is important since reduction of intraocular pressure (IOP), achieved with medical therapy, laser surgery, and/or incisional surgery, slows the rate of further deterioration.

Once a patient is diagnosed with glaucoma, clinicians must balance the risks and expenses of advancing levels of medical and surgical intervention with the risk of vision loss due to disease progression. The clinician must determine, at regular intervals, whether the disease is stable or not. If the evidence for progression is sufficiently strong and the apparent rate of progression is such that the patient is at risk of symptomatic vision loss in his or her lifetime, the physician must more aggressively lower the IOP. During most stages of the disease, assessment of the VF is the most important factor that must be considered in determining whether the disease process is stable in an individual patient. As a result, much effort has been devoted to determining if damage to an individual's VF is progressing over time at a rate requiring treatment adjustment.

The majority of past studies either model each VF location separately over time or aggregate over the

translational vision science & technology

Table 1. Data Summaries from the Modeling and Validation Datasets

Dataset	Patients (Eyes)	% Progressing (n)	Number of Tests		Follow-Up, y	
			Mean	Min/Max	Mean	Min/Max
Modeling	97 (191)	26.18 (50)	7.42	(2, 21)	2.55	(0.23, 9.38)
Validation	83 (100)	60.00 (60)	5.00	(5, 5)	4.34	(0.04, 9.67)

entire VF at each time point when determining progression status.[1–3] Both methods are potentially inefficient and neglect important spatial information contained in the data. We hypothesized that jointly considering changes in sensitivities at each VF location over time when determining progression and identifying regions across the VF where clinicians place increased emphasis when judging progression would preserve important spatial relationships, increasing the explanatory and predictive capabilities of the model. Such an approach increases the amount of available information used in determining progression and more closely resembles the decision-making process used by clinicians who focus on changes across the entire VF and patterns of change as opposed to individual locations separately. The purpose of this study was to develop, implement, and validate a statistical model for detecting glaucoma progression, as defined by the consensus of expert clinicians, and compare its performance to existing approaches.

Our newly developed method for jointly incorporating highly spatially correlated predictors in the modeling framework provides insight into how expert clinicians determine the glaucoma progression status based on observing VF measurements over time for an individual and provides a framework for improved prediction of the outcome for future patients.

Methods

Data Sources

Two VF datasets consisting of reliable VFs (fixation loss rate less than 20% and false-positive and false-negative response rates less than 15%) obtained with the 24-2 Swedish Interactive Thresholding Algorithm (SITA) of the Humphrey Visual Field Analyzer (Carl Zeiss Meditec, Inc., Dublin, CA) were used in this study.[4,5] In the first dataset, each of the VFs were obtained using the SITA standard strategy and VF progression status was determined based on the independent reviews of two expert clinicians. In the case of disagreement, a third clinician was consulted. Reviewers had access to an event-based determination of progression, the Glaucoma Progression Analysis output (GPA; Carl Zeiss Meditec, Inc.), and the print versions of the VF tests over time. Clinicians were instructed to define each series of VF tests as definitely progressing or not.[4] Table 1 displays the summary information from this modeling dataset.

The second dataset contained approximately 1% of VF tests performed with the full threshold strategy and approximately 1% performed with the 30-2 test pattern. It represents VF series and responses from an independent validation group of glaucoma patients. These data have been described previously.[5] Progression status was once again determined based on review of the VF series and GPA output for each patient, this time by a different set of five expert clinicians. The intraobserver and interobserver agreements were good to excellent and moderate, respectively.[5] Table 1 displays the summary information from this validation dataset.

Statistical Model

In our model, the probability that an eye is determined to be progressing based on the clinician expert consensus is a function of the rates of change in sensitivities across time at each VF location, where the contribution of a particular location is related to the contributions at nearby spatially-related locations through the use of a spatially referenced prior distribution. We model the underlying probability associated with the binary clinician expert consensus on glaucoma progression status (yes/no) for a patient/eye as a function of VF sensitivity changes over time at each location on the VF, such that $Y_i \mid \boldsymbol{\beta}, \boldsymbol{\theta} \sim$ Bernoulli$\{p_i(\boldsymbol{\beta}, \boldsymbol{\theta})\}$, $i = 1, \ldots, n$ and

$$\Phi^{-1}\{p_i(\boldsymbol{\beta}, \boldsymbol{\theta})\} = \beta_0 + \beta_1 d_i + \sum_{j=1}^{m} \theta(\boldsymbol{s}_j) z_i(\boldsymbol{s}_j) \qquad (1)$$

where Y_i is the progression status of patient/eye i, n is the total number of unique patient/eye combinations included in the study (see Table 1), $p_i(\boldsymbol{\beta}, \boldsymbol{\theta})$ is the probability that patient/eye i is diagnosed as progressing, $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution
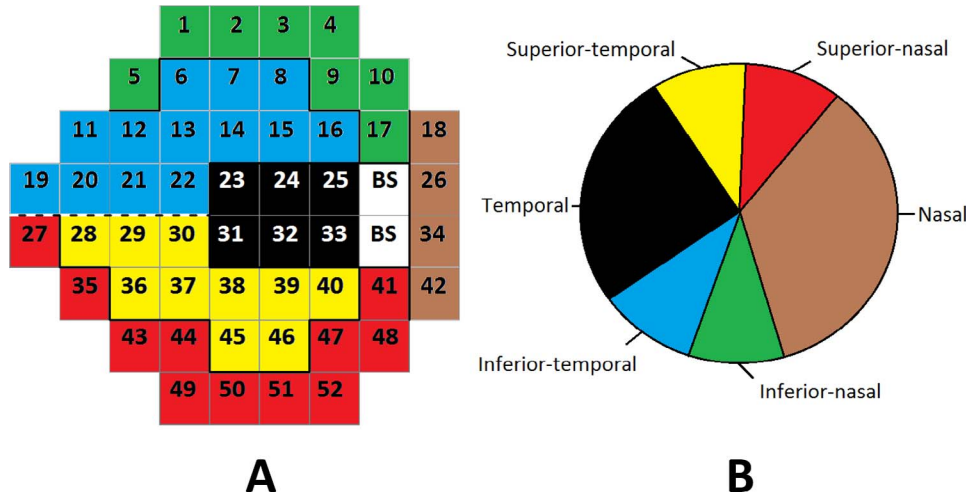
Figure 1. Visual field regions (**A**) mapped to the optic disc (**B**). The numbers represent the spatial locations associated with the $\eta(s_j)$ parameters. BS, blind spot.

function of the standard normal distribution, d_i is the length of time in years that patient/eye i has been followed in the study, $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, β_1 is the parameter describing the association between follow-up time and the probability of being diagnosed as progressing by the clinician expert consensus, β_0 is the intercept parameter, $\boldsymbol{\theta} = \{\theta(s_1), \ldots, \theta(s_m)\}^T$, and $m = 52$ is the number of VF locations (after removing the two locations within the physiologic blind spot; see Fig. 1). We note that progression often occurs at different rates in different eyes of the same individual and that the clinicians assessed progression for each eye individually. Therefore, we chose to model each eye independently.

The $z_i(s_j)$ term is defined as the estimated slope from a simple linear regression analysis of VF sensitivities at VF location s_j for person/eye i. Increasingly negative values of this metric indicate deterioration in vision at a given location while values near zero indicate little change over time. The sum, $\sum_{j=1}^{m} \theta(s_j) z_i(s_j)$, represents the total impact of changes in the VF over time at each of the individual VF locations on the probability of patient/eye i being diagnosed as progressing by the clinician expert consensus.

The main parameters of interest in the study are represented by $\theta(s_j)$, $j = 1, \ldots, m$. These spatially varying regression parameters describe the association between a change in VF sensitivities over time at each location and the probability of interest. Including a separate parameter for each VF location allows for increased modeling flexibility and for the possibility that deterioration in the VF at different locations is

more (or less) informative with respect to being diagnosed as progressing by the clinician expert consensus. A priori we assume a constant association between sensitivity changes at each location and the probability of interest along with location-specific deviations such that $\theta(s_j) = \theta_0 + \eta(s_j)$, where θ_0 represents the constant mean of the spatial process and $\eta(s_j)$ is the slope deviation specific to VF location s_j. The constant prior mean reflects our initial beliefs that changes in sensitivities over time at any location should have a similar impact on progression diagnosis by the clinician expert consensus, with the possibility that certain locations are more (or less) meaningful to clinicians being represented by the spatial deviations. The $\eta(s_j)$ parameters allow for location-specific slopes for each VF location, potentially leading to a model that better describes the process used by clinicians to diagnose glaucoma progression. We allow these parameters to be spatially correlated as detailed in the Prior Specification Subsection so that estimation of their values will depend on the parameter values in surrounding VF locations if necessary. Equation 3 describes how these parameters are related to each other a priori in more detail. Therefore, the probit regression model in Equation 1 can be rewritten as:

$$
\begin{aligned}
\Phi^{-1}\{p_i(\boldsymbol{\beta}, \theta_0, \boldsymbol{\eta})\} &= \beta_0 + \beta_1 d_i + \sum_{j=1}^{m}\{\theta_0 + \eta(s_j)\}z_i(s_j) \\
&= \beta_0 + \beta_1 d_i + \theta_0 \sum_{j=1}^{m} z_i(s_j) \\
&\quad + \sum_{j=1}^{m} \eta(s_j) z_i(s_j),
\end{aligned}
$$

(2)

where $\sum_{j=1}^{m} z_i(s_j)$ represents a measure of the total level of deterioration across the entire VF over time

and θ_0 describes the relationship between this measure and the probability of interests.

Prior Specification

We assigned prior distributions to the introduced model parameters to complete the model specification. A spatially referenced prior distribution was required for the introduced location specific deviations, $\eta(s_j)$. Providing spatial structure in this context helped to control the effect of multicollinearity which often can inflate standard errors and change the sign of parameter estimates. We included a separate predictor at each of the 52 VF locations for a patient/eye with the understanding that sensitivities in similar VF locations are likely positively correlated. Conditional on the spatial proximity matrix, W, which describes the spatial relationship between each of the 52 VF locations, the $\eta(s_j)$ parameters were jointly assigned a prior distribution that allowed neighboring parameters to share information spatially, if appropriate. We specified an intrinsic conditional autoregressive (ICAR) prior model such that $\eta|\sigma_\eta^2 \sim \text{ICAR}(\sigma_\eta^2)$ and

$$\eta(s_i)|\eta(-s_i) \sim \text{N}\left(\sum_{j=1}^{m} \frac{w_{ij}}{w_{i+}}\eta(s_j), \frac{\sigma_\eta^2}{w_{i+}}\right) \qquad (3)$$

where $\eta = \{\eta(s_1),\ldots,\eta(s_m)\}^T$, $\eta(-s_i) = \{\eta(s_1),\ldots,\eta(s_{i-1}),\eta(s_{i+1}),\ldots,\eta(s_m)\}^T$, σ_η^2 is the variance parameter for the spatial process, w_{ij} is the row i, column j entry of W describing the proximity between VF locations s_i and s_j, and w_{i+} represents the row i sum of W.[6]

For spatial data on a grid, the neighborhood definition for W is typically specified as $w_{ij} = 1$ if s_i and s_j share a common border and $w_{ij} = 0$ otherwise. Locations are not considered to be neighbors of themselves, resulting in $w_{ii} = 0$ for all i. However, the spatial structure across the VF is more complex due to each location's structural relationship with the optic disc. We, therefore, used a similar proximity matrix to that previously introduced.[4] In Figure 1, we display the optic disc regions (Fig. 1B) and their spatial association with each location of the VF (Fig. 1A).[7] Based on this structural mapping, we defined $W(\delta)$ as:

$$w_{ij}(\delta) =$$
$$\begin{cases} 1, & \text{if } s_i \text{ and } s_j \text{ share a common border and are in the same optic disc region;} \\ \delta, & \text{if } s_i \text{ and } s_j \text{ share a common border and are in adjacent optic disc regions;} \\ 0, & \text{otherwise} \end{cases}$$

with $w_{ij}(\delta) = 0$ for all i. The δ parameter describes the strength of spatial association between neighboring VF locations in adjacent optic disc regions. A number of past studies also have investigated the use of a partially unknown proximity matrix while working in different settings.[8–10]

We selected weakly informative prior distributions for the hyperparameters to allow the data to drive the inference rather than our prior beliefs. The three nonspatial regression parameters were given independent, vague yet proper prior distributions, such that $\beta_0, \beta_1, \theta_0 \sim \text{N}(0, 10^{10})$. The variance parameter for the ICAR(σ_η^2) process was given a vague yet proper uniform prior distribution, such that $\sigma_\eta^2 \sim \text{Unif}(10^{-100}, 10^4)$. The neighborhood proximity matrix parameter was given a uniform prior distribution such that $\delta \sim \text{Unif}(0,1)$.

Computing

Posterior inference is based on 990,000 samples from the posterior distribution of interest after discarding the first 10,000 draws as a burn-in period. The number of samples to obtain was determined based on the calculation of batch means Monte Carlo (MC) standard errors.[11] Summaries of the MC standard errors for presented posterior mean estimates are displayed in Table 2 and Figure 2. In Supplementary Figure S1, we present trace plots for all nonspatial parameters and randomly selected spatial deviations ($\eta(s_i)$). We noted that these plots are representative of the remaining parameters not displayed. All analyses were done using R Statistical Software and the model fitting details are presented in the Supplementary Material.[12]

Model Comparisons

The benefits of the newly developed model, referred to as "Model 1" hereafter, are most apparent when results are compared with competing models. We considered the following models in this study:

- Model 1: The newly developed probit regression model for spatially correlated predictors with spatially varying VF parameters modeled using the ICAR(σ_η^2) prior distribution and partially unknown spatial proximity matrix.
- Model 2: A probit regression model with separate parameters for VF location that are modeled using exchangeable and independent normally distributed prior distributions.
- Model 3: A probit regression model including only a single spatially aggregated predictor without considering location.
- Models P1 to P4: A probit regression model using

Table 2. Posterior Summaries for the Nonspatial Parameters in Model 1

Parameter	Mean	SD	Quantiles		
			0.025	0.50	0.975
Intercept, $\widetilde{\beta}_0$	−1.25	0.22	−1.70	−1.25	−0.82
Follow-up time, in years, β_1	0.17	0.06	0.04	0.17	0.30
Spatial mean, θ_0	−0.52	0.16	−0.84	−0.51	−0.21
ICAR variance parameter/100, $\sigma_\eta^2/100$	1.72	1.52	0.21	1.29	5.69
Proximity matrix parameter, δ	0.68	0.23	0.19	0.72	0.99

The Monte Carlo standard errors for the presented estimated posterior means ranged from 0.00 to 0.05 with a median value of 0.02.

the first (Model P1), second (Model P2), third (Model P3), and fourth (Model P4) smallest P values from simple pointwise linear regressions (PLRs), fit separately at each location on the VF, as a single predictor.

- Models PS1 to Model PS4: A probit regression model using the first (Model PS1), second (Model PS2), third (Model PS3), and fourth (Model PS4) smallest estimated slopes that are associated with a P value less than or equal to 0.01 from simple PLRs, fit separately at each location in the VF, as a single predictor.

All of the introduced models were fit in the Bayesian setting for comparison purposes. Model 1 represents the newly introduced spatial model previously described. Model 2 uses the same probit regression model form as Model 1 (see Equation 1) while ignoring the spatial structure of the individual parameters. Instead, exchangeable normally distributed prior distributions with common variance were assigned to the parameters. We expected this model to perform poorly due to the lack of sharing of information between the introduced parameters, relatively small sample size, and high correlation between the predictors. This model served as a baseline to investigate what happens when all slopes are included in the model without accounting for spatial correlation.

Model 3 represents a simplified version of Model 1, ignoring the spatial information in the data. Model 3 is nested within Model 1 with $\sum_{j=1}^{m}\eta(s_j)z_i(s_j)$ removed from Equation 2. Therefore, Model 3 assumes a constant effect (θ_0) across all VF locations. We expected this model to outperform Model 1 if there was no spatial variability across the VF in parameter estimates. This would suggest that clini-
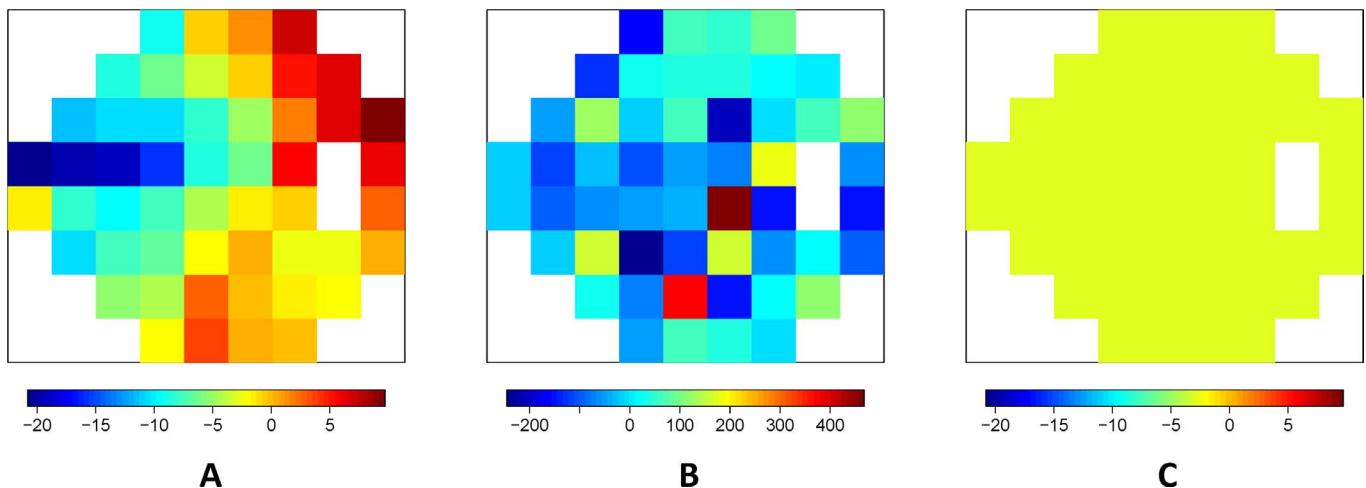
Figure 2. Posterior means from (**A**) Model 1, (**B**) Model 2, and (**C**) Model 3 for the spatial slopes across the visual field. The posterior standard deviations for the presented parameters ranged from 6.98 to 11.34 with a median value of 8.51 for Model 1, 34.52 to 122.50 with a median value of 56.43 for Model 2, and was 0.75 for Model 3. The Monte Carlo standard errors for the presented estimated posterior means ranged from 0.01 to 0.05 with a median value of 0.02 for Model 1, 0.15 to 1.15 with a median value of 0.39 for Model 2, and was 0.001 for Model 3.

cians weigh each VF location equally when determining progression status.

Models P1 to P4 represent common analyses in the progression determination setting.[13] Simple PLR analyses are done at each of the available 52 locations on the VF. From each of these separately run analyses, a frequentist P value is obtained indicating if the regression slope is significantly less than zero. In Model P1, we calculated the minimum P value over the entire VF for a patient and entered it as a predictor in a probit regression model with the progression status, as defined by the consensus of expert clinicians, as the outcome. We then used the second (Model P2), third (Model P3), and fourth (Model P4) smallest P values as predictors separately in their own models and compared results. Models PS1 to PS4 are similar except that they use the slopes from the PLR as the predictor and restrict to only those slopes that have a P value less than or equal to 0.01. Patients who do not have a VF location that meets this P value restriction are given a covariate value of zero to indicate this. These models account for statistical significance through the P value restriction as well as the rate of progression through use of the slope values and, therefore. more closely represent progression methods used in practice.

We considered multiple model comparison and validation measures to differentiate the models in terms of fit and prediction. The deviance information criterion (DIC) is a commonly considered model comparison tool used in the Bayesian setting where the posterior mean of the deviance statistic, $\bar{D}$, is used to assess model fit while the effective number of parameters, p_D, is used to describe the model complexity. The DIC then is defined as DIC $= \bar{D} + p_D$, with smaller values being preferred.[14]

D_k is a posterior predictive loss model comparison metric that considers the balance between model fit (G_k) and complexity (P).[15] We work with the deviance of the Bernoulli distribution (with continuity correction) and set $k = 10^{10}$ to calculate D_k. As with DIC, smaller values of D_k suggest an improved model fit. D_k is most useful in comparing the predictive performance of a model while DIC is preferred if the model is used mainly for explanatory purposes.

The receiver operating characteristic (ROC) curve also was calculated for each model along with the area under the ROC curve (AUC), the optimal sensitivity, and optimal specificity values. We formally compared the AUC values from the different models using a paired statistical test and reported the P values. The

optimal sensitivity and specificity values were determined using Youden's Index.[16]

Simulation Study

We carried out a simulation study to demonstrate the ability of our model to efficiently and accurately estimate the individual regression parameters across the VF in a number of different data generation settings and to compare the performance with alternative models. These details and results are presented in the Supplementary Material. The findings suggest that Model 1 is robust to the data generation setting and is preferred most often to the competing models in the simulation framework.

Sensitivity Analyses

We performed a number of sensitivity analyses to assess the adequacy and robustness of our model to the choice of VF predictor (slopes), fixed versus random settings for δ, selected σ_η^2 prior distribution, selected neighborhood definition, and choice of the ICAR(σ_η^2) prior distribution. These details and results also are presented in the Supplementary Material. The findings suggested that the observed inference is robust to each of these modeling choices. We also presented an exploratory analysis of the robustness of our findings based on the number of available VF tests in Supplementary Table S5. The results appear to be robust to the number of available VF tests overall.

Model Validation

To validate the newly introduced model, we predicted the probability of being diagnosed as progressing by the clinician expert consensus for patient/eyes in the validation dataset [$\hat{p}_{i0}(\boldsymbol{\beta},\theta_0,\boldsymbol{\eta})$] using the posterior predictive distribution. Recall that this dataset was not used in building the predictive model. We similarly predicted these same probabilities using the competing models.

We compared the predicted probabilities with the actual progression determinations for the patient/eyes by defining $\hat{d}_{i0} = \frac{Y_{i0}-\hat{p}_{i0}(\boldsymbol{\beta},\theta_0,\boldsymbol{\eta})}{\sqrt{\hat{p}_{i0}(\boldsymbol{\beta},\theta_0,\boldsymbol{\eta})\{1-\hat{p}_{i0}(\boldsymbol{\beta},\theta_0,\boldsymbol{\eta})\}}}$, $i = 1,\ldots,n_0$, where n_0 is number of patient/eyes in the validation dataset ($n_0 = 100$). Values of $\hat{d}_{i0}$ near zero indicate accurate prediction while large values indicate the opposite. We explored the distribution of these diagnostic measures for each model to determine if any outliers exist. We also calculated and compared $\sum_{i=1}^{n_0}|\hat{d}_{i0}|$ and $\sum_{i=1}^{n_0}\hat{d}_{i0}^2$ for each model to identify which

model provides the best predictions overall. Smaller values for these summary measures indicate a model is predicting well.

Results

Analysis of Modeling Dataset

We first applied Model 1 to the modeling dataset. Posterior inference for the nonspatial parameters of Model 1 are shown in Table 2. The 95% credible interval (CI) for β_1 does not include zero and is positive. This indicates that a patient/eye followed for a longer period of time is more likely to be diagnosed as progressing by the group of clinicians. The 95% CI for θ_0 does not include zero and is negative. This indicates that a decrease in any of the VF sensitivities over time (negative slope) leads to an increased probability of being diagnosed as progressing by the clinician expert consensus, regardless of spatial location. In Figure 2A, we display the posterior means for the location-specific slope parameters, $\theta_0/\sigma_z + \eta(s_j)$, for Model 1. These parameters represent the increase in an individual's latent variable (Y_i^*) with a 1 dB decrease in sensitivity per year at the respective VF location for a patient/eye. A significant increase in Y_i^* significantly increases the probability of progression diagnosis. Therefore, more negative estimates of $\theta_0/\sigma_z + \eta(s_j)$, as shown in the blue and green regions of Figure 2A, indicate that decreases in VF sensitivities over time at these locations have a greater impact on the clinicians' progression determination. Parameter estimates near zero indicate that decreases in these areas have less impact on the progression determination.

In Figure 2A, in the nasal, inferior-nasal, and superior-nasal regions of the optic disc (see Fig. 1) there is very little impact of decreasing sensitivities over time on the probability of diagnosed progression. However, in the temporal, inferior-temporal, and superior-temporal regions of the optic disc, the parameter estimates are further from zero. This indicates that decreasing sensitivities over time in these regions may be more meaningful in terms of glaucoma progression diagnosis. Locations s_{19}, s_{20}, s_{21}, and s_{22} of the inferior-temporal optic disc region are the only locations to have 95% CIs which fail to include zero [$\theta_0/\sigma_z + \eta(s_{19})$: (−43.89, −1.29); $\theta_0/\sigma_z + \eta[s_{20}]$: (−37.26, −4.64); $\theta_0/\sigma_z + \eta(s_{21})$: (−38.25, −4.48); $\theta_0/\sigma_z + \eta(s_{22})$: (−35.37, −1.19)].

The findings in Figure 2A have an anatomic explanation. The retinal verve fiber layer (RNFL)

bundle is thickest and converges in the superior-temporal and inferior-temporal regions of the optic disc. As a result, early glaucoma VF loss as well as RNFL defects are predominantly observed in these two regions and, typically, glaucomatous damage to the optic disc advances in a sequence beginning with the same regions.

Model Comparisons

In Figure 2, the posterior means of the location-specific slope parameters are displayed for models 1 to 3. The Model 1 results have been described previously. In Figure 2B, the posterior means for Model 2 are displayed. When compared to Model 1 (Fig. 2A), the range of the means are extreme and the posterior standard deviations are larger at each VF location. Figure 2B also lacks a clear pattern in the means across the VF, unlike in Model 1. These results are expected for Model 2 since multicollinearity is known to change the signs of parameter estimates and inflate standard errors. For Model 3 (Fig. 2C), only one parameter is estimated across the entire VF. Figures 2C and 2A are shown on the same scale for comparison purposes. Recall that if Model 3 was the true underlying model, we would expect Figures 2A and 2C to look similar. However, the plots indicate that there are changes across the VF that Model 3 is unable to account for given its constant effect specification.

In Table 3, we display the model fit/prediction results from all considered models. The ROC curve plots are displayed in Figure 3. For Models P1 to P4 and Models PS1 to PS4, we only display the ROC curve plots for the models with the highest AUC value in each group due to the large number of models being compared. Model 2 has the largest DIC and p_D values, indicating poor overall fit to the data. The predictive measures suggest that Model 2 has improved predictive ability when compared to the other models. However, given the DIC results, Model 2 is likely overfitting the data and resulting in improved predictions as a result. In terms of model fit, Model 1 has the smallest DIC when compared to the competing models. Due to the increased flexibility of Model 1, we now have improved predictive performance when compared to the other models (excluding Model 2) as suggested by G_k, P, and D_k ($k = 10^{10}$). The AUC value for Model 1 is significantly larger than the values from all other models (excluding Model 2) with the P vaues from the paired statistical tests ranging from 0.000 to 0.018. Model 1 also has the highest optimal sensitivity among all

Table 3. Model Fit and Prediction Comparisons between the Considered Models

Method	AUC	Sensitivity	Specificity	DIC (p_D)	G_k	P	D_k
Model 1	0.80*	0.92	0.60	198.15 (12.61)	64.50	65.58	130.08
Model 2†	0.91	0.88	0.79	222.04 (47.80)	45.76	43.00	88.76
Model 3	0.72	0.70	0.73	207.58 (3.00)	75.89	70.93	146.83
Model P1	0.65	0.82	0.45	212.11 (1.93)	80.25	74.80	155.05
Model P2	0.67	0.56	0.72	213.38 (1.94)	80.46	75.13	155.59
Model P3	0.66	0.54	0.72	212.34 (1.94)	80.18	74.70	154.88
Model P4	0.68	0.54	0.75	210.49 (1.95)	79.23	74.15	153.38
Model PS1	0.70	0.56	0.80	201.29 (2.00)	73.62	68.45	142.08
Model PS2	0.67	0.50	0.83	202.82 (2.01)	73.85	69.07	142.92
Model PS3	0.59	0.32	0.90	210.80 (2.00)	77.65	72.44	150.09
Model PS4	0.60	0.28	0.91	210.13 (2.00)	77.37	72.05	149.42

Larger values of AUC, sensitivity, and specificity are preferred while smaller values of DIC and D_k are preferred ($k = 10^{10}$).
* Model 1 AUC value is statistically larger than the AUC values from all other models (excluding Model 2 due to overfitting).
† Unstable results due to overfitting. Results will not be generalizable to different datasets.

models. The competing models are outperformed by Model 1 in terms of model fit and prediction, indicating that accounting for spatial variability across the VF is an important feature of the analysis.
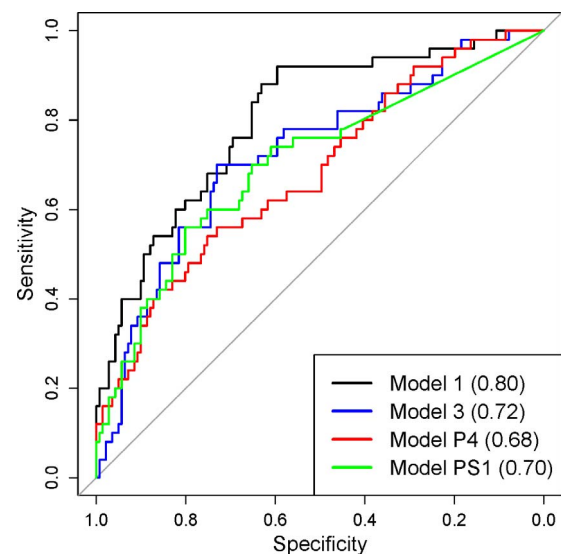
Model Validation

The results from the model validation analyses can be seen in Table 4. As suspected, Model 2 is overfitting and provides extremely poor predictions for the validation dataset. Model 3 provides similar predictive ability to Model 1 overall. Models P1 to P4 result in large outlying values of $\hat{d}_{i0}$, indicating poor prediction for those patient/eyes. Models PS2 to PS4 predict slightly better than Models P1 to P4 overall but are still outperformed by Model 1. These results provide further evidence that the newly introduced model is not suffering from overfitting and suggests that it may be clinically useful.

Discussion

Because of the chronic and progressive nature of glaucoma, an important requirement of the long-term follow-up is establishing whether the disease is stable or has progressed, which oftentimes is challenging to the clinician. One of the methods commonly used in clinical practice to judge progression is visual observation of serial automated VF test results along with other clinical and nonclinical information. This approach is quicker and easy to perform for experienced clinicians, but it is subjective without quantification of the change and, therefore, the

outcome may be very variable even among glaucoma specialists.[17,18]

This has led to the introduction of a number of objective methods to assess VF progression in glaucoma patients. The majority of these methods use statistical modeling in the form of simple linear regression analyses where VF measures collected in frequent follow-up visits are analyzed over time. These measures include the raw sensitivities at individual VF locations as well as global indices, such as the mean deviation (MD), pattern standard

Figure 3. Receiver operating characteristic (ROC) curve plots with AUC curves listed in parentheses. Model 2 is omitted due to overfitting.

Table 4. Distribution of Model Validation Measures. Smaller Values Indicate Improved Prediction.

Method	$\sum_{i=1}^{n_0}|\hat{d}_{i0}|$	$\sum_{i=1}^{n_0}\hat{d}_{i0}^2$	$\min\{\hat{d}_{i0}\}$	$\max\{\hat{d}_{i0}\}$
Model 1	100.02	122.23	−1.22	3.02
Model 2*	412.41	16,430.20	−10.37	70.00
Model 3	101.15	116.62	−1.35	2.03
Model P1	125.90	907.30	−1.26	22.29
Model P2	112.95	260.84	−0.85	11.09
Model P3	110.06	215.12	−0.86	8.74
Model P4	109.70	215.34	−0.91	8.66
Model PS1	335.12	49,677.15	−222.48	2.42
Model PS2	104.32	156.06	−1.09	2.17
Model PS3	107.16	154.48	−0.75	1.93
Model PS4	113.26	174.05	−0.93	1.91

* One data point had to be removed from the Model 2 results because the prediction was in complete disagreement with the outcome, resulting in ∞ for $\hat{d}_{i0}$.

deviation (PSD), and visual field index (VFI).[1,2] Following glaucomatous VF over time using these global indices leads to a reduction of the data to a single measure that ignores the spatial nature of VF loss of glaucomatous progression.[19] The pitfall with PLR is that it is based on the erroneous assumption of progression independence between tests locations and only considers test points that show deepening and not enlargement; this assumption is contradicted by available evidence.[20] Other studies, such as the Advanced Glaucoma Intervention Study[21] and the Collaborative Initial Glaucoma Treatment Study,[22,23] have used scoring systems to condense VF data into a single numeric value. The scoring techniques so far have been used only in clinical trials and not in the clinical practice because they display larger long-term variability and fail to provide information with regard to the spatial characteristics of the defects.

Alternative regression models also have been implemented and compared to PLR, including pointwise exponential and quadratic regression.[3] Event-based methods, which define progression based on a prespecified level of deterioration, also have been considered.[24] Spatial filters have been offered as a method to reduce variability in VF sensitivity measurements before determining progression status.[25] Spatial disease mapping techniques also have been used to model the raw VF sensitivities over time with the goal of identifying significant progression.[4] We proposed a new analytical method that uses the sensitivities of all VF test locations and takes into account their spatial correlation. This model outper-

formed competing models in terms of model fit and predictive ability in the modeling and validation datasets.

Our newly developed method most closely resembles the previously introduced pointwise methods. In the pointwise methods, a progression determination is made based on the location and severity of decreasing sensitivities across the VF. At each VF location, the slope and/or P value are evaluated and a decision about progression is made based on these 52 values. Our method also attempts to summarize the information contained in these 52 values. However, instead of offering a new deterministic method to make the progression decision, our method uses the introduced statistical model to optimize the decision-making process using all of the information simultaneously. This data-driven method is shown to outperform competing methods with similar forms. We also note that without the use of the spatial prior distribution (Model 2), the method would be essentially ignoring the spatial location of vision deterioration. So while the forms of the model are similar conditional on the spatial random effects, unconditionally Model 1 allows for sharing of information across the VF during estimation. As shown in Figure 2, this results in stable and interpretable spatial patterns that are driven by the observed data as well as our prior understanding of the structure-function correlation between the VF and optic nerve head.

The major limitation of VF testing is short-term (at each test location) and long-term fluctuation (between tests) of the sensitivity thresholds as a result of measurement error and the combination of the psychophysical nature of the test and biological factors. The long-term variability between tests is concerning and becomes an issue when serial VFs are used to monitor disease progression. The fact that reduction of the long-term fluctuation of VF sensitivities is not a component of this analysis may be regarded as a limitation. However, if one considers that there is no standard method for determining glaucoma progression by evaluation of longitudinal VF data and that most widely used methods, such as the GPA and the Octopus EyeSuite (Haag-Streit, Koenig, Switzerland), do not attempt to reduce the long-term variability, our method is acceptable based on its performance. In addition, while developing and validating the model described herein using spatially correlated VF data from clinically confirmed perimetric glaucoma patients only may be viewed also as a limitation, it is, however, important to note that this method also may be used successfully in glaucoma

suspects, preperimetric glaucoma patients, and other conditions where VF is used to monitor disease progression, such as nonglaucomatous optic neuropathies. Although we chose the Garway-Heath's map because it is the most widely used based on the good structure-function relationship it yields,[26] it is worth noting that a positive attribute of the proposed model is its flexibility as it may be applied easily to other glaucoma VF cluster map schemes. Studies comparing different cluster map schemes may be needed in the future to determine the scheme with the strongest model fit and highest predictive ability to detect glaucoma progression.

The complexity of the model is not a factor when classifying new patients. The VF data for a new patient can easily be entered into the regression framework and a predicted probability of progression, as defined by the consensus of expert clinicians, is calculated within a few seconds using the previously obtained posterior samples from the model parameters. This predicted probability then can be used to classify the new patient based on the optimal threshold value determined during the ROC curve analysis. This speed will allow for real time decision making in the clinical setting, which constitutes a real strength of the proposed model.

In conclusion, we presented and described a tool for the analysis of spatially correlated longitudinal VF data and demonstrated its usefulness in the modeling and validation datasets. With both datasets, we have established that the proposed method outperformed competing models in terms of model fit and predictive ability without sacrificing information. Using the introduced regression modeling framework, our model can be extended in future work to incorporate other covariates, such as RNFL, ganglion cell-inner plexiform layer and optic disc topographic measures, and patient level attributes that could provide additional information for decision making. Our newly developed model was able to jointly consider highly correlated changes across the entire VF when determining the progression status based on clinician expert consensus, similar to the actual process used by expert clinicians. Results from the validation dataset suggest that the model may also be clinically useful.

Acknowledgments

References

1. Fitzke FW, Hitchings RA, Poinoosawmy D, McNaught AI, Crabb DP. Analysis of visual field progression in glaucoma. *Br J Ophthalmol*. 1996;80:40–48.

2. Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression. *Am J Ophthalmol*. 2008;145:343–353.

3. Caprioli J, Mock D, Bitrian E, et al. A method to measure and predict rates of regional visual field decay in glaucoma. *Invest Ophthalmol Vis Sci*. 2011;52:4765–4773.

4. Betz-Stablein BD, Morgan WH, House PH, Hazelton ML. Spatial modeling of visual field data for assessing glaucoma progression. *Invest Ophthalmol Vis Sci*. 2013;54:1544–1553.

5. Tanna AP, Bandi JR, Budenz DL, et al. Interobserver agreement and intraobserver reproducibility of the subjective determination of glaucomatous visual field progression. *Ophthalmology*. 2011;118:60–65.

6. Besag J. Spatial interaction and the statistical analysis of lattice systems. *J R Stat Soc Series B Stat Methodol*. 1974;36:192–236.

7. Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology*. 2000;107:1809–1815.

8. Lu H, Carlin BP. Bayesian areal wombling for geographical boundary analysis. *Geogr Anal*. 2005;37:265–285.

9. Lu H, Reilly CS, Banerjee S, Carlin BP. Bayesian areal wombling via adjacency modeling. *Environ Ecol Stat*. 2007;14:433–452.

10. Lee D, Mitchell R. Boundary detection in disease mapping studies. *Biostatistics*. 2011;15:457–469.

11. Roberts GO. Markov chain concepts related to sampling algorithms. In: Gilks W, Richardson S, Spiegelhalter D, eds. *Markov Chain Monte Carlo*

in Practice. Suffolk, UK: Chapman & Hall; 1996: 45–57.

12. R Core Team. R: *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2014. Available at: http://www.R-project.org/.

13. Smith SD, Katz J, Quigley HA. Analysis of progressive change in automated visual fields in glaucoma. *Invest Ophthalmol Vis Sci*. 1996;37: 1419–1428.

14. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Series B Stat Methodol*. 2002;64:583–639.

15. Gelfand AE, Ghosh SK. Model choice: a minimum posterior predictive loss approach. *Biometrika*. 1998;85:1–11.

16. Youden, WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–35.

17. Viswanathan AC, Crabb DP, McNaught AI, et al. Inter-observer agreement on visual field progression in glaucoma: a comparison of methods. *Br J Ophthalmol*. 2003;87:726–730.

18. Iester M, Corallo G, Capris E, Capris P. Agreement in detecting glaucomatous visual field progression by using guided progression analysis and Humphrey overview printout. *Eur J Ophthalmol*. 2011;21:573–579.

19. Chauhan BC, Drance S. The use of visual field indices in detecting changes in the visual field in glaucoma. *Invest Ophthalmol Vis Sci*. 1990;31: 512–520.

20. Boden C1, Blumenthal EZ, Pascual J, McEwan G, Weinreb RN, Medeiros F, Sample PA. Patterns of glaucomatous visual field progression identified by three progression criteria. *Am J Ophthalmol*. 2004;138:1029–1036.

21. Advanced Glaucoma Intervention Study. 2. Visual field test scoring and reliability. *Ophthalmology*. 1994;10:144–155.

22. Gillepsie BW, Musch DC, Guire KE, et al. The CGITS Study Group. The collaborative initial glaucoma treatment study: baseline visual field and test-retest variability. *Invest Ophthalmol Vis Sci*. 2003;44:2613–2620.

23. Musch DC, Lichter PR, Guire KE, Standardi CL. The CIGTS Study Group. The Collaborative Initial Glaucoma Treatment Study: study designs, methods, and baseline characteristics of enrolled patients. *Ophthalmology*. 1999;106:653–62.

24. Bengtsson B, Lindgren A, Heijl A, Lindgren G, Asman P, Patella M. Perimetric probability maps to separate change caused by glaucoma from that caused by cataract. *Acta Ophthalmol Scand*. 1997; 75:184–188.

25. Gardiner SK, Crabb DP, Fitzke FW, Hitchings RA. Reducing noise in suspected glaucomatous visual fields by using a new spatial filter. *Vision Res*. 2004;44:839–848.

26. Gardiner SK, Johnson CA, Cioffi GA. Evaluation of the structure-function relationship in glaucoma. *Invest Ophthalmol Vis Sci*. 2005;46: 3712–3717.

translational vision science & technology