Improved Estimates of Visual Field Progression Using Bayesian Linear Regression to Integrate Structural Information in Patients with Ocular Hypertension

Richard A. Russell,^{1,2} Rizwan Malik,¹ Balwantray C. Chauhan,³ David P. Crabb,² and David F. Garway-Heath^{1,2}

PURPOSE. To assess whether neuroretinal rim area (RA) measurements of the optic disc could be used to improve the estimate of the rate of change in visual field (VF) mean sensitivity in patients with ocular hypertension (OHT) using a Bayesian linear regression (BLR), compared to a standard ordinary least squares linear regression (OLSLR) of mean sensitivity (MS) measurements alone.

METHODS. MS and RA measurements were analyzed from a longitudinal series of 179 patients with OHT visiting Moorfields Eye Hospital between 1992 and 2000. For each patient, linear regression of RA was computed after an appropriate transformation to “scale” RA with MS measurements, and the slope coefficient from this regression was used as a prior for BLR of MS. The BLR then was compared with the OLSLR approach by evaluating how accurately each regression technique predicted future MS measurements.

RESULTS. On average, BLR was significantly more accurate than OLSLR for series up to 8 measurements long (root-mean-square prediction error [RMSPE] was 0.14 decibels [dB] smaller with BLR than OLSLR; $P < 0.001$, Wilcoxon signed-rank test), with OLSLR of VF data alone being more accurate for longer series (RMSPE was 0.06 dB smaller with OLSLR than BLR).

CONCLUSIONS. BLR provides a significantly more accurate estimate of the rate of change in MS than the standard OLSLR approach, especially in short time series, suggesting that structural measurements can be used successfully in statistical models to assist clinicians monitoring VF progression in patients with OHT. Further studies are necessary to validate

the method in glaucoma patients. (*Invest Ophthalmol Vis Sci.* 2012;53:2760–2769) DOI:10.1167/iovs.11-7976

The detection and accurate quantification of visual field (VF) progression is one of the most important tasks when managing glaucoma. Standard automated perimetry (SAP) remains the gold standard for tracking functional damage in glaucoma, but quantitative imaging measurements also are crucial for identifying and monitoring structural aspects of the disease.^{1–3} At present, glaucomatous eyes generally are identified and monitored using either structural or functional change.^{3–5} Change in both modalities may occur, though not always identified at the same time within the relatively short time span of typical longitudinal studies. One explanation for this dissociation is that VF sensitivity, optic nerve head (ONH) topographic parameters, and retinal nerve fiber layer (RNFL) thickness measurements all are affected by measurement variability,^{6–17} which may confound attempts to identify progression when it is present. Structural changes not linked directly to retinal ganglion cell (RGC) loss, such as conformational changes in the lamina cribrosa or in glial support tissue, and changes in function not linked directly to RGC loss, such as the development of media opacity or RGC dysfunction, are further possible explanations for structure-function dissociation.

Nevertheless, across the range of glaucoma severity, structure and function measurements are associated with healthy eyes having a full VF and normal structure measurements, and eyes with advanced glaucoma having marked VF loss, thinning of the RNFL, and rim loss in the ONH. Statistical methods that formally pool different sources of information into a single model may help in identifying and monitoring glaucoma patients more effectively if functional and structural changes occur as a result of the same biological process that constitutes glaucomatous progression.^{3,4,18} Such models not only are quantitative and objective, but also can reduce the effects of measurement variability. Bayesian approaches are useful for integrating data in this manner, yet there has been little application of Bayesian methodology to integrate data in different domains in glaucoma research.^{4,18,19} Bayesian statistics are based on the notion that *prior* beliefs about the data under investigation should be considered in the statistical analysis to modify *posterior* beliefs. Bayesian analysis specifies this modification statistically using a formula known as *Bayes' theorem*. This theorem expresses the posterior probability of a given hypothesis in terms of the probability of the hypothesis (before integrating new evidence), the prior probability of the data under all possible hypotheses, and the conditional probability of observing the data given the specified hypothesis is true (known as the *likelihood function*).

To date, changes in VF measurements over time have been investigated largely using standard ordinary least squares linear regression (OLSLR) methodology; such analyses have become

From the ¹NIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom; ²Department of Optometry and Visual Science, City University London, United Kingdom, and ³Department of Ophthalmology and Visual Sciences, Dalhousie University, Halifax, Nova Scotia, Canada.

Supported in part by the Department of Health's National Institute for Health Research (NIHR) Biomedical Research Centre for Ophthalmology at Moorfields Eye Hospital NHS Foundation Trust and the UCL Institute of Ophthalmology (RAR, RM and DFG-H). Also supported by an Investigator Initiated Research grant from Allergan, Inc. (RAR), and by funding from the International Glaucoma Association (DFG-H's chair at UCL).

Submitted for publication June 2, 2011; revised August 16 and November 13, 2011, and February 13, 2012; accepted March 16, 2012.

Disclosure: **R.A. Russell**, Allergan (F); **R. Malik**, None; **B.C. Chauhan**, None; **D.P. Crabb**, None; **D.F. Garway-Heath**, None

Corresponding author: David F. Garway-Heath, Moorfields Eye Hospital, 162 City Road, London, EC1V 2PD, UK. david.garway-health@moorfields.nhs.uk.

known as “trend-based.” These analyses examine point-wise VF sensitivity or global VF indices over time, and evaluate the rate (slope) and significance (P value) of change.^{16,20,21} The objective of our study was to assess whether Bayesian normal error linear regression (BLR), which integrates imaging measurements of global neuroretinal rim area (RA), as an informative prior, could be used to detect more accurately a change in perimetric mean sensitivity (MS) over time compared to a standard OLSLR of mean sensitivity measurements alone in patients with ocular hypertension (OHT).

METHODS

Study Sample

A retrospective analysis was carried out on a prospective longitudinal study of patients with OHT attending Moorfields Eye Hospital between 1992 and 2001. Full details of the study sample have been reported previously.²² Briefly, OHT was defined as a previously untreated intraocular pressure >22 mmHg and <35 mmHg on two or more occasions within a 2-week period, and a baseline mean Advanced Glaucoma Intervention Study (AGIS) VF score of 0.²³ Patients had visual acuity at recruitment of 6/12 or better, with no coexistent ocular or neurologic disease. Patients who met the inclusion criteria and gave their informed consent were enrolled. The study adhered to the tenets of the Declaration of Helsinki and had local ethics committee approval.

In our study, the same eye was selected for analysis as had been randomized in the original study; only paired confocal scanning laser tomography (CLST) and VF tests (performed on the same day) were included in analyses.

Visual Field Testing

Visual field testing (24-2 pattern, Full-Threshold program) was carried out with the Humphrey Field Analyzer (Carl Zeiss Meditec, Dublin, CA) at approximately 4-month intervals from the time of recruitment until September 2001. Unreliable VFs were discarded according to the following criteria: $\geq 25\%$ fixation losses, $\geq 30\%$ false negative errors or $\geq 30\%$ false positive errors. The primary outcome measure in our study was decibel (dB) MS, which was calculated as the arithmetic mean. Point-wise sensitivities of all test locations in the 24-2 visual field (excluding blind spot) were first unlogged and then averaged. Finally, MS was calculated as the logarithm (base 10) of this average value.²⁴

CLST Imaging

ONH imaging with the Heidelberg Retina Tomograph “Classic” (Heidelberg Engineering, Heidelberg, Germany) was introduced into the study protocol in 1994. Imaging initially took place annually for the first two years, and subsequently at approximately 4-month intervals until September 2001. Each patient had at least five CLST mean topographies in their series; three single topographies were acquired at each visit. Images with a mean pixel height standard deviation greater than $50\ \mu\text{m}$ were excluded to best reflect clinical practice. Images also were excluded if the contour lines exported from the baseline mean topography could not be aligned satisfactorily in the follow-up mean topography. In total, 8 (from 1467) mean topographies were excluded for this reason. Single topographies were imported as HRTport files into a beta version of Heidelberg Retina Tomograph Explorer Version 3.1.2.0, and RA was calculated using the Moorfields reference plane, which has been shown to improve repeatability and the detection of progression compared to the “standard” and $320\text{-}\mu\text{m}$ reference planes.^{25,26}

Analyses

All statistical analyses were carried out in R.²⁷ R code from the “MCMCregress” function in the “MCMCpack” package was used to carry out BLR.²⁸

Linear Regression Analysis. BLR of each patient’s MS measurements against time were computed using corresponding RA measurements to derive a prior probability distribution (see Appendix for a brief description of Bayesian statistics). RA measurements were transformed first so that a change in RA over time scaled to a change in MS over time (see Appendix for details of transformation applied). Next, a standard OLSLR of transformed RA over time was computed. The slope coefficient from this regression and its standard error then were used as a prior for the slope coefficient in a BLR of corresponding MS measurements.

Comparison of the BLR Approach Against the Standard OLSLR Method. The study hypothesis was that BLR of structure-function measurements would offer a more accurate quantification of glaucomatous functional progression compared to OLSLR of MS measurements alone. Accuracy is difficult to measure in this context due to the absence of a gold-standard for measuring true change in signal (i.e., differential light sensitivity). Thus, to compare the accuracy of the BLR with the OLSLR, an iterative method was developed to examine the prediction error (residual) associated with each method when forecasting the subsequent MS measurement. A linear regression was fitted to each patient’s first 3 MS measurements, and the residual (difference between actual and predicted measurements) corresponding to the patient’s fourth measurement was recorded, then a second linear regression was fitted to the same patient’s VF data using their first 4 MS measurements, and the residual corresponding to the patient’s fifth (subsequent) measurement was recorded, and so on up to the total series length minus 1. For BLR, only the corresponding RA measurements were used in the model. When VFs or HRT images were excluded for not meeting the defined reliability criteria, the corresponding structure or function measurement (RA or MS) also was discarded in the regression models. Finally, the root-mean-square prediction error (RMSPE) for BLR and corresponding OLSLR was calculated for each patient. RMSPE is defined as the square root of the mean of squared prediction errors (over all iterations).

RESULTS

The study included 179 eyes from 179 patients. Table 1 summarizes the characteristics of the study sample at baseline. Figure 1 illustrates the frequency of the total number of paired CLST and VF tests for the 179 patients examined.

Accuracy and Precision of the BLR and OLSLR

Using the iterative procedure outlined above, the prediction errors associated with each linear regression method were compared for all patients. First, prediction errors were stratified according to the number of measurements (series length) inputted into each regression approach, to determine whether the number of measurements affected the accuracy and relative accuracy of the methods to predict the subsequent measurement. Figure 2 indicates the results of this analysis as a series of box plots. As expected, for each regression approach, absolute prediction error decreases with the number of measurements. On average, BLR was significantly more accurate for series of up to 8 VF/CLST pairs (mean difference in RMSPEs, OLSLR minus BLR, was $0.14\ \text{dB}$; $n_1 = n_2 = 179$, $P < 0.001$, Wilcoxon signed-rank test). OLSLR of VF data alone was more accurate than BLR for longer series (mean difference in RMSPEs, OLSLR minus BLR, was $-0.06\ \text{dB}$; $n_1 = n_2 = 179$, $P > 0.1$, Wilcoxon signed-rank test), although the difference in RMSPEs was not significant. While it is not recommended to deduce a trend from as few as three VFs, a Bayesian regression analysis including three MS/RA measurement pairs results in a median prediction error smaller than that from OLSLR of a series of five or six MS measurements; however, as seen in

TABLE 1. Characteristics of Study Sample at Baseline; Median (Inter-Quartile Range)

Measurement	Median (Inter-Quartile Range)
Age (y)	60.0 (53.4, 67.7)
Arithmetic mean sensitivity (dB)	29.4 (28.5, 30.5)
Global rim area (mm ²)	1.19 (1.04, 1.37)
Follow-up period (y)	5.8 (4.8, 6.4)
CLST tests (<i>n</i>)	8 (7, 10)
VF tests (<i>n</i>)	8 (7, 10)

Figure 2, the range of prediction error is greater for BLR of three MS/RA measurements than OLSLR of longer series of MS measurements. BLR of a series of four MS/RA pairs leads to a median absolute prediction error approximately equivalent to that from OLSLR of a series of seven MS measurements, as well as a narrower error range (see dashed lines in Fig. 2).

Considering all prediction errors for each patient, the two-sided Wilcoxon matched pairs signed rank sum test indicated that the RMSPEs resulting from OLSLR were significantly greater than those from BLR (Wilcoxon signed-rank test: $V = 12,556$, $n_1 = n_2 = 179$, $P < 0.001$). This suggests that, overall, BLR better predicts mean sensitivity and, therefore, more accurately estimates the rate of change in MS. Furthermore, the RMSPE for the BLR was smaller than that for the OLSLR in 134 out of 179 patients (74.9%). Figure 3A is a scatterplot (with marginal histograms) of the difference in RMSPE (over all series lengths) for the two regression approaches (BLR minus OLSLR) against the OLSLR slope derived from each patient's full series of MS measurements. Negative values in the Y-axis indicate that the RMSPE for BLR is smaller than the corresponding error for OLSLR; hence, 74.9% of the points lie below the $y = 0$ line. Across all patients, the mean difference between the RMSPEs for the two linear regression approaches was -0.13 dB (shown by the dashed line in Fig. 3A), indicating that, on average, prediction error for OLSR was 0.13 dB greater than for BLR. Furthermore, the mean absolute difference in predicted MS using BLR as opposed to OLSLR was 0.74 dB.

For those patients for whom BLR performed less accurately than OLSLR, the OLSLR RMSPEs were significantly smaller than the corresponding OLSLR RMSPEs in the patients for whom BLR provided a more accurate prediction (Wilcoxon rank sum

FIGURE 1. Histogram of the frequency of total number of paired CLST and VF measurements for 179 patients investigated.

test: $W = 1334$, $n_1 = n_2 = 45$, $P < 0.01$), suggesting that VF variability was much lower in these patients. There was no statistical difference in the mean pixel height standard deviation between these two groups.

In the BLR model, the posterior precision equals the prior precision plus the precision of the likelihood. Hence, the Bayesian confidence intervals for slope (rate of progression) and prediction of MS are inherently smaller than those from OLSLR. This is illustrated in Figure 4, which plots the magnitude of the 95% confidence interval for the slope (Fig. 4A) and the prediction of MS (Fig. 4B) for the BLR and OLSLR methods as a function of the series length.

Linear Regression Examples

Illustrative examples of the BLR and OLSLR are shown in Figures 5 and 6. The top panel of Figure 5 indicates the patient's transformed RA measurements that are paired over time with corresponding MS measurements; an OLSLR is fitted to these data to derive the prior for the BLR. The middle panel illustrates the OLSLR of a patient's MS measurements over time (this regression is shown by the solid line) and the accompanying BLR (indicated by the dashed line). The filled circle indicates the future MS measurement (from which the prediction error or residual was calculated), while the open circles correspond to the MS measurements that were analyzed in each regression approach. Finally, the bottom panel displays the prior, likelihood and posterior probability distributions that are used in the BLR computation.

In the example shown in Figure 5, it is evident that the BLR offers a more accurate prediction of the future MS measurement than the standard OLSLR approach since the residual between the closed circle and the BLR (dashed line) is smaller than the residual between the same point and the associated OLSLR (solid line). In Figure 6, the predictive accuracy of the BLR approach far exceeds the OLSLR when only the first 4 data points are considered in the regression (top panel). However, when the number of measurements is increased to 6 (middle panel) or 8 (bottom panel), the difference in predictive accuracy of BLR compared to OLSLR is decreased markedly compared to the shorter series, so that there is little difference in the two regression approaches.

FIGURE 2. Box plots of the difference in absolute prediction errors between OLSLR and BLR for series lengths from 3 to 11 measurements predicting the mean sensitivity in the $n^{th} + 1$ measurement (i.e., the fourth to 12th visual field). The top and bottom dashed lines equate to the median and upper quartile absolute prediction error for BLR of 4 measurements, respectively.

FIGURE 3. (A) Scatterplot of the OLSLR slope from each patient's full series of MS measurements against the difference in RMSPE (over all series lengths) for the two regression approaches (with marginal histograms). The mean difference between the RMSPEs for the two linear regression approaches is shown by the *dashed line*. (B) Scatterplot of rates of change in MS derived from OLSLR against corresponding rate resulting from BLR (with marginal histograms). The *solid diagonal line* represents the line of equality.

DISCUSSION

In recent years, there has been a large interest in the application of Bayesian statistics in science, medical, and engineering disciplines.^{29–31} This has been intensified by the growing availability of Bayesian statistical software and the speed of modern computers. Bayesian approaches offer the opportunity to add more variables to an inference model. In monitoring glaucomatous progression, clinicians are interested in a range of clinical measurements. New research evaluated a Bayesian

FIGURE 4. (A) Box plots of the magnitude of the 95% confidence interval (CI) for the slope for the OLSLR and BLR methods for series lengths from 3 to 11 measurements. (B) Box plots of the magnitude of the 95% CI for the prediction of mean sensitivity in the $n^{th} + 1$ measurement (i.e., the fourth to 12th visual fields) for the OLSLR and BLR methods.

hierarchical model for combining structural (average retinal nerve fiber layer thickness measurements from scanning laser polarimetry) and functional (visual field index [VFI] measurements from SAP) information, and found that the method improved detection of glaucoma progression compared with standard OLSLR of VFI measurements alone.⁴ The Bayesian hierarchical model is based a cohort of subjects and uses information from the whole sample to update the posterior estimate in each patient. Conversely, in our study, the BLR approach is not based on a hierarchical model and, instead, uses each patient's prior (structural) information separately. The BLR method was compared with the OLSLR approach, and the results suggested that visual field progression rates are more estimated accurately with the Bayesian method.

Until recently, VF trend-analyses have used standard OLSLR methodology. OLSLR is a frequentist approach to statistics; significantly, frequentist statistics assume there are sufficient measurements to infer something meaningful or, in other words, that signal is distinguishable from noise in the data. In circumstances where variability is pronounced and there is prior information to add to the model, Bayesian statistics offer a particularly apt alternative to frequentist inference. In the follow-up of glaucoma patients, there often are insufficient VF

FIGURE 5. *Top panel:* Transformed-RA measurements over time and fitted OLSLR for a given OHT patient. *Middle panel:* Corresponding BLR and OLSLR of three MS measurements (*open circles*) over time. The BLR is indicated by the *dashed line* while the OLSLR is shown by the *solid line*. The *filled circle* indicates the future MS measurement not examined in each linear regression approach. *Bottom panel:* Prior, likelihood, and posterior probability distributions contributing the BLR shown in top panel. In this example the prior distribution has a large effect on the posterior distribution, since RA measurements display a strong relationship with time, while corresponding MS measurements display a much weaker relationship, hence the likelihood distribution is concentrated weakly and the posterior distribution is dominated by the prior.

measurements to estimate progression accurately; this is due to the imprecision (large variability) of these measurements. Moreover, there are substantial difficulties in obtaining an adequate amount of VF data to identify glaucomatous progression that are clinically significant. These obstacles include financial resources, physician patient loads and patients' time.³² Supplementing perimetry with imaging measurements allows clinicians to explore a different manifestation of progression, and is relatively quick to carry out. Images can be acquired on the same occasion as a VF test, whereas it is difficult in clinical practice to undertake two VF tests on a single visit; in addition, imaging generally is preferred by patients.³³

A further difficulty for clinicians in routine glaucoma management is interpreting the considerable volume of data generated by SAP and imaging devices.³² There is a significant, growing need for clinically relevant automated, objective tools to define progression based jointly on structural and functional information together. At present, analysis software tends to judge structure and function test data separately in a binary "progression present" or "absent" manner. The approach outlined in this study provides an objective, quantitative framework for integrating structural and functional measurements that gives an output in the domain of greatest interest—the rate of visual field loss.

To derive a prior for BLR, it was necessary to apply a transformation to RA. In our study, it is assumed that RA declines linearly over time in linear (mm²) units.⁵ However, studies of structure-function associations almost invariably report a non-linear relationship between visual function scaled in decibels and linearly scaled structural parameters.³⁴ Thus, we investigated transforming RA in linear units into dB units, since the associations become more linear when structure and function are scaled in the same units.^{34–36} Furthermore, the relationship between visual field sensitivity and ganglion cell number is linear when both are expressed in decibels.³⁷ Interestingly, the results largely were unchanged, which suggests that scaling is less important over the short stage of the disease process examined in our study. It also is assumed that MS declines linearly over time; linear regression models have proved superior and practical compared to other more complex models for predicting VF sensitivity in time series.³⁸

MS was chosen as a global index for progression over mean deviation (MD), since the latter is derived from age-corrected normal sensitivity and is weighted according to eccentricity,¹⁶ while RA (subsequently used to derive a prior for Bayesian analysis) is not age-corrected or weighted. Furthermore, research into structure-function models suggests arithmetic MS approximates better to retinal ganglion cell counts than indices based on a geometric mean, such as MD.²⁴ This hypothesis is supported by analyzing the correlation between MD and RA for the data under investigation, which indicates that there is no significant correlation between these two variables (Kendall's τ test: $\tau = 0.004$, $Z = 0.257$, $n_1 = n_2 = 1467$, $P = 0.797$), while MS and RA do correlate significantly (see Appendix). In addition, recent research suggests that MS may predict better future glaucomatous functional change compared with MD, and the authors suggest that this may be due to imperfect age-correction implicit in the calculation of MD.^{39,40}

Care was taken in our study to avoid sources of variability in CLST measurements. Firstly, contour lines were drawn by a single expert, and secondly, the Moorfields reference plane was used to improve rim area repeatability and the ability to detect structural progression.²⁵ Nevertheless, CLST rim area measurements are subject to errors when used to track glaucomatous progression. Errors include the incorporation of blood vessels, and other structures that may be unrelated to axon loss, into the RA measurement. Pulsations of retinal

FIGURE 6. (A1 and A2) OLSLR (solid lines) and BLR (dashed line) of a given OHT patient's first four RA and MS measurements, respectively. (B1 and B2) OLSLR and BLR of the patient's first six RA and MS measurements. (C1 and C2) OLSLR and BLR of the patient's first eight RA and MS measurements. Filled circles indicate the future MS measurement not examined in each linear regression approach.

arteries in the optic nerve head also can produce variability in CLST stereoscopic measurements.^{41,42} An interesting extension to the current study would be to use other structural measurements to derive a prior for BLR that are likely to be

related better to VF MS, such as RNFL thickness. Furthermore, patients often exhibit focal, rather than diffuse rim loss during the course of progressive disease. There is an unfortunate trade-off between the detection of shallow localized defects

and variability. Global indices, such as MS and RA, have less variability than point-wise or sectoral measurements, but are more likely to overlook focal damage. Conversely, point-wise analyses may detect better subtle abnormalities compared with global indices, but are associated with greater variability. It is possible that structural information about focal damage could be used as a prior to reduce the large variability associated with functional analyses of localized defects, and overcome this trade-off. It remains to be ascertained whether the BLR approach can be applied successfully to RA sectors (or sectorial RNFL thickness measurements) as well as global structural measures; however, a preliminary analysis was performed using BLR on a sector-by-sector basis, and the results were very encouraging suggesting that the BLR approach offered similar improvements in accuracy and precision compared to the global approach.

A limitation of our study was that, at baseline, the patient sample consisted entirely of OHT patients. BLR could be used to predict progression in patients with established glaucoma, although further evaluation of the methodology in a longitudinal data set of patients with glaucoma is required. However, in contrast to patients with OHT, patients with glaucoma have lower MS measurements and, therefore, higher variability.¹⁷ Consequently, integrating structural information may prove even more valuable in patients with glaucoma. Moreover, patients with OHT represent a very important demographic, comprising a large proportion of patients in glaucoma clinics. Thus, methods to identify patients with OHT who have glaucoma as accurately as possible are understandably very important. The fewer number of visits, for equivalent prediction accuracy and precision, required for joint structure and function analysis would relieve a significant part of this burden of care.

The RMSPE for the BLR was smaller than that for the OLSLR in 75% of patients. Therefore, in 25% of patients, the OLSLR provided a better estimate of the rate of progression. This could be due to inherent variability in the measurements. Moreover, the summed OLS prediction errors were significantly smaller in these patients than in the 75% of patients in whom BLR was more accurate, indicating that VF measurements were inherently less variable in these patients. Consequently, the usefulness of BLR was reduced. Importantly, Figure 3A suggests that the magnitude of RMSPE for BLR is not affected by the rate of change in MS (since the points are scattered homogeneously over the range of OLSLR slopes). This implies that the RA prior is informative across the range of rates of change in MS. It should be noted that RMSPE is useful only for evaluating the accuracy of the two methods and is not informative about the difference in prediction of MS and the derived rates of progression. This is apparent in Figure 3B, which plots the rates of change in MS for BLR against the corresponding rate for OLSLR (for all patients and all measurements); the figure illustrates considerable disagreement between the derived rates. From a clinical perspective, the rate of change in VF sensitivity is more important than predicting a future VF measurement when managing an individual patient. Indeed, the rates of change in MS for BLR and OLSLR may be very different. For example, in Figure 5 the absolute prediction error for the OLSLR is not much greater than that for the BLR, yet the derived rates of change in MS are very different (OLSLR slope 0.31 dB/year, BLR slope -0.62 dB/year).

On average, BLR was more accurate for series of up to about 8 VF/CLST pairs, with OLSLR of VF data alone being more accurate for longer series. This has important implications for the potential use of BLR in clinical practice. In early glaucoma monitoring, there are insufficient VF measurements to estimate accurately the rate of progression using OLSLR due to the high variability of VF measurements, especially when learning

effects are prominent. In such circumstances, BLR is significantly more accurate, on average, than OLSLR for estimating change in MS. Improved detection of VF progression arguably is most important in early testing since clinicians' decisions on subsequent monitoring will be affected greatly by initial test results. For example, in Figure 5, OLSLR of the patient's three VF measurements suggests that the patient's MS is improving, whereas BLR suggests a considerable decline in MS. In such instances, it would be expected that this patient's monitoring frequency (and perhaps even treatment) would change to reflect this.

When the number of MS measurements is increased, the performance of BLR is reduced gradually compared to OLSLR. This may be due to the absence of a one-to-one linear relationship in MS and transformed RA measurements, or simply that a sufficient number of VF measurements are available to mitigate the "noise" in the measurements so that the trend in MS loss is estimated more accurately (for example, see Fig. 6). This hypothesis is supported by Figure 4, which indicates that the precision of the OLSLR method is very low in early testing but similar to the BLR method for series length above seven measurements.

In our study, the modeled linear relationship between transformed RA and MS is based on the sample average. However, the relationship between structural and functional progression may deviate from this average for an individual patient or over the course of time. Therefore, if change occurs more quickly in one domain (structural or functional) compared with the average, or the rate of progression varies with disease severity, then the accuracy of the BLR will be reduced. In early testing, the assumption that individual structure/function relations conform to the average appears acceptable, since functional variability is very large, and the prior structure information helps to overcome this. However, when more perimetric measurements are observed, the inherent accuracy and precision of OLSLR increases, and so the prior becomes less useful. Studies suggest that structural variability may be as large as functional variability, if not larger.⁴³ However, a great strength of the BLR method is that the Bayesian posterior precision is equal to the prior precision plus the precision of the likelihood. In early testing, the Bayesian confidence intervals for the slope and prediction of MS are markedly smaller than the corresponding confidence intervals derived from OLSLR of the VF data alone (see Fig. 4).

Over all iterations for all patients (930 linear regressions), BLR resulted in 60 significantly negative slopes (rate less than -0.25 dB/year, $P < 0.05$) from 27 patients compared to 71 slopes from 26 patients using OLSLR. Moreover, BLR resulted in only 8 significantly positive slopes from 7 patients compared to 29 slopes from 13 patients using OLSLR (rate greater than 0.25 dB/year, $P < 0.05$). In total, 8 patients were flagged as having progression using both methods. In these patients, on average, BLR flagged progression 0.45 years earlier than OLSLR. In those 19 patients in whom only BLR was significant, the mean time to flag progression was 2.5 years, while the corresponding time for OLSLR (15 patients) was 4.1 years. These results suggest that BLR detects glaucomatous progression sooner and with fewer false positives than OLSLR. Examining a rate of loss greater than 0.5 dB/year, BLR resulted in 14 significantly negative slopes from 11 patients compared to 49 slopes from 17 patients using OLSLR. However, the greater number of progressing outcomes using OLSLR may be a result of false positives, and this is supported by examining the number of significant positive slopes from the two methods; BLR resulted in only 4 significantly positive slopes from 3 patients compared to 20 slopes from 8 patients using OLSLR.

Our study suggests that it is useful to integrate structural and functional measurements from glaucoma patients into a

single model for progression. It is unsurprising that, with reliable prior information, the Bayesian regression model provides a more accurate prediction of change in MS than the standard OLSLR approach. Furthermore, the accuracy of such a model may be improved greatly by using a different structural measurement that relates more closely to function. Evidently, with unreliable prior information (or structure-function dissociation) BLR may not be appropriate. The results of our and another recent study,⁴ provide compelling evidence that structural measurements are informative about visual function, and moreover, that integrated structure-function statistical models can be useful tools to assist clinicians monitoring glaucomatous progression. Further research is necessary to develop such statistical methodologies for application in the clinic. However, preliminary results demonstrate clearly the feasibility and usefulness of these models.

APPENDIX

LINEAR REGRESSION ANALYSIS

For the interested reader, the BLR model used in our study has been described previously.²⁸ In addition, the R²⁷ code to carry out BLR, as performed in our study, is available freely online as the 'MCMCregress' function in the 'MCMCpack' package;²⁸ default settings for the function were used unless otherwise specified.

In a linear model, the relationship between two variables is given by:

$$y_i = \beta x_i + e_i$$

where y is the response (outcome) variable, x is the predictor variable, β is a vector of the intercept term and the slope coefficient, e is the error term and i refers to the i^{th} observation in a sample of n observations ($i = 1, 2, \dots, n$). In this model, the error term stipulates that the relationship between the response variable and the predictor variable is not exact, and generally is assumed to be Gaussian distributed with mean zero and constant variance (σ^2):

$$e_i \sim \text{Normal}(0, \sigma^2)$$

In a Bayesian analysis of a linear model, priors must be specified for β and e_i . In our study, the BLR model was implemented using semi-conjugate priors; a multivariate Gaussian prior was used for the intercept and slope coefficients and an inverse Gamma prior was used on the conditional error variance:

$$\beta \sim \text{Normal}(b_0, B_0^{-1})$$

$$\sigma^2 \sim \text{InverseGamma}\left(\frac{c_0}{2}, \frac{d_0}{2}\right)$$

where b_0 is the vector of prior means and B_0^{-1} is the prior precision of the multivariate Gaussian prior on β , while c_0 is the shape parameter and d_0 is the scale parameter for the inverse Gamma prior on σ^2 . In our study, the mean and precision of the prior for the slope coefficient were derived from OLSLR of the patient's corresponding RA measurements over time. A non-informative prior was used for the intercept term (precision equal to zero). This assigns equal likelihood to all possible values of the intercept, and so has no impact on the posterior. The prior on the conditional error variance was specified using standard non-informative parameters – $c_0 = 0.001$ and $d_0 = 0.001$.

The Markov Chain Monte Carlo (MCMC) algorithm known as Gibbs sampling was implemented to approximate the joint

posterior distribution. To ensure convergence, 200,000 Gibbs draws were taken with an initial 50,000 burn-in draws and a thinning interval of five (i.e., only every fifth draw was saved). Before making inference on the posterior density sample, diagnostics were performed to assess whether the MCMC converged to a stationary distribution. Convergence was checked using a trace plot and a density plot for each parameter; these revealed no evidence for non-stationarity (indicating that the burn-in period was sufficient), and suggested that the posterior distributions of the model coefficients were normal. In addition, autocorrelation plots implied that there was no autocorrelation.

In the BLR model, the posterior density is a weighted average of the prior and likelihood, and will place greater weight on the prior when its precision is high.⁴⁴ Conversely, if the likelihood probability distribution is strong, that is the observed dependent variable (MS) displays a strong relationship with the independent variable (follow-up time) or the prior probability distribution is concentrated weakly, then the posterior probability distribution is largely subjugated by the likelihood. This is particularly pertinent to the application of BLR for monitoring glaucoma patients, since structure-function measurements are highly variable, and progression may be characterized better by one or the other quantity — whichever has least variability.

TRANSFORMATION OF RA

RA measurements were transformed to "scale" with MS measurements. The slope coefficient from "Passing and Bablok" linear regression (PBLR) of MS against RA (from all measurements of the 179 patients examined) provided this scaling factor. PBLR is a non-parametric procedure based on the rank principle that holds no special assumptions regarding the distribution of the samples or their measurement errors.⁴⁵ The procedure requires that there is a significant, positive correlation using Kendall's τ test between the two variables, which we demonstrate for the two variables here (Kendall's τ test: $\tau = 0.12$, $Z = 6.79$, $n_1 = n_2 = 1544$, $P < 0.001$).

Figure A1 illustrates the results of PBLR of MS against RA measurements (dashed line). Conventional OLSLR is inappro-

FIGURE A1. RA measurements (X-axis) were transformed to scale with MS measurements (Y-axis) using the slope coefficient from PBLR (dashed line). The slope coefficient from PBLR was 6.06, indicating that RA measurements should be multiplied by this value to scale suitably with MS measurements. Darker points in the plot indicate overlapping data.

priate to derive a scaling factor, since it assumes an error-free predictor variable and normally-distributed error terms in the response variable, both of which are untrue in this instance. The slope coefficient from PBLR was 6.06 (95% confidence interval: 5.71–6.44), indicating that RA measurements should be multiplied further by this value to scale suitably with MS measurements.

References

- Weinreb RN, Khaw PT. Primary open-angle glaucoma. *Lancet*. 2004;363:1711–1720.
- Garway-Heath DE. Early diagnosis in glaucoma. *Prog Brain Res*. 2008;173:47–57.
- Xin D, Greenstein VC, Ritch R, Liebmann JM, De Moraes CG, Hood DC. A comparison of functional and structural measures for identifying progression of glaucoma. *Invest Ophthalmol Vis Sci*. 2011;52:519–526.
- Artes PH, Chauhan BC. Longitudinal changes in the visual field and optic disc in glaucoma. *Prog Retin Eye Res*. 2005;24:333–354.
- Strouthidis NG, Scott A, Peter NM, Garway-Heath DE. Optic disc and visual field progression in ocular hypertensive subjects: detection rates, specificity, and agreement. *Invest Ophthalmol Vis Sci*. 2006;47:2904–2910.
- Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from full threshold, SITA standard, and SITA fast strategies. *Invest Ophthalmol Vis Sci*. 2002;43:2654–2659.
- Tan JC, Garway-Heath DE, Fitzke FW, Hitchings RA. Reasons for rim area variability in scanning laser tomography. *Invest Ophthalmol Vis Sci*. 2003;44:1126–1131.
- Budenz DL, Fredette MJ, Feuer WJ, Anderson DR. Reproducibility of peripapillary retinal nerve fiber thickness measurements with stratus OCT in glaucomatous eyes. *Ophthalmology*. 2008;115:661–666.
- DeLeón Ortega JE, Sakata LM, Kakati B, et al. Effect of glaucomatous damage on repeatability of confocal scanning laser ophthalmoscope, scanning laser polarimetry, and optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2007;48:1156–1163.
- Strouthidis NG, White ET, Owen VM, Ho TA, Hammond CJ, Garway-Heath DE. Factors affecting the test-retest variability of Heidelberg retina tomograph and Heidelberg retina tomograph II measurements. *Br J Ophthalmol*. 2005;89:1427–1432.
- Owen VM, Strouthidis NG, Garway-Heath DE, Crabb DP. Measurement variability in Heidelberg retina tomograph imaging of neuroretinal rim area. *Invest Ophthalmol Vis Sci*. 2006;47:5322–5330.
- Lusky M, Boscem ME, Weinreb RN. Reproducibility of optic nerve head topography measurements in eyes with undilated pupils. *J Glaucoma*. 1993;2:104–109.
- Chauhan BC, LeBlanc RP, McCormick TA, Rogers JB. Test-retest variability of topographic measurements with confocal scanning laser tomography in patients with glaucoma and control subjects. *Am J Ophthalmol*. 1994;118:9–15.
- Mikelberg FS, Wijsman K, Schulzer M. Reproducibility of topographic parameters obtained with the Heidelberg retina tomograph. *J Glaucoma*. 1993;2:101–103.
- Piltz JR, Starita RJ. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol*. 1990;109:109–111.
- Heijl A, Lindgren G, Olsson J. A package for the statistical analysis of visual fields. *Doc Ophthalmol Proc Ser*. 1987;49:153–168.
- 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.
- 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.
- Tucker A, Vinciotti V, Liu X, Garway-Heath D. A spatio-temporal Bayesian network classifier for understanding visual field deterioration. *Artif Intell Med*. 2005;34:163–177.
- Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression. *Am J Ophthalmol*. 2008;145:343–353.
- Viswanathan AC, Fitzke FW, Hitchings RA. Early detection of visual field progression in glaucoma: a comparison of PROGRESSOR and STATPAC 2. *Br J Ophthalmol*. 1997;81:1037–1042.
- Kamal D, Garway-Heath D, Ruben S, et al. Results of the betaxolol versus placebo treatment trial in ocular hypertension. *Graefes Arch Clin Exp Ophthalmol*. 2003;241:196–203.
- Advanced Glaucoma Intervention Study. 2. Visual field test scoring and reliability. *Ophthalmology*. 1994; 101:1445–1455.
- Swanson WH, Feliuss J, Pan F. Perimetric defects and ganglion cell damage: interpreting linear relations using a two-stage neural model. *Invest Ophthalmol Vis Sci*. 2004;45:466–472.
- Asaoka R, Strouthidis NG, Kappou V, Gardiner SK, Garway-Heath DE. HRT-3 Moorfields reference plane: effect on rim area repeatability and identification of progression. *Br J Ophthalmol*. 2009;93:1510–1513.
- Poli A, Strouthidis NG, Ho TA, Garway-Heath DE. Analysis of HRT images: comparison of reference planes. *Invest Ophthalmol Vis Sci*. 2008;49:3970–3975.
- Team. *RDC. R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2010.
- Martin AD, Quinn KM, Park JH. MCMCpack: Markov Chain Monte Carlo in R. *J Stat Software*. 2011;42:1–21.
- Whyte S, Walsh C, Chilcott J. bayesian calibration of a natural history model with application to a population model for colorectal cancer. *Med Decis Making*. 2010;31:625–641.
- Holland DJ, Blake A, Tayler AB, Sederman AJ, Gladden LFA. Bayesian approach to characterising multi-phase flows using magnetic resonance: application to bubble flows. *J Magn Reson*. 2010;209:83–87.
- López-Cruz PL, Bielza C, Larrañaga P, Benavides-Piccione R, Defelipe J. Models and simulation of 3D neuronal dendritic trees using bayesian networks. *Neuroinformatics*. 2011;9:347–369.
- Chauhan BC, Garway-Heath DE, Goñi FJ, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol*. 2008;92:569–573.
- Gardiner SK, Demirel S. Assessment of patient opinions of different clinical tests used in the management of glaucoma. *Ophthalmology*. 2008;115:2127–2131.
- Reus NJ, Lemij HG. The relationship between standard automated perimetry and GDx VCC measurements. *Invest Ophthalmol Vis Sci*. 2004;45:840–845.
- Garway-Heath DE, Holder GE, Fitzke FW, Hitchings RA. Relationship between electrophysiological, psychophysical, and anatomical measurements in glaucoma. *Invest Ophthalmol Vis Sci*. 2002;43:2213–2220.
- Schlottmann PG, De Cilla S, Greenfield DS, Caprioli J, Garway-Heath DE. Relationship between visual field sensitivity and retinal nerve fiber layer thickness as measured by scanning laser polarimetry. *Invest Ophthalmol Vis Sci*. 2004;45:1823–1829.
- Harwerth RS, Carter-Dawson L, Smith EL 3rd, Barnes G, Holt WF, Crawford ML. Neural losses correlated with visual losses in

- clinical perimetry. *Invest Ophthalmol Vis Sci.* 2004;45:3152-3160.
38. McNaught AI, Crabb DP, Fitzke FW, Hitchings RA. Modelling series of visual fields to detect progression in normal-tension glaucoma. *Graefes Arch Clin Exp Ophthalmol.* 1995;233:750-755.
  39. Gardiner SK, Demirel S, Johnson CA. Perimetric indices as predictors of future glaucomatous functional change. *Optom Vis Sci.* 2011;88:56-62.
  40. Gardiner SK, Demirel S, Johnson CA, Swanson WH. Assessment of linear-scale indices for perimetry in terms of progression in early glaucoma. *Vision Res.* 2011;51:1801-1810.
  41. Chauhan BC, McCormick TA. Effect of the cardiac cycle on topographic measurements using confocal scanning laser tomography. *Graefes Arch Clin Exp Ophthalmol.* 1995;233:568-572.
  42. Stein DM, Wollstein G, Schuman JS. Imaging in glaucoma. *Ophthalmol Clin North Am.* 2004;17:33-52.
  43. Jampel HD, Vitale S, Ding Y, Quigley H, Friedman D, Congdon N, et al. Test-retest variability in structural and functional parameters of glaucoma damage in the glaucoma imaging longitudinal study. *J Glaucoma.* 2006;15:152-157.
  44. Bolstad WM. *Introduction to Bayesian statistics.* 2nd ed. Hoboken, N.J.: John Wiley; 2007.
  45. Bablok W, Passing H, Bender R, Schneider B. A general regression procedure for method transformation. Application of linear regression procedures for method comparison studies in clinical chemistry, Part III. *J Clin Chem Clin Biochem.* 1988;26:783-790.