

UCLA

UCLA Previously Published Works

Title

A Bayesian Hierarchical Spatial Longitudinal Model Improves Estimation of Local Macular Rates of Change in Glaucomatous Eyes.

Permalink

<https://escholarship.org/uc/item/5qd4j8v9>

Journal

Translational Vision Science and Technology, 13(1)

Authors

Su, Erica

Mohammadzadeh, Vahid

Mohammadi, Massood

et al.

Publication Date

2024-01-02

DOI

10.1167/tvst.13.1.26

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

A Bayesian Hierarchical Spatial Longitudinal Model Improves Estimation of Local Macular Rates of Change in Glaucomatous Eyes

Erica Su¹, Vahid Mohammadzadeh², Massood Mohammadi², Lynn Shi², Simon K. Law², Anne L. Coleman², Joseph Caprioli², Robert E. Weiss¹, and Kouros Nouri-Mahdavi²

¹ Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, California, USA

² Glaucoma Division, Stein Eye Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA

Correspondence: Kouros Nouri-Mahdavi, 100 Stein Plaza, Los Angeles, CA 90095, USA. e-mail: nouri-mahdavi@jsei.ucla.edu

Received: November 13, 2022

Accepted: December 6, 2023

Published: January 29, 2024

Keywords: macular OCT; ganglion cell complex; Bayesian; hierarchical model; simple linear regression

Citation: Su E, Mohammadzadeh V, Mohammadi M, Shi L, Law SK, Coleman AL, Caprioli J, Weiss RE, Nouri-Mahdavi K. A bayesian hierarchical spatial longitudinal model improves estimation of local macular rates of change in glaucomatous eyes. *Transl Vis Sci Technol.* 2024;13(1):26, <https://doi.org/10.1167/tvst.13.1.26>

Purpose: Demonstrate that a novel Bayesian hierarchical spatial longitudinal (HSL) model improves estimation of local macular ganglion cell complex (GCC) rates of change compared to simple linear regression (SLR) and a conditional autoregressive (CAR) model.

Methods: We analyzed GCC thickness measurements within 49 macular superpixels in 111 eyes (111 patients) with four or more macular optical coherence tomography scans and two or more years of follow-up. We compared superpixel-patient-specific estimates and their posterior variances derived from the latest version of a recently developed Bayesian HSL model, CAR, and SLR. We performed a simulation study to compare the accuracy of intercept and slope estimates in individual superpixels.

Results: HSL identified a significantly higher proportion of significant negative slopes in 13/49 superpixels and a significantly lower proportion of significant positive slopes in 21/49 superpixels than SLR. In the simulation study, the median (tenth, ninetieth percentile) ratio of mean squared error of SLR [CAR] over HSL for intercepts and slopes were 1.91 (1.23, 2.75) [1.51 (1.05, 2.20)] and 3.25 (1.40, 10.14) [2.36 (1.17, 5.56)], respectively.

Conclusions: A novel Bayesian HSL model improves estimation accuracy of patient-specific local GCC rates of change. The proposed model is more than twice as efficient as SLR for estimating superpixel-patient slopes and identifies a higher proportion of deteriorating superpixels than SLR while minimizing false-positive detection rates.

Translational Relevance: The proposed HSL model can be used to model macular structural measurements to detect individual glaucoma progression earlier and more efficiently in clinical and research settings.

Introduction

Detection of change over time is often important for the proper treatment of chronic diseases and is crucial to the management of glaucoma. Glaucoma is particularly suited to applying statistical and machine learning models to detect disease and to identify its progression, because most disease-related outcome measures are quantifiable.^{1–6} Relevant outcomes include the optic disc size, neuroretinal rim, retinal nerve fiber layer,

or macular thickness measures along with a host of quantitative variables related to visual fields.

Global, regional, or local structural and functional measures have frequently been used as metrics to detect differences of a specified amount from baseline, commonly called *event analyses*, or to estimate rates of change, also known as *trend analyses*.^{7–9} Longitudinal analyses of local structural or functional measures frequently rely on repeated simple linear regression (SLR) of such measures from single eyes against time to estimate eye-specific rates of change at macular

superpixels, optic disc or retinal nerve fiber layer sectors, or visual field test locations or clusters.^{9–13} Although SLR is easy to do and seems intuitive, there are multiple issues with this approach, which although occasionally acknowledged, have mostly been ignored.^{14–16} Simple linear regression models ignore the spatial relationships among local structural (sectors, superpixels) or functional (visual field clusters or locations) measures. Another shortcoming of SLR is that valuable population information from the cohort is not used to refine estimated rates of change. Similarly, there is no formal way to specify the correlation of baseline measurements (the intercepts) with slopes (rates of change). This is an important issue because baseline measurements influence the magnitude of rates of change in an optical coherence tomography (OCT) sector or superpixel.⁹ Accounting for baseline structural measurements can lead to a reduction of the estimated variability for rates of change (slopes). A few prior studies have proposed linear mixed or hierarchical models to address some of the above shortcomings.^{14,15,17,18} For example, Montesano et al.¹⁴ applied a hierarchical linear mixed model to 24-2 visual field data to estimate global and local rates of change in individual eyes. To account for spatial correlation, Betz-Stablein et al.¹⁹ developed a model using conditional autoregressive (CAR) priors on intercepts and slopes for individual-level 24-2 visual field data. However, in either case, no attempt was made to include the cohort's population data to help estimate individual eye parameters. The cohort's data has information on the distribution of possible slopes, and using this information as a prior results in more accurate slope estimates. Such models are challenging to specify correctly and run and can require a significant investment of time and computer CPU power.

Our team has developed several versions of a novel Bayesian hierarchical spatial longitudinal (HSL) model to improve estimation of local macular thickness rates of change in a prospective cohort of glaucoma patients.^{4,20} This ongoing project aims to provide a longitudinal framework to estimate global and most importantly local rates of change more precisely across the macula within individual eyes while at the same time overcoming the inadequacies of the SLR approach. Each successive model improves on the previous version with additional modeling features. The current model adds spatially correlated patient-specific random effects and spatially correlated visit effects. The goal of the current work is to compare the performance of the latest HSL model to that of SLR and a CAR model. We hypothesized that the HSL model would provide more accurate estimates of local macular rates of change thanks to reduced variance

for those rates of change compared to SLR and more appropriate shrinkage toward the population averages by location than CAR.

Methods

One hundred eleven eyes (111 patients) from the Advanced Glaucoma Progression Study (AGPS) were included in this study. The AGPS is an ongoing longitudinal study at the University of California Los Angeles. The university's Institutional Review Board approval was obtained for this study. The study adhered to the tenets of the Declaration of Helsinki and conformed to Health Insurance Portability and Accountability Act policies. All patients provided written informed consent at the time of enrollment in the study. Inclusion criteria for enrolled eyes were as follow: (a) clinical diagnosis of primary open-angle glaucoma, pseudoexfoliative glaucoma, pigmentary glaucoma, or primary angle-closure glaucoma and (b) evidence of either central damage on 24-2 visual field (VF), defined as two or more points within the central 10° with $P < 0.05$ on the pattern deviation plot or VF mean deviation (MD) worse than -6 dB. Exclusion criteria consisted of baseline age less than 40 years or greater than 80 years, best-corrected visual acuity $<20/50$, refractive error exceeding 8 diopters of sphere or 3 diopters of cylinder, or significant retinal or neurological disease affecting OCT measurements. Eyes with four or more OCT scans and two or more years of follow-up were included. We analyzed observations up to 4.25 years after baseline. Data from visits less than 0.2 years after a previous visit were omitted.²⁰

Macular OCT Imaging

Macular volume scans were obtained with a Spectralis spectral-domain OCT (Heidelberg Engineering, Heidelberg, Germany). The Posterior Pole Algorithm of the Spectralis OCT acquires $30^\circ \times 25^\circ$ volume scans of the macula (61 B-scans spaced approximately 120 μm apart) centered on the fovea and repeated nine to 11 times to reduce speckle noise. Proprietary software of Spectralis OCT, the Glaucoma Module Premium Edition, was used to automatically segment individual retinal layers before data export. Images were reviewed for segmentation errors and image artifacts. Segmentation errors were manually rectified with the OCT device's built-in software. After segmentation, the individual layer thickness measurements are provided as 8×8 arrays of $3^\circ \times 3^\circ$ superpixels for the central $24^\circ \times 24^\circ$ region centered on

the fovea. Because of substantial measurement noise in the peripheral macular regions, we only included the ganglion cell complex (GCC) thickness within 7×7 arrays of superpixels from Spectralis OCT macular volume scans after excluding the most inferior row and nasal column of superpixels. The GCC thickness was calculated by adding the thickness measurements of the retinal nerve fiber layer, ganglion cell layer, and inner plexiform layer.

Data Management and Statistical Modeling

Our methods have been described previously.²⁰ We identified potential outliers as observations with very large increases or decreases between consecutive measurements and removed observations that produced the greatest reduction in the sum of absolute consecutive differences resulting in removing approximately 0.5% of observations as outliers. We removed all observations for a single person in a single superpixel if we identified two or more outliers in that superpixel.²⁰ We then fit a Bayesian normal hierarchical random effects model using the JAGS package in R (*R2jags*).^{21,22} The current version of our Bayesian HSL model includes (a) a macula-wide global intercept and slope, (b) superpixel-specific intercept and slope deviations from the global intercept and slope (superpixel-level random effects), (c) macula-wide patient-specific intercept and slope effects (global patient random effects), (d) patient-superpixel specific intercept and slope random effects, and (e) macula-wide visit effects. Letting y_{ijk} denote a single observation of GCC thickness (μm) for patient i at time t_{ij} in superpixel k , the model is

$$y_{ijk} = \mu_0 + \mu_1 t_{ij} + \alpha_{0k} + \alpha_{1k} t_{ij} + \beta_{0ik} + \beta_{1ik} t_{ij} \\ + \phi_{0k} PC_{0i} + \phi_{1k} PC_{1i} t_{ij} + \phi_{2k} VE_{ij} + \epsilon_{ijk}$$

$$\epsilon_{ijk} \sim N(0, \sigma_k^2 \exp(PC_{2i})),$$

where μ_0 is the global macula-wide intercept; μ_1 is the global macula-wide slope; α_{0k} is the population average intercept in superpixel k ; α_{1k} is the population average slope in superpixel k ; β_{0ik} is the patient-superpixel intercept for the i th patient; β_{1ik} is the patient-superpixel interaction slope for the i th patient; ϕ_{0k} is the spatial effects eigenvector component for the patient-specific intercepts; ϕ_{1k} is the spatial effects eigenvector component for the patient-specific slopes; ϕ_{2k} is the spatial visit effects eigenvector component; PC_{0i} is the macula-wide patient-specific intercept random effect; PC_{1i} is the macula-wide patient-specific slope random effect; PC_{2i} is the macula-wide

patient-specific log residual variance random effect; VE_{ij} is the macula-wide visit random effect; the population component of the patient-superpixel log residual variance is $\log \sigma_k^2$. Our model uses novel, semi-parametric spatial effects for the intercept $\phi_{0k} PC_{0i}$, slope $\phi_{1k} PC_{1i}$, residual variance $\sigma_k^2 \exp(PC_{2i})$, and visit effects $\phi_{2k} VE_{ij}$. In each component, terms with a k subscript ϕ_{0k} , ϕ_{1k} , σ_k^2 , and ϕ_{2k} define the spatial pattern of variation across the macula whereas the random effects PC_{0i} , PC_{1i} , $\exp(PC_{2i})$, and VE_{ij} are random patient effects, or visit effects, that indicate how much patient i 's observations deviate from the population averages. Each of these four terms are factor analytic models with one factor each where the ϕ_{0k} , ϕ_{1k} , σ_k^2 , and ϕ_{2k} are called *loadings* and the random effects are also called *factors*.²³ Superpixel-specific population intercepts and slopes are the sum of the macula-wide global intercept and slope plus the superpixel-specific intercept and slope deviations. Similarly, a superpixel-patient slope (intercept) is the sum of (a), (b), (c), and (d): the sum of the macula-wide global slope (intercept), superpixel-specific slope (intercept), the macula-wide patient-specific intercept slope (intercept), and the patient-superpixel specific slope (intercept) random effect. Patient-superpixel log residual variances are modeled as the sum of a superpixel component and a patient-specific component. The outlier removal algorithm and full model and priors are given in the Supplemental material.

In Bayesian models, inference is made by summarizing the posterior distributions of the parameters of interest. The posterior distribution is obtained by combining the likelihood (information in the data) with the prior (prior knowledge about the unknown parameters) and quantifies the uncertainty in the parameters after observing the data. The posterior mean (i.e., mean of the posterior distribution of a parameter of interest) is often taken as the Bayesian point estimate; similarly, the posterior standard deviation (SD), that is, the SD of the posterior distribution, summarizes the uncertainty in the posterior. Posterior means and SDs of superpixel-patient-specific intercepts and slopes were calculated from the HSL model and compared to superpixel-patient-specific intercepts and slopes obtained from a Bayesian SLR model with a flat prior analyzing longitudinal data from each superpixel-patient separately and a Bayesian CAR model with intrinsic CAR priors analyzing data from each patient separately.

The Bayesian SLR model produces posterior means and posterior Credible Intervals (CrI) that are identical to the classical point estimate and confidence intervals (CI) from classical least squares regression on the same data. The Bayesian SLR posterior SDs are slightly

larger than classical SLR standard errors (SEs) by a factor of $(\frac{n_i-4}{n_i-2})^{(1/2)}$, where n_i is the number of observations for patient i because SEs are not standard deviations and SEs do not account for the classical estimates being t distributed with $n_i - 2$ degrees of freedom. Using the posterior SD from this Bayesian SLR model puts SLR, CAR, and HSL on an equal footing.

For the CAR model inspired by Betz-Stablein et al.,¹⁹ we assumed GCC thicknesses in neighboring superpixels would be more similar than in more distant superpixels. We defined the neighborhood structure, or adjacency matrix, such that superpixels sharing an edge are weighted as 1. For each patient i , we model spatial dependence in intercepts and slopes

$$y_{jk} = \alpha_0 + \alpha_1 t_j + \beta_{0k} + \beta_{1k} t_j + \epsilon_{jk}$$

$$\epsilon_{jk} \sim N(0, \sigma^2),$$

where α_0 is the overall eye intercept; α_1 is the overall eye slope; β_{0k} is the superpixel-specific spatial deviation from the overall eye intercept; β_{1k} is the superpixel-specific spatial deviation from the overall eye slope; and σ^2 is the residual variance specific to patient i . We fit the CAR models using the R package NIMBLE v0.13.^{24,25} The full CAR model and priors are given in the Supplemental material.

We identified superpixel-patient-specific rates of change as worsening or improving (significantly negative or positive) when the upper or lower limit of the symmetric 95% CrI was less than or greater than zero, respectively. We compared the proportion of significant negative or significant positive slopes at each superpixel from the HSL and SLR models using an $\alpha = 0.05$ level classical McNemar's test with the null hypothesis being that the two proportions were equal.

We summarize differences (SLR – HSL and CAR – HSL) of posterior means and of SDs for individual patient-superpixel intercepts and slopes. We compared differences of posterior means averaged within each of 49 superpixels and also averaged across all 5419 patient-superpixel profiles. We similarly compared the SLR/HSL and CAR/HSL ratios of posterior SDs for superpixel averaged intercepts and slopes, and patient-superpixel intercepts and slopes. Ratios of posterior variances (i.e., ratios of squared SDs) are measures of the improved estimation efficiency of the better model. Individual patient-superpixel intercept and slope posterior means, posterior SDs, and 95% CrI lengths from HSL, CAR, and SLR were plotted against each other to compare the inferences from the three approaches. A *patient-superpixel profile* is the set of measurements over time for a single patient in a single superpixel.

We omitted 18 patient-superpixel profiles in the data cleaning step because there were two or more outliers in a single patient-superpixel profile and omitted two more patient-superpixel profiles with constant GCC at all time points, as SLR, unlike the HSL model or CAR, is unable to provide an appropriate inference. Thus 5419 is slightly smaller than $111 \times 49 = 5439$ patient-superpixel profiles.

Simulation Study

We ran a simulation study to provide comparisons between the accuracy of HSL, CAR, and SLR estimates for rates of change (slopes); this comparison is only possible in a simulation study where true individual eye slopes are known. In the initial step, a true slope and intercept for each superpixel-patient combination (49 superpixels in each of 111 patient-eyes) was set by sampling each from a normal distribution with mean equal to the sum of the superpixel-specific population and macula-wide patient-specific parameters and variance equal to the superpixel intercept and slope random effects variance. All superpixel-specific and patient-specific parameters were set as the posterior means from our model fit to the actual data; log residual variance was set as the posterior mean of the patient plus superpixel log components; visit times were copied from the data set and visit effect parameters were set to the posterior mean of those parameters. In the next step, we generated 100 data sets with random residual errors and visit effects.

We recorded bias, estimator variance, and root mean squared error (RMSE) from the simulation for all patient-superpixel intercepts and slopes. We took the posterior means from the HSL and CAR models as the estimates, and the least squares estimate was the SLR estimate. Bias was calculated as the *average estimate over simulated datasets minus truth*, which is a measure of whether the model estimates the true value on average. Estimator variance is the variance over simulations of the estimate around its average estimate over simulated data sets, a measure of precision. Estimator variance is estimator SD squared. RMSE is the square root of the sum of squared bias and estimator variance

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{\sum_{m=1}^{100} (\text{estimate}_m - \text{truth}_m)^2}{100}} \\ &= \sqrt{\text{bias}^2 + \text{estimator variance}}, \end{aligned}$$

where m indexes data sets. The RMSE is the gold standard for assessing the accuracy of a model for estimating the true parameter value. We calculated average over simulations of 95% Credible Interval

Length (CrIL), which is the length of the interval with 2.5% posterior probability to the left and to the right of the interval for HSL and CAR and the usual 95% CI for SLR. Finally, 95% CrI Coverage is the proportion of the time where the 95% CrI contains the truth. For bias, estimator SD, RMSE, and 95% CrIL, smaller is better. For 95% CrI Coverage, values close to 95% are preferred. For the estimator SD, RMSE, and 95% CrIL, we report the SLR/HSL and CAR/HSL ratios; ratios greater than one represent better performance of HSL. We report superpixel summaries by averaging reported measures over all patients at each superpixel location and global summaries by averaging each component over all patients and superpixel locations.

Results

A total of 39,625 GCC superpixel measurements in 49 superpixels from 815 visits were included in the analysis. Supplementary Table S1 presents clinical and demographic characteristics of the study sample. Mean (SD) baseline 10-2 visual field mean deviation was -8.9 (5.9) dB. Mean (SD) follow-up time was 3.6 (0.4) years with an average (SD) of 7.3 (1.1) OCT scans per eye.

Simulation Results

Table 1 presents the average RMSE, 95% CrIL, and 95% CrI coverage probability for SLR, CAR, and HSL

Table 1. Averages Over all Patient-Superpixel Intercepts and Slopes of RMSE, CrIL, CrI Coverage Probability

Model	RMSE	95% CrIL	95% CrI Coverage Probability
Intercept			
HSL	1.07	4.21	0.949
CAR	1.31	4.24	0.882
SLR	1.44	7.17	0.951
Ratio (CAR/HSL)	1.25	1.03	
Ratio (SLR/HSL)	1.38	1.70	
Slope			
HSL	0.38	1.55	0.947
CAR	0.58	1.59	0.810
SLR	0.69	3.44	0.951
Ratio (CAR/HSL)	1.65	1.08	
Ratio (SLR/HSL)	2.05	2.25	

Lower RMSE and 95% CrIL indicates better model performance and ratios greater than one favor HSL over CAR or SLR. The HSL model outperforms SLR for RMSE and 95% CrIL for both intercepts and slopes. CAR does substantially worse on 95% CrI coverage probability compared to HSL and SLR.

averaged over all 5439 patient-superpixel for intercepts and for slopes. Compared to SLR, the HSL model had substantially better RMSE (intercept SLR/HSL mean ratio: 1.38; slope SLR/HSL mean ratio: 2.05) and 95% CrIL (intercept SLR/HSL mean ratio: 1.70; slope SLR/HSL mean ratio: 2.25), and similar 95% CrI coverage probability. Compared to CAR, the HSL model had better RMSE (intercept CAR/HSL mean ratio: 1.25; slope CAR/HSL mean ratio: 1.65), similar 95% CrIL, and much better 95% CrI coverage probability (HSL vs. CAR coverage probability for intercepts: 0.949 vs. 0.882; for slopes: 0.947 vs. 0.810).

Patient-Superpixel Intercepts and Slopes

Supplementary Table S2 presents summaries of the grand mean and tenth to ninetieth percentile range of posterior means, posterior SDs, estimator SDs, 95% CrILs, RMSEs, and 95% CrI coverage probabilities for intercepts and slopes across all 5439 patient-superpixel combinations for HSL, CAR, SLR, and the differences between the means (SLR – HSL and CAR – HSL) and ratios (SLR/HSL and CAR/HSL) of the SDs, CrILs, and RMSEs.

Intercepts. For the 5439 individual patient-superpixel intercepts, SLR, CAR, and HSL models demonstrated similar average, tenth-, and ninetieth-percentile posterior means. The SLR model had larger (worse) estimator SDs (ratio range 1.12–2.31) and larger (worse) 95% CrILs (ratio range 1.23–3.09) than HSL for all patient-superpixels. Although CAR had similar posterior SDs and 95% CrILs to HSL, HSL had higher (better) 95% CrI coverage probabilities for 80.2% (4364 out of 5439) of patient-superpixel intercepts. Compared to SLR and CAR, the HSL model had lower (better) RMSE values for 95.1% (5174/5439) and 92.7% (5028/5439) of patient-superpixel intercepts, respectively.

Slopes. Averaged across all 5439 estimated patient-superpixel slopes, SLR, CAR, and HSL had similar average slope estimates. However, the tenth to ninetieth percentile range of SLR slope posterior means ($-1.36, 0.48 \mu\text{m}/\text{year}$) was much wider than HSL ($-1.17, 0.19 \mu\text{m}/\text{year}$) and CAR ($-1.01, 0.14 \mu\text{m}/\text{year}$). The SLR model had larger (worse) estimator SDs than HSL (SLR/HSL ratio range 1.17–11.84) and 95% CrILs (SLR/HSL ratio range 1.32–6.44) for all patient-superpixel slopes. The CAR model had larger estimator SDs than HSL (CAR/HSL ratio range 0.71–7.57) with comparable 95% CrILs (CAR/HSL ratio range 0.55–2.75); however, HSL had better 95% CrI coverage probabilities for 4965/5439 (92.1%) patient-superpixel slopes. The average 95% CrI coverage probabilities for patient-superpixel slopes were 0.951, 0.810, and 0.949 for SLR, CAR, and HSL, respectively. Compared to

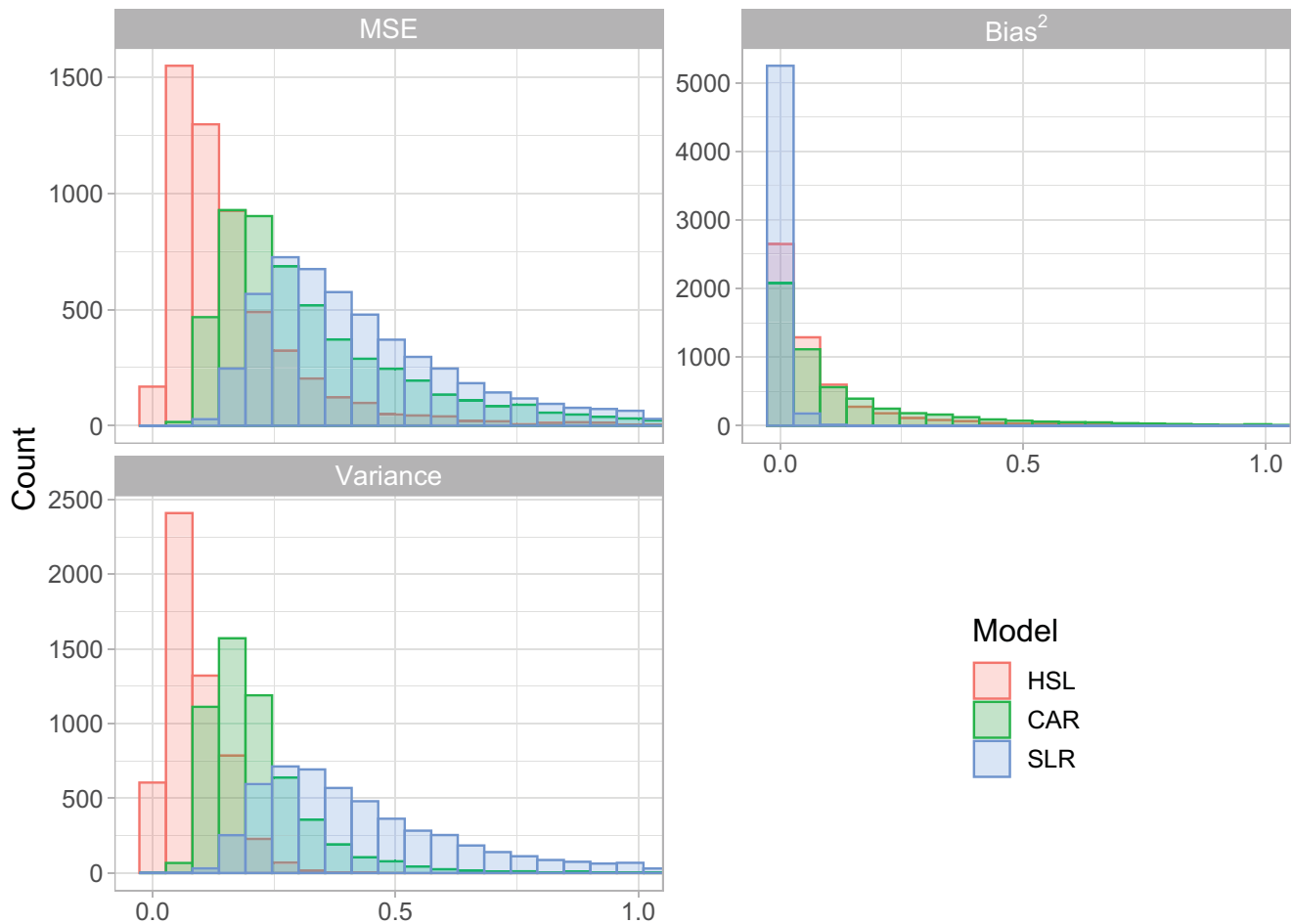


Figure 1. Histograms of MSE (upper left), squared bias (upper right), and estimator variance (lower left) for patient-superpixel slopes comparing SLR, CAR, and HSL models pooled across all superpixels from the simulation study. Mean squared error is squared bias plus estimator variance; the x-axis values are directly comparable in all three plots. Lower MSE indicates better model performance. Counts of large values omitted due to truncating the x-axis at 1: MSE (HSL = 43; CAR = 241; SLR = 500), squared bias (HSL = 31; CAR = 120; SLR = 0), and estimator variance (HSL = 0; CAR = 8; SLR = 492).

SLR and CAR, the HSL model had lower RMSE values for 5211/5439 (95.8%) and 5097/5439 (93.7%) patient-superpixel slopes, respectively. Figure 1 plots histograms of mean squared error (MSE), squared bias, and estimator variance for patient-superpixel slopes for all patients and superpixels for the SLR (blue), CAR (green), and HSL (red) models. The HSL model had noticeably smallest MSE because of having much smaller estimator variance compared to CAR and SLR. Equivalent histograms are plotted for the intercepts in Supplementary Figure S1, where the HSL model also demonstrates noticeably smaller MSE compared to CAR and SLR.

Superpixel-Averages of Intercepts and of Slopes

For the intercepts, there was significant bias averaged across patients for HSL (range -0.054 to $0.082 \mu\text{m}$) and CAR (range -0.486 to $0.290 \mu\text{m}$) for 28/49 and 45/49 superpixels, respectively, while SLR

had no bias in any superpixels. While the magnitude of the observed bias for HSL was not clinically relevant both in comparison to the size of the intercepts (posterior mean range 52.9 – $101.6 \mu\text{m}$) and compared to the posterior SDs, the magnitude of bias for CAR was more comparable to the posterior SDs. In contrast, SLR had substantially larger (worse) estimator SD (ratio range 1.29 – 1.74), RMSE (ratio range 1.19 – 1.58), and 95% CrILs (ratio range 1.47 – 1.94) than HSL for all 49 superpixels. The CAR model had larger estimator SD (ratio range 0.97 – 1.40) and RMSE (ratio range 1.15 – 1.36) than HSL across superpixels, but similar 95% CrILs (ratio range 0.76 – 1.33). There were no important differences (fraction coverage = 0.95) in 95% CrI coverage probability for HSL and SLR, but noticeable differences in parafoveal and nasal superpixels for CAR (HSL range -0.014 to 0.008 ; CAR range -0.169 to 0.012 ; SLR range -0.004 to 0.006). Superpixel averages of estimator SD, RMSE, 95% CrIL, and

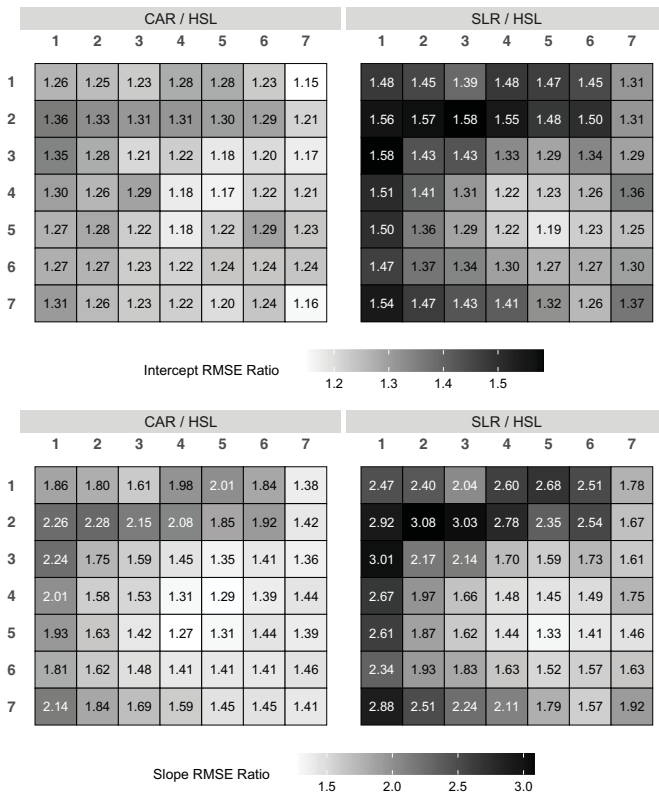


Figure 2. Heat map of the average RMSE ratio for SLR over HSL or CAR over HSL for intercepts (top) and slopes (bottom) by superpixel location from the simulation study. HSL outperforms SLR and CAR in all superpixels as ratios greater than one favor HSL over the alternative model.

95% CrI coverage probability for intercepts and slopes are shown in Supplementary Figures S2, S3, S4 and S5, respectively.

For the slopes, there was significant bias for HSL (range -0.040 to 0.038 $\mu\text{m}/\text{year}$) and CAR (range -0.155 to 0.252 $\mu\text{m}/\text{year}$) for 27/49 and 45/49 superpixels, respectively. The magnitude of the bias for HSL was generally modest in comparison to the slopes (range -0.982 to 0.025 $\mu\text{m}/\text{year}$) and was small in comparison to posterior SDs; however, the magnitude of the bias for CAR was on par with the slope estimates. For the 49 superpixel slopes, SLR had larger (worse) estimator SD (ratio range 1.55–4.56), RMSE (ratio range 1.33–3.08), and 95% CrILs (ratio range 1.63–3.14) than HSL across superpixels. CAR had larger estimator SD (ratio range 0.94–3.05) and RMSE (ratio range 1.27–2.28) than HSL across superpixels, but similar 95% CrILs (ratio range 0.73–1.69). There were no substantive differences (coverage -0.95) in 95% CrI coverage probability for HSL and SLR, but noticeable differences in parafoveal and nasal superpixels for CAR (HSL range -0.033 to 0.024 ; CAR range -0.273 to -0.006 ; SLR range -0.004 to 0.005). Figure 2 displays

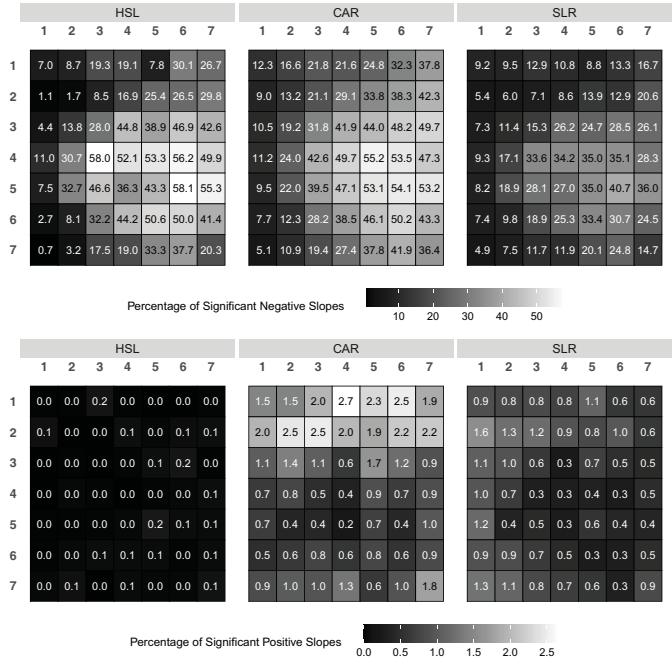


Figure 3. Heat map of the percentage of significant negative slopes (top) and significant positive slopes (bottom) detected by SLR, CAR, and HSL models when the true slope is negative in the simulation study.

heat maps of the intercept and slope average RMSE ratios (CAR/HSL and SLR/HSL) across superpixels. In all superpixels, HSL outperforms CAR and SLR with all ratios >1 . Figure 3 gives the proportion of significant negative and positive slopes detected by HSL, CAR, and SLR when the true slope is negative. The HSL model detected a higher proportion of significant negative slopes than SLR in 38/49 superpixels, with notably larger differences in the central superpixels. It also detected a lower proportion of significant positive slopes than SLR in all superpixels. In contrast, CAR detected a higher proportion of significant negative slopes in 36/49 superpixels, but also a higher proportion of significant positive slopes in all 49 superpixels than HSL when the true slopes were negative. The proportions of significant negative and positive slopes detected by HSL, CAR, and SLR when the true slope is positive are shown in Supplementary Figure S6. The HSL model detected a lower proportion of significant negative slopes in 42/49 superpixels than SLR when the true slopes were positive and a lower proportion of significant positive slopes in 43/49 superpixels than SLR. In addition, HSL detected a lower proportion of significant negative slopes in all 49 superpixels and a lower proportion of significant positive in 39/49 superpixels than CAR. A detailed breakdown of the proportion of significant slopes detected by superpixel is shown in Supplementary Table S3.

Table 2. Summary of the Grand Mean of Posterior Means and Posterior SD for Intercepts and Slopes for all 5419 Patient-Superpixel Combinations for SLR, CAR, and HSL Models for the Data From the Patient Cohort

	Mean	Tenth Percentile	Ninetieth Percentile
Intercept posterior mean			
HSL	72.95	52.73	102.97
CAR	72.96	52.49	102.92
SLR	72.96	52.29	103.32
Difference (CAR – HSL)	0.01	–1.36	1.39
Difference (SLR – HSL)	0.01	–1.53	1.55
Intercept posterior SD			
HSL	1.09	0.82	1.40
CAR	1.03	0.72	1.49
SLR	1.61	0.78	2.76
Ratio (CAR/HSL), median	0.92	0.71	1.28
Ratio (SLR/HSL), median	1.35	0.76	2.43
Slope posterior mean			
HSL	–0.39	–1.19	0.26
CAR	–0.41	–1.40	0.60
SLR	–0.41	–1.80	0.87
Difference (CAR – HSL)	–0.02	–0.76	0.71
Difference (SLR – HSL)	–0.02	–0.85	0.82
Slope posterior SD			
HSL	0.40	0.27	0.56
CAR	0.41	0.23	0.59
SLR	0.77	0.35	1.33
Ratio (CAR/HSL), median	0.96	0.63	1.55
Ratio (SLR/HSL), median	1.69	0.93	3.28

The difference in posterior means for intercepts and slopes is defined as SLR minus HSL or CAR minus HSL. The ratio in posterior SDs is defined as SLR over HSL or CAR over HSL and we report the median (tenth, ninetieth percentile) ratio of posterior SDs.

Analysis of the AGPS Data

Table 2 presents summaries of the grand mean and tenth to ninetieth percentile range of posterior means and SDs for intercepts and slopes across all 5419 patient-superpixel combinations for HSL, CAR, and SLR and the differences between the means (CAR – HSL and SLR – HSL) and ratios of the SDs. Globally, HSL, CAR, and SLR posterior means were similar on average for both the intercepts (CAR – HSL mean difference [tenth, ninetieth percentile]: 0.01 [–1.36, 1.39 μm]; SLR – HSL difference: 0.01 [–1.53, 1.55 μm]) and slopes (CAR – HSL mean difference [tenth, ninetieth percentile]: –0.02 [–0.76, 0.71 $\mu\text{m}/\text{year}$]; SLR – HSL difference: –0.02 [–0.85, 0.82 $\mu\text{m}/\text{year}$]); however, substantial differences existed across patients and superpixels. The striking difference between HSL and SLR was the systematically higher posterior SDs for SLR compared to HSL particularly for the slopes (median 1.69, tenth, ninetieth percentile = 0.93, 3.28).

The mean and tenth and ninetieth percentiles of posterior means and SDs of slopes by superpixel for HSL, CAR, and SLR are shown in Supplementary Table S4. Although the posterior means of the slopes are similar, the ranges are noticeably smaller and the posterior SDs are also uniformly smaller across all superpixels for HSL and CAR. Across 5419 patient-superpixel curves, the HSL posterior SDs were smaller than those of SLR for 75.2% of intercepts and 87.1% of slopes but larger than those of CAR for 63.4% of intercepts and 54.6% of slopes. Figure 4 displays the median and tenth and ninetieth percentile of posterior SD ratios of CAR/HSL and SLR/HSL. Across all 49 superpixels, HSL has smaller posterior SDs than SLR; HSL has smaller posterior SDs than CAR in temporal superpixels and larger posterior SDs in nasal superpixels. Figure 5 shows scatter plots of the posterior means from SLR against HSL. There is noticeable shrinkage towards the population mean in the HSL estimates in the peripheral superior and temporal superpixels. The

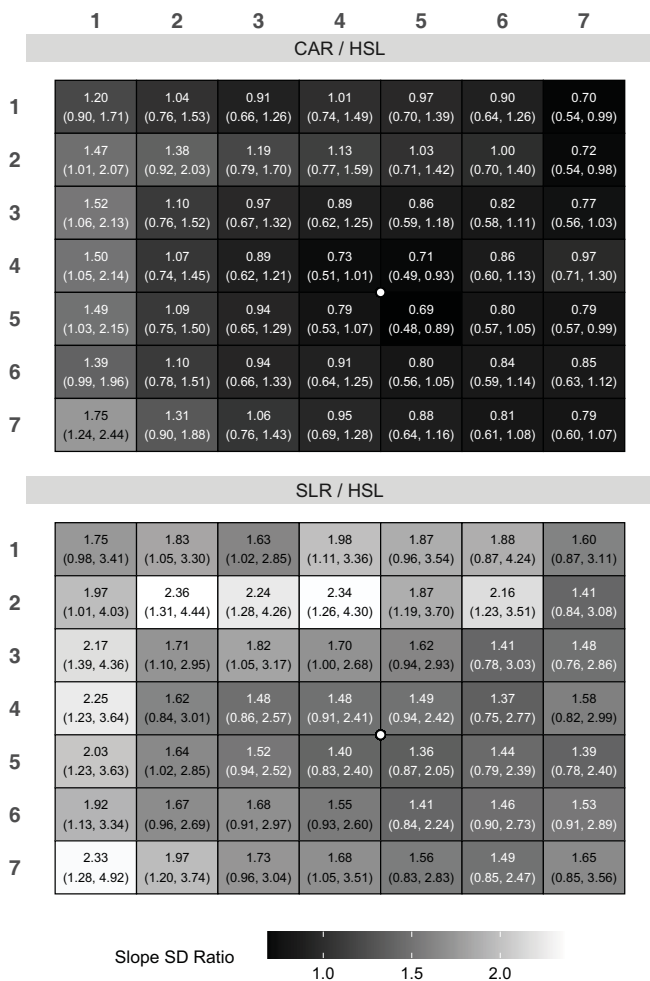


Figure 4. Heat map of the median (tenth, ninetieth percentile) slope SD ratio as a function of superpixel location for the patient cohort data. The ratio was defined as either CAR over HSL or SLR over HSL. Ratios > 1 indicate better performance for HSL. On average, HSL outperforms SLR at all superpixel locations.

scatter plots of the posterior means from CAR against HSL are shown in Supplementary Figure S7, where there is noticeable shrinkage in the HSL estimates in the temporal regions as well. We provide the scatter plots of posterior means from SLR against HSL and from CAR against HSL on the same axes for all 49 superpixels in Supplementary Figures S8 and S9, respectively.

The HSL model identified a higher proportion of significant negative slopes compared to SLR and a lower proportion compared to CAR (HSL = 17.6%; CAR = 26.6%; SLR = 15.6%); it detected a lower proportion of significant positive slopes compared to both CAR and SLR (HSL = 1.2%; CAR = 6.9%; SLR = 4.6%). Supplementary Figure S10 presents the McNemar's test results comparing the proportion of significant negative slopes between HSL and SLR or HSL and CAR, where a higher proportion of signifi-

cant negative slopes was identified in 13/49 superpixels for HSL compared to SLR, 4/49 superpixels for SLR compared to HSL, and 31/49 superpixels for CAR compared to HSL. Supplementary Figure S11 shows the McNemar's test results for comparing the proportion of significant positive slopes, where a lower proportion of significant positive slopes was identified in 21/49 superpixels for HSL compared to SLR and 36/49 for HSL compared to CAR.

Discussion

We examined the ability of SLR, CAR, and HSL to accurately estimate rates of GCC thinning within macular superpixels in a cohort of eyes with central or moderate to advanced glaucoma damage at baseline. Our novel HSL model resulted in lower posterior SD for both intercepts (SLR/HSL median ratio: 1.35) and slopes (SLR/HSL median ratio: 1.69, Table 1) indicating a marked improvement in the certainty in estimated intercepts and rates of change for HSL over SLR. The simulation study showed a significantly higher performance by HSL compared to SLR in terms of detecting actual change; on simulated models, HSL detected 21% of slopes as significantly negative while SLR detected only 13% when the true slopes were negative. A smaller and still significantly better performance was observed with the cohort data. Although CAR also reduced the posterior SD, CAR estimates were more biased by superpixel location and offered substantially reduced 95% CrI coverage probability (mean intercept probability = 0.882; slope probability = 0.810). Based on the simulation study findings, CAR under-reported posterior SDs or overestimated coverage, implying that in the AGPS data analysis, the more frequent declarations of significance are overly optimistic. Therefore, with the advantage of the reduction in noise and appropriate coverage, HSL could detect changes in GCC more efficiently and earlier with relevant clinical and research implications for earlier detection of glaucoma progression.

Simple linear regression is still frequently used to estimate global or local rates of change of structural or functional measures in the field of glaucoma. We show that SLR has numerous weaknesses in this context. Data from the patient cohort are not used to help draw inferences about individual eyes, sectors or superpixels. Spatial correlations are ignored, and correlation of baseline thickness and slopes are not accounted for. Visit effects cannot be accommodated in SLR. These limitations in SLR lead to substantially larger uncertainty in estimating individual patient-superpixel rates



Figure 5. Scatter plots of slope posterior means from simple linear regression model (SLR, y-axis) against those from the hierarchical spatial longitudinal model (HSL, x-axis) in each superpixel for the patient cohort data (right eye format). The SLR posterior means are much more variable than the HSL means. There is noticeable shrinkage towards the population mean in the HSL estimates in the peripheral superior and temporal regions. Each plot is square with its own axes with the x- and y- axes having the same range. The red dashed line represents the $x = y$ diagonal.

of change. In contrast, our HSL model accommodates all these features of the data and thus reduces uncertainty in estimating rates of change.

Our Bayesian hierarchical spatial longitudinal model addresses these shortcomings in SLR by modeling the structure of the data with patients nested in a cohort, spatial correlations across the macula, and

visit effects. The HSL model provides intercept and slope estimates with substantially higher accuracy by incorporating information from the cohort and other superpixels from the same person and reduces the posterior variances for both intercepts and slopes. We implemented this updated version of the HSL model developed in our research laboratory in this study. The

current version allows random variation across superpixels of individual intercepts and slopes from a global patient estimate. Similarly, intercepts and slopes within individual superpixels are allowed to randomly vary from the population intercept and slope. Population level intercepts and slopes vary across the macula and log residual variances are modeled with superpixel and individual patient components. We recently identified spatial distributions of patient-intercepts, patient-slopes, patient-residual variances, and residuals across superpixels through residual analysis of a prior model that did not include spatial effects except for superpixel level parameters.²⁶ The current model in this study improves on the model in Mohammadzadeh et al.²⁶ by including such spatial effects for patient-superpixel intercepts, slopes, and residual variances. Moreover, we now include visit effects, which model correlations in residuals across superpixels for a single patient-visit. These spatial effects are modeled as factors in a factor analysis with one factor per dimension.

Visit effects can be estimated with our HSL model but are impossible to implement in SLR. Visit effects model the correlation in regression residuals across superpixels in a single visit. Presence of visit effects means that all GCC thickness measurements from a single patient-visit tend to be randomly above or below the patient's actual GCC thickness. Estimating visit effects requires multivariate modeling of all superpixel thickness measures in one model. We believe this is the first study to identify visit effects in structural OCT data. Visit effects in visual field data are well known,^{27–29} though modeling of visual field visit effects has been rarely done.³⁰

Our findings indicate a more than twofold reduction in the variance of slope estimates with the HSL model compared to SLR. This significant reduction in posterior variance shows that the HSL model is much more efficient than SLR in using the available data and it allows HSL to detect significant rates of change earlier as compared to SLR. For example, HSL identified a higher proportion of significant negative slopes (17.6% vs. 15.6%) and lower proportion of significant positive slopes (1.2% vs. 4.6%) as compared to SLR. Although these numbers may not seem clinically impressive, the overall superior performance of the HSL is clinically relevant as it not only increases negative hit rates, an indication of higher sensitivity for identifying actual decreasing thicknesses, but it also results in marked reduction of significant positive slopes, an indication of potentially higher specificity as GCC thickness is not expected to increase over time. These findings were confirmed in the simulation study.

There are previous studies that used hierarchical linear models for detection of longitudinal change

in visual fields; our study is unique as it addresses detection of longitudinal changes in macular structural measurements.^{14,16,19,30,31} Montesano et al.¹⁴ developed a Bayesian model for visual field data accounting for the within-eye hierarchical structure, data censoring, and the heteroskedastic variance as a function of the mean threshold sensitivity; they found that time to detect progression was shorter for Bayesian models compared to SLR with permutation analysis of point-wise linear regression.¹⁴

In the article by Betz-Stablein et al.,¹⁹ a conditional autoregressive prior was used to model spatial correlation across the visual field intercepts and slopes within a single person but data from multiple patients were not included in a single model. Intercept and slopes were not modeled as correlated, and conditional on neighboring superpixels, distant superpixels were considered independent. Our CAR model, although inspired by Betz-Stablein et al.,¹⁹ is novel; like in Betz-Stablein et al., our CAR model was fit to data from each patient separately. Although the variances in slope estimates were on par with HSL, there was substantially more bias and reduced 95% CrI coverage probability, translating to a higher proportion of significant slopes identified even when the true slopes were of the opposite sign; thus the CAR model had higher error rates. In contrast, our model also has random patient-superpixel intercepts and slopes, but then allows for finding global patterns of associations in both intercepts and slopes, as we demonstrated in our previous work.²⁶ It also accommodates residual variance varying by both superpixel and patient and allows for visit effects as well.

Based on the simulation study, HSL has substantially lower RMSE than SLR and CAR for both individual patient intercepts and slopes and for superpixel aggregated effects compared to SLR. The SLR model does not have population or superpixel parameters that summarize the population of patients globally or in a superpixel, so any superpixel level inference from SLR is ad hoc; by taking a Bayesian approach with our SLR models, we were able to create an appropriate, if simplistic, inference for SLR results aggregated across patients within a superpixel. For patient-superpixel intercepts and slopes, HSL had lower RMSE in an overwhelming proportion of intercepts (95%) and slopes (96%). Estimates from HSL had thus lower noise as compared to SLR.

Our simulation results demonstrated a small bias in some estimated HSL intercepts and slopes. This phenomenon of shrinkage toward the prior mean is a well-known feature of Bayesian estimators. By borrowing information from the population of patients, the Bayesian models shrink all estimates toward popula-

tion averages. Shrinkage depends on the play between uncertainty in the estimate without population information and the variability in the population. Unusually high or low estimates without shrinkage are typically due to noise in addition to possibly higher or lower underlying true values; these values are shrunk by greater absolute amounts in hierarchical Bayesian models than estimates nearer to the center of the distribution. Posterior means from a hierarchical model are more stable than classical non-Bayesian estimates such as from SLR. This shrinkage mitigates against erroneous high/low slope estimates and, hence, helps prevent making aggressive therapeutic decisions when relying on uncertain and possibly erroneous estimates, pending additional data. The magnitude of the superpixel-averaged slope bias was at most 10% of the average slope across all patients and superpixels ($-0.040 \mu\text{m}/\text{year}$ vs. $-0.39 \mu\text{m}/\text{year}$). Despite this bias, the HSL model was still able to identify a higher proportion of worsening slopes while at the same time minimizing the significant positive slopes, which is desirable within both clinical and research frameworks.

Our model, as with any model, assumes data generated by the model. Because we have normality of residuals built into the model, it does not necessarily model well a distribution with long tails. The obvious outliers are quite extreme, and we tried to remove all the obvious outliers. We believe these are due to measure error. Undoubtedly there are still outliers in the data set. A future extension of our methods may allow for diagnosing outliers via the model itself, or we may develop a novel and improved outlier removal algorithm.

Optical coherence tomography data are somewhat noisy. If clinicians use raw data with outliers, they will not be using the best data possible, and the conclusions could be flawed. Undoubtedly OCT image quality will improve with time and outliers will lessen or disappear. In the meantime, our outlier removal algorithm does not require model fitting; clinicians could use it in clinical practice before trying to fit any model including SLR.

In clinical practice, the HSL model can be fit to longitudinal data from the entire cohort. The estimated rates of change for the cohort of patients provide information on progression rates at the patient-superpixel level in the same cohort. Once posterior samples are obtained, observations at future time points for any patient can be predicted using the posterior predictive distribution. For a new patient from the same population, intercepts and slopes can be predicted using the posterior samples without refitting the model. Periodically, as more data is collected over time, the HSL

model can be refitted to update model parameters and better guide predictions. For clinicians at a new institution, it would be best to refit the model to data from the new institution and then continue along the same path.

The implications of our findings go beyond macular structural measures. The proposed framework can be applied to other structural measures such as retinal nerve fiber layer or neuroretinal rim measurements with modifications for differing geometry and scale of measurements given the fact that all the limitations of SLR apply to those structural measures as well. We are also developing a similar hierarchical model for visual field measurements. The proposed framework is a useful starting point for analysis of visual fields; however, modifications will be required given properties of visual fields such as increased variability with worsening thresholds (heteroscedasticity), censoring, and possible loss of information once threshold sensitivity at individual test locations drops to 15 to 19 dB or below.^{14,32}

In conclusion, we present a novel Bayesian HSL model that improves estimation accuracy of local GCC rates of change. In a simulation study, SLR and CAR have median MSE ratios over the proposed model of 3.3 and 2.4, respectively, for estimating superpixel-patient slopes; in both the simulation study and in the patient cohort data, HSL identifies a higher proportion of deteriorating superpixels when compared to SLR while minimizing positive detection rates. This efficiency is found by more fully utilizing already available information from measurements on a cohort of glaucoma patients and jointly analyzing measurements on all superpixels. Our findings have important implications for improved detection of glaucoma deterioration both clinically and in the research setting.

Acknowledgments

Supported by an NIH R01 grant (R01-EY029792), an unrestricted Departmental Grant from Research to Prevent Blindness, and an unrestricted grant from Heidelberg Engineering.

Presented as a poster at the Annual Meeting of the Association for Research in Vision and Ophthalmology, May 2022, Denver CO.

Disclosure: **E. Su**, None; **V. Mohammadzadeh**, None; **M. Mohammadi**, None; **L. Shi**, None; **S.K. Law**, None; **A.L. Coleman**, Alcon Foundation and Laboratories Thea SAS (2022-2023) (C); **J. Caprioli**, None;

R.E. Weiss, None; **K. Nouri-Mahdavi**, Topcon Healthcare (R), Heidelberg Engineering (F)

References

1. Thompson AC, Jammal AA, Medeiros FA. A review of deep learning for screening, diagnosis, and detection of glaucoma progression. *Transl Vis Sci Technol.* 2020;9(2):1–19.
2. Liu YY, Ishikawa H, Chen M, Wollstein G, Schumman JS, Rehg JM. Longitudinal modeling of glaucoma progression using 2-dimensional continuous-time hidden Markov model. *Med Image Comput Comput Assist Interv.* 2013;16(Pt 2):444–451.
3. Abu SL, Marín-Franch I, Racette L. A framework for assessing glaucoma progression using structural and functional indices jointly. *PLoS One.* 2020;15(7):1–15.
4. Mohammadzadeh V, Su E, Rabiolo A, et al. Ganglion cell complex: the optimal measure for detection of structural progression in the macula. *Am J Ophthalmol.* 2022;237:71–82.
5. Kass MA, Heuer DK, Higginbotham EJ, et al. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol.* 2002;120:701–713; discussion 829–830.
6. 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;120:714–720; discussion 829–830.
7. Rao HL, Kumbar T, Kumar AU, Babu JG, Senthil S, Garudadri CS. Agreement between event-based and trend-based glaucoma progression analyses. *Eye.* 2013;27:803–808.
8. Wu Z, Medeiros FA. Comparison of visual field point-wise event-based and global trend-based analysis for detecting glaucomatous progression. *Transl Vis Sci Technol.* 2018;7(4):20.
9. Rabiolo A, Mohammadzadeh V, Fatehi N, et al. Comparison of rates of progression of macular OCT measures in glaucoma. *Transl Vis Sci Technol.* 2020;9(7):1–14.
10. Mohammadzadeh V, Rabiolo A, Fu Q, et al. Longitudinal macular structure–function relationships in glaucoma. *Ophthalmology.* 2020;127:888–900.
11. Medeiros FA, Zangwill LM, Bowd C, Mansouri K, Weinreb RN. The structure and function relationship in glaucoma: implications for detection of progression and measurement of rates of change. *Invest Ophthalmol Vis Sci.* 2012;53:6939–6946.
12. Mansouri K, Leite MT, Medeiros FA, Leung CK, Weinreb RN. Assessment of rates of structural change in glaucoma using imaging technologies. *Eye.* 2011;25:269–277.
13. Gracitelli CPB, Tatham AJ, Zangwill LM, Weinreb RN, Liu T, Medeiros FA. Estimated rates of retinal ganglion cell loss in glaucomatous eyes with and without optic disc hemorrhages. *PLoS One.* 2014;9(8):e105611.
14. Montesano G, Garway-Heath DF, Ometto G, Crabb DP. Hierarchical censored bayesian analysis of visual field progression. *Transl Vis Sci Technol.* 2021;10(12):1–16.
15. Swaminathan SS, Berchuck SI, Jammal AA, Rao JS, Medeiros FA. Rates of glaucoma progression derived from linear mixed models using varied random effect distributions. *Transl Vis Sci Technol.* 2022;11(2):16.
16. Zhu H, Russell RA, Saunders LJ, Ceccon S, Garway-Heath DF, Crabb DP. Detecting changes in retinal function: analysis with non-stationary Weibull error regression and spatial enhancement (ANSWERS). *PLoS One.* 2014;9(1):1–11.
17. Russell RA, Malik R, Chauhan BC, Crabb DP, Garway-Heath DF. Improved estimates of visual field progression using bayesian linear regression to integrate structural information in patients with ocular hypertension. *Invest Ophthalmol Vis Sci.* 2012;53:2760–2769.
18. Medeiros FA, Leite MT, Zangwill LM, Weinreb RN. Combining structural and functional measurements to improve detection of glaucoma progression using Bayesian hierarchical models. *Invest Ophthalmol Vis Sci.* 2011;52:5794–5803.
19. 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.
20. Mohammadzadeh V, Su E, Heydar Zadeh S, et al. Estimating ganglion cell complex rates of change with bayesian hierarchical models. *Transl Vis Sci Technol.* 2021;10(4):1–13.
21. R Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* 2023.
22. Su YS, Yajima M. R2jags: using R to run ‘JAGS’. R package version 0.7-1. 2021.
23. Kline P. *An Easy Guide to Factor Analysis.* Oxfordshire, UK: Routledge; 2014.

24. de Valpine P, Turek D, Paciorek C, Anderson-Bergman C, Temple Lang D, Bodik R. Programming with models: writing statistical algorithms for general model structures with NIMBLE. *J Comput Graph Stat.* 2017;26:403–413.
25. de Valpine P, Paciorek C, Turek D, et al. NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling. Available at: <https://zenodo.org/records/5562925>.
26. Mohammadzadeh V, Su E, Shi L, et al. Multivariate longitudinal modeling of macular ganglion cell complex: spatiotemporal correlations and patterns of longitudinal change. *Ophthalmol Sci.* 2022;2(3):100187.
27. Phu J, Kalloniatis M. The frontloading fields study (Ffs): detecting changes in mean deviation in glaucoma using multiple visual field tests per clinical visit. *Transl Vis Sci Technol.* 2021;10(13):1–17.
28. Chauhan BC, Garway-Heath DF, Goñi FJ, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol.* 2008;92:569–573.
29. Rui C, Montesano G, Crabb DP, et al. Improving event-based progression analysis in glaucomatous visual fields. *Sci Rep.* 2021;11(1):1–9.
30. Bryan SR, Eilers PHC, Lesaffre EMEH, Lemij HG, Vermeer KA. Global visit effects in pointwise longitudinal modeling of glaucomatous visual fields. *Invest Ophthalmol Vis Sci.* 2015;56:4283–4289.
31. Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci.* 2000;41:417–421.
32. Russell RA, Crabb DP, Malik R, DF Garway-Heath. The relationship between variability and sensitivity in large-scale longitudinal visual field data. *Invest Ophthalmol Vis Sci.* 2012;53:5985–5990.