Bayesian Hierarchical Modeling of Longitudinal Glaucomatous Visual Fields using a Two-Stage Approach

S. R. Bryan[a,b], P.H.C. Eilers[a], B. Li[a], D. Rizopoulos[a], K.A. Vermeer[b],

H.G. Lemij[c] and E.M.E.H. Lesaffre[a,d]

a Department of Biostatistics, Erasmus MC, Rotterdam, the Netherlands

b Rotterdam Ophthalmic Institute, Rotterdam, The Netherlands

c Glaucoma Service, Rotterdam Eye Hospital, Rotterdam, The Netherlands

d L-Biostat, KU Leuven, Leuven, Belgium

Corresponding author: Susan Bryan, Department of Biostatistics, Erasmus MC, PO Box

2040, 3000 CA Rotterdam, The Netherlands

email: s.bryan@eramsusmc.nl, Tel: +31/(0)10/7044231, Fax: +31/(0)10/7043014

Abstract

The Bayesian approach has become increasingly popular because it allows to model quite complex models via Markov chain Monte Carlo (MCMC) sampling. However, it is also recognized nowadays that MCMC sampling can become computationally prohibitive when a complex model needs to be fit to a large data set. To overcome this problem, we applied and extended a recently proposed two-stage approach to model a complex hierarchical data structure of glaucoma patients who participate in an ongoing Dutch study. Glaucoma is one of the leading causes of blindness in the world. In order to detect deterioration at an early stage, a model for predicting visual fields (VF) in time is needed. Hence, the true underlying VF progression can be determined, and treatment strategies can then be optimized to prevent further VF loss. Since we were unable to fit these data with the classical one-stage approach upon which the current popular Bayesian software is based, we made use of the two-stage Bayesian approach. The considered hierarchical longitudinal model involves estimating a large number of random effects and deals with censoring and high measurement variability. In addition, we extended the approach with tools for model evaluation

KEY WORDS: Bayesian modeling, Hierarchical structure, Longitudinal data analysis, Two-stage approach.

1 Introduction

Since the introduction of MCMC sampling by Gelfand and Smith [8] and the development of the BUGS software [19] the Bayesian approach has become tremendously popular in various application areas, but especially to fit models to complex data structures. But with the years it also became clear that MCMC sampling can be computationally quite cumbersome, and even prohibitive, for fitting complex models to relatively large data sets. Several attempts have been made to look for alternative computational procedures and software, with notable examples such as INLA [24] and STAN [11]. While this newly software can sometimes speed up the computations considerably, the computational gain is not always obvious upfront and for some advanced models the new developments may not be suitable yet. In addition, the majority of the practical Bayesians still use BUGS-related software. In this context, Lunn et al. [18] proposed to fit a hierarchical model in two stages. The authors claim more model flexibility in this way, but advocate the use of their procedure especially for its computational properties. In this paper we further illustrate the use of the two-stage approach on a far more complex hierarchical data structure of glaucoma patients. In addition, we extended the approach with an additional sampling step to allow for the calculation of model selection and model evaluation criteria.

Our modeling approach is motivated by data from the Glaucoma Study conducted by the Rotterdam Eye Hospital in the Netherlands. According to the World Health Organization (WHO), glaucoma is one of the leading causes of irreversible blindness in the world [14]. Adequate treatment may slow down the disease, possibly even halting its progression. Evaluation of a longitudinal series of visual fields (VF), as measured by standard automated perimetry (SAP), provides a way to detect early evidence of glaucoma and to determine functional deterioration. However, due to the subjective nature of this technique, SAP is prone to large variability. In order to measure the true progression of the disease, this variability needs to be taken into account. The Glaucoma Study provides a unique database with a long follow up time. Although methods may have existed to model this type of data, the difficulties in extracting it from the device has made this type of data rare and hence has prevented much research on the topic.

The response variable of interest are the sensitivity estimates which describe the level of differential light sensitivity at different locations within each eye. The sensitivity estimates are left-censored due to the limitation of the device. Models which take into account

this type of censoring, such as the Tobit model, have been described in the literature [28]. Our interest lies in modeling the latent, true values rather than the observed sensitivity estimates for two reasons. Firstly, clinical interest lies in predicting the disease progression rather than the observed sensitivity estimates. Secondly, using the latent scale allows us to use a simpler model than when directly modeling the observed data. The hierarchical structure of the data consists of 4 levels, namely, (1) the individual, (2) the eye (3) the hemifield and (4) the location. There is a vast amount of literature that addresses hierarchical mixed effects models, for both frequentist [29] and Bayesian [17, 21] approaches. We model this complex data structure using a Bayesian hierarchical mixed effects model with cross-classified random effects. Hence, we combine both spatial and time effects. One of the difficulties in modeling VF data is the amount and type of measurement error or variability in the sensitivity estimates. This may be due to measurable factors, such as season, time of day and reliability indices, or unknown transient factors, such as fatigue, lack of concentration, or delayed reaction time. Although their magnitudes may vary, these factors affect all locations belonging to the same VF. We propose to model them as Global Visit Effects (GVEs). Furthermore, there is an inverse relationship between sensitivity and variability. For example, measurement error in the VFs increases with damage, and hence low sensitivity estimates have high variability. Therefore, it is naive to assume a constant variance over the wide range of sensitivity estimates. In this paper, we relax this assumption in order to incorporate this relationship. A problem with high dimensional data and complex data structures, is that it is sometimes difficult or even impossible to model them with standard MCMC algorithms. Lunn et al. [18] proposed a two-stage approach, which allowed us to simplify the problem while still benefiting from the advantages of a full Bayesian model. However, one of the disadvantages of this approach, is that it is not possible to directly obtain the random effects estimates needed for most model evaluations. We address this issue by extending the two-stage approach to be able to determine these estimates.

Our aim is to model this complex data structure in order to obtain better estimates of the true evolution of the sensitivity over time, so that treatment strategies can be optimized to prevent further progression of VF loss. The structure of the paper is as follows. In Section 2 we give further details on the motivating data set and introduce the research questions that triggered our modeling approach(es). In Section 3 we describe the

models used in the analysis. In the subsequent section we briefly review computational aspects of the analysis. Model comparison is dealt with in Section 5. In Section 6 we apply our models to the Glaucoma Study data. Section 7 contains a concluding discussion. Further details regarding the modeling approach are provided in an appendix.

2 Motivating data set: the Glaucoma Study

2.1 Description of the project

The Glaucoma Study is a prospective cohort study conducted by the Rotterdam Eye Hospital in the Netherlands. This is an ongoing study which began in 1998. Inclusion criteria included glaucoma diagnosis and an age range of 18 to 85 years. In total, 139 patients, consisting of 80 (57.6%) men and 59 (42.4%) women, were recruited with a mean follow-up of 10.5 years. Follow-up data were collected at approximately 6-monthly intervals. All patients gave their written informed consent for participation. All research procedures followed the tenets set forth in the Declaration of Helsinki. Furthermore, all of the data that was used in this analysis has been made available online at http://rod-rep.com.

Sensitivity estimates were measured at 52 test locations within each eye, or 26 test locations within each hemifield (excluding two locations corresponding to the blind spot) as shown in Figure 1. The VFs were tested using the Humphrey Field Analyzer with the 24-2, white-on-white test strategy using the Full Threshold algorithm. The light source can be attenuated in the range from 1 to 10,000 times. On the decibel (dB) scale an attenuation x is defined as $s = 10 \, \texttt{log}_{10}(x)$, or $x = 10^{s/10}$. The lowest sensitivity that can be detected by this perimeter is 0 dB, although negative values could in fact occur if it were not for the limitations of this device. The highest sensitivity that can be detected is 50 dB, however few humans are capable of seeing a stimulus less than 40 dB, which is 1/10,000 of the maximum intensity of the instrument (or 1 asb). Thus, for practical purposes, the useful intensity range for white light testing is from 0 to 40 dB with a background illumination of 31.5 asb. [1].
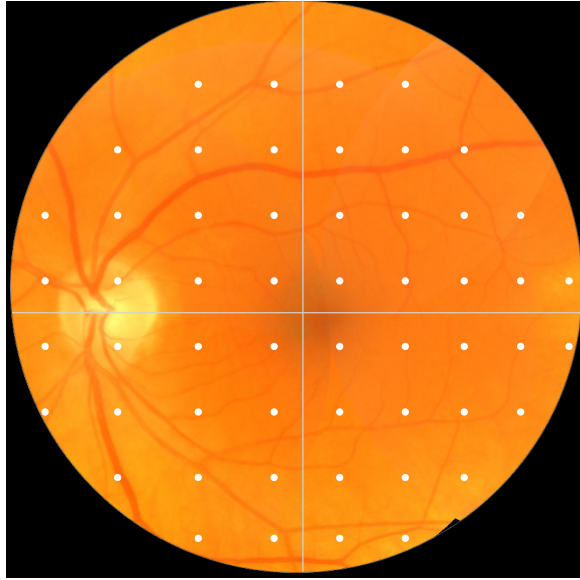
Figure 1: Fundus photo of a left human eye with the 54 test locations for the VF test represented by white dots.

2.2 Previous research

Parameters such as the mean deviation (MD) and visual field index (VFI) summarize the 52 sensitivity estimates into single values which can be used by the clinicians when optimizing treatment strategies. Longitudinal modeling of these VF summary parameters has been done before [2, 4, 13, 15]. Modeling of individual test locations is potentially of greater interest, because it provides additional information such as the spatial nature of the fields which is otherwise lost in global parameters. In previous research, each location was analyzed as an independent sample [5, 6, 20]. However, separate location-specific regression models are not able to use any information from the data set as a whole. Multilevel mixed-effects models provide a better fit to the data than separate regression models by accounting for group effects and/or within-group correlation [29]. This was shown in the context of global VF measurements by Pathak et al. [22].

In glaucoma, variability is presumably related to fatigue effects and response errors, whereby sensitivity estimates decrease over time [3, 12]. Differences in fatigue effects, between the inferior and superior hemifields within an eye have been demonstrated [12]. Furthermore, this effect may differ between the first and second eye at the same visit. The number of false-negative answers have been shown to be higher in eyes with field loss. This may be explained by an increased variability in sensitivity estimates typically found in such eyes [3, 25]. A common approach to reduce measurement variability is to average multiple

measurements. For example, random uncorrelated measurement errors that are present in the point-wise sensitivity estimates are reduced when calculating summary parameters such as the mean deviation (MD). Other errors, however, are spatially correlated and affect the whole VF. One group of such errors are measurable factors, including season, time of day and reliability indices, which have been evaluated before [13]. Although these factors are statistically significant, they are rather small and hence only explain a small part of the observed global variation in VFs.

The inverse relationship between variability and sensitivity has been described in the literature. Henson et al. (2000) [10] found that this relationship is well represented by the function $\log(SD) = A + B \times \texttt{sensitivity}(dB)$, where A and B are 2.81 dB and -0.066 dB respectively for normal eyes and 3.62 dB and -0.098 dB for glaucomatous eyes. Russell et al. (2012) [25] showed that the distribution of residuals is relatively concentrated at high VF sensitivities (26 to 36 dB) but stretches substantially as the sensitivity estimates decrease to a level of 10 dB. Sensitivity estimates near 10 dB are associated with residuals spanning almost the entire dynamic range of the instrument. This could be caused by a loss of ganglion cells (due to glaucomatous damage), or relocation of the stimulus to the peripheral visual field where there are fewer ganglion cells [27]. Zhu et al. (2014) [30] describe a method to detect change using an inferential statistical model which incorporates the non-stationary variability using a mixture of Weibull distributions.

Although there is a wide range of literature which discusses these aspects, the majority of previous work deals with the global indices or treats each point-wise estimate as an independent sample. Furthermore, these aspects have been addressed separately. Hence, it is clear that an approach which takes into account the complex structure of the data and considers all of the aforementioned problems, is needed. We will address censoring, the hierarchical structure, the global variation as well as the relationship between the variability and sensitivity.

3   Statistical Models

Modeling the sensitivity estimates is beneficial for the evaluation of the progression of VF loss. By incorporating biological effects into the model, we aimed to improve the model fit and hence provide a better method for modeling this progression. This was done by building the model up sequentially.

3.1 Censoring

It is important to note that unseen sensitivity estimates are indicated on the VF print out as < 0. They are smaller than zero because the instrument is unable to determine such sensitivities. Thus a model which defines the relationship between time and the latent, true sensitivity value is needed. The relationship between the observed y^* and the latent, true sensitivity value y is given by,

$$y^* = y \times I(y \geq 0) \; + \; 0 \times I(y < 0).$$

3.2 Hierarchical Model

We propose using a Bayesian hierarchical mixed effects model [17, 21] to analyze the glaucoma data. This model is able to take into account both the within subject and between subject variability. Furthermore, we capitalized on the common features within each eye by taking into account the correlation between measurements belonging to the same eye. In addition, correlation of VF measurements within the inferior and superior hemifields, separated by the horizontal raphe, was assumed to be higher than between hemifields. Hence, the hierarchical structure of the data consists of 4 levels, namely, (1) the individual, (2) the eye (3) the hemifield and (4) the location. Let β correspond to the regression parameters and $\texttt{years}_{it}$ represent the time between measurement t and the first measurement for each individual i, ranging from 0 to 10.5 years. The individual-specific intercept and slope are represented by α, the eye-specific intercept and slope by γ, the hemifield-specific intercept and slope by η, and the location-specific intercept and slope by λ. We then have, for individual $i = 1, \dots, N$; eye $e = 1, 2$; hemifield $h = 1, 2$; location $l = 1, \dots, 26$ and timepoint $t = 1, \dots, T_i$,

Model 1:

$$
\begin{aligned}
y_{iehlt} &= \beta_0 + \beta_1 \texttt{years}_{it} + \alpha_{0i} + \alpha_{1i}\texttt{years}_{it} + \gamma_{0ie} + \gamma_{1ie}\texttt{years}_{it} + \eta_{0ieh} + \eta_{1ieh}\texttt{years}_{it} + \\
&\quad \lambda_{0iehl} + \lambda_{1iehl}\texttt{years}_{it} + \epsilon_{iehlt} \hspace{4cm} (1) \\
&= \mu^{(1)}_{iehlt} + \epsilon_{iehlt},
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha_i &= \left(\begin{smallmatrix}\alpha_{0i}\\\alpha_{1i}\end{smallmatrix}\right) \sim N\left(\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right), \Sigma_\alpha = \left(\begin{smallmatrix}\Sigma_{\alpha 11} & \Sigma_{\alpha 12}\\\Sigma_{\alpha 21} & \Sigma_{\alpha 22}\end{smallmatrix}\right)\right); \\
\gamma_{ie} &= \left(\begin{smallmatrix}\gamma_{0ie}\\\gamma_{1ie}\end{smallmatrix}\right) \sim N\left(\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right), \Sigma_\gamma = \left(\begin{smallmatrix}\Sigma_{\gamma 11} & \Sigma_{\gamma 12}\\\Sigma_{\gamma 21} & \Sigma_{\gamma 22}\end{smallmatrix}\right)\right); \\
\eta_{ieh} &= \left(\begin{smallmatrix}\eta_{0ieh}\\\eta_{1ieh}\end{smallmatrix}\right) \sim N\left(\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right), \Sigma_\eta = \left(\begin{smallmatrix}\Sigma_{\eta 11} & \Sigma_{\eta 12}\\\Sigma_{\eta 21} & \Sigma_{\eta 22}\end{smallmatrix}\right)\right); \\
\lambda_{iehl} &= \left(\begin{smallmatrix}\lambda_{0iehl}\\\lambda_{1iehl}\end{smallmatrix}\right) \sim N\left(\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right), \Sigma_\lambda = \left(\begin{smallmatrix}\Sigma_{\lambda 11} & \Sigma_{\lambda 12}\\\Sigma_{\lambda 21} & \Sigma_{\lambda 22}\end{smallmatrix}\right)\right) \text{ and} \\
\epsilon_{iehlt} &\sim N(0, \sigma^2).
\end{aligned}
$$

3.3   Visit Effect

Junoy Montolio et al. (2012) [13] explicitly modeled the global variations with known factors such as season, time of day and reliability indices. However, we speculated that other transient factors, such as fatigue, lack of concentration, or delayed reaction time may play a more important role. Since all these (as well as possibly other) factors, affect all locations belonging to the same VF, we propose to take them together and to call them, as well as model them as the Global Visit Effects (GVEs). In this way, we can account for both the known and the unknown factors. Hence, the GVE accounts for all factors that affect all measurements of the same eye at each visit. To illustrate the importance of these factors, we show in Figure 2 the VFs over time of one eye, where all locations have a drastic decrease in sensitivity at around 1 year. From the longitudinal profiles, it is evident that this decrease is caused by something that affected all VF measurements of that visit, rather than by actual damage. To account for the visit-dependent offset at all locations, or GVE, we included a parameter, ϕ_{iet}, in the model to capture the offset at every visit j for each eye k within each individual i. This gives,

Model 2:

$$
\begin{aligned}
y_{iehlt} &= \mu^{(1)}_{iehlt} + \phi_{iet} + \epsilon_{iehlt} \\
&= \mu^{(2)}_{iehlt} + \epsilon_{iehlt}.
\end{aligned} \tag{2}
$$

From an initial exploratory analysis, we observed a number of spikes in the distribution of the visit effects. To accommodate these spikes, we assumed a t-distribution for ϕ_{iet}. The t-distribution allows greater flexibility in the distribution of random effects compared
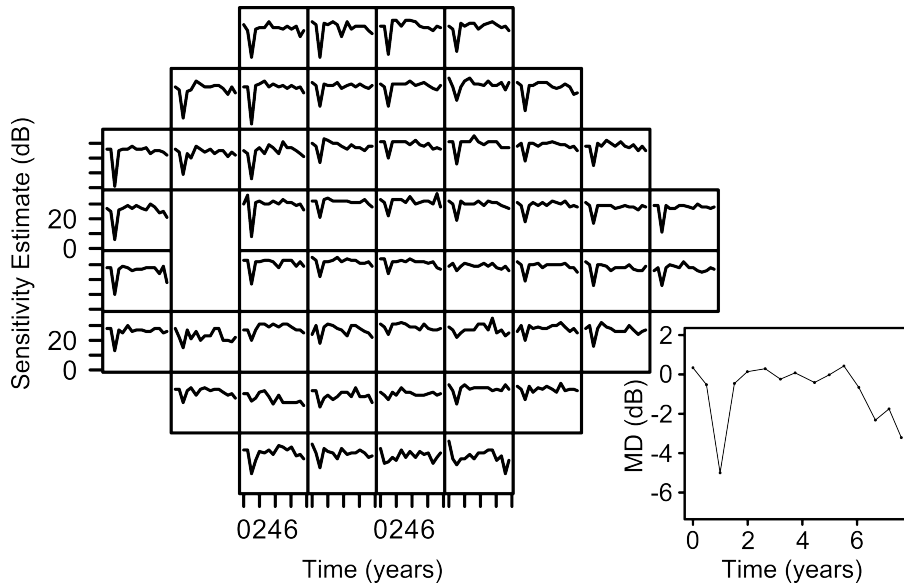
Figure 2: Retinal sensitivity estimates over time for each location of the visual field in the left eye of a single glaucoma patient. A decrease in the sensitivity estimates can be seen in all locations at around 1 year. The longitudinal profile of the MD values over time are shown on the right. The visit-dependent decrease is also clear at around 1 year for the MD.

to the normal distribution, and can handle heavy tails in random effects distributions [16]. Hence, we let,

$$\phi_{iet} \sim t(0, \sigma_\phi^2, 3),$$

where $t(\mu, \sigma^2, df)$ denotes the generalized t-distribution with mean μ, scale parameter σ, and df degrees of freedom.

3.4 Relationship between Variability and Sensitivity

There is an association between a decline in VF sensitivity and an increase in response variability. However, values lower than 0 dB cannot be measured. This inherent censoring process introduces a positive bias at low sensitivity estimates, which is made worse by the increased variability for low sensitivity estimates. We assumed a linear relationship between the expected values of the sensitivity estimates and the logarithm of the standard deviation. However, since we were interested in modeling the latent sensitivity estimates, we extrapolated this linear relationship for predicted sensitivity estimates below 10 dB. This can be seen in Figure 3. In this exploratory analysis, we found that the relationship was well represented by the function $\log(SD) = A + B \times \texttt{sensitivity}(dB)$, where A

and B are 2.60 dB and -0.06 dB respectively. We extended Model 2 to incorporate this relationship such that,

Model 3:

$$y_{iehlt} \quad = \quad \mu_{iehlt}^{(2)} + \epsilon_{iehlt},$$

and

$$
\begin{aligned}
\log(\sigma_{iehlt}) \quad &= \quad f\{E(y_{iehlt})\} \\
&= \quad \beta_0^* + \beta_1^* \mu_{iehlt}^{(2)},
\end{aligned}
\tag{3}
$$

where f is a linear function. A summary of all the parameters and their definitions for all the models is shown in Table 1.
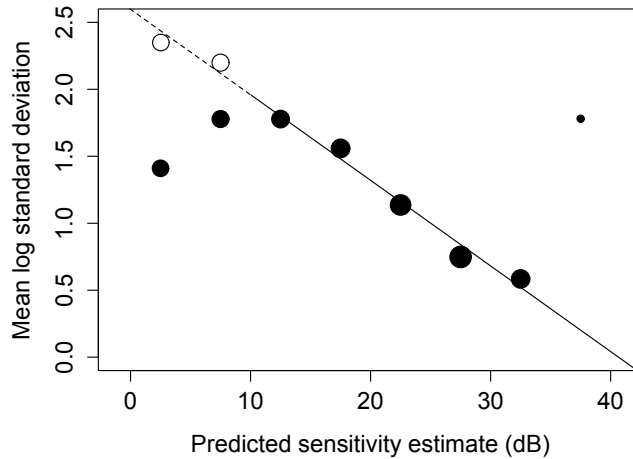

Figure 3: Bubbleplot representing the mean logarithm of the standard deviation for different predicted sensitivity estimates determined using linear regression for each location. The predicted values were subdivided into groups with width 5 dB. The empty bubbles correspond to the hypothetical values, corresponding to the censored measurements, for the mean logarithm of the standard deviation for the predicted sensitivity estimates after extrapolation. The bubbles are scaled to the logarithm of the number of observations.

Table 1: Summary of parameters included cumulatively in each of the models

Model	Parameter	Definition
Model 1	y_{iehlt}	Latent sensitivity estimate
	β_0	Population-averaged intercept
	β_1	Population-averaged slope
	α_{0i}	Individual-specific intercept
	α_{1i}	Individual-specific slope
	γ_{0i}	Eye-specific intercept
	γ_{1i}	Eye-specific slope
	η_{0i}	Hemifield-specific intercept
	η_{1i}	Hemifield-specific slope
	λ_{0i}	Location-specific intercept
	λ_{1i}	Location-specific slope
	σ^2	Variance
Model 2	ϕ_{iet}	Global visit effect
	σ_ϕ^2	Global Visit Effect variance
Model 3	β_0^*	Intercept in logarithm of the standard deviation
	β_1^*	Slope in logarithm of the standard deviation

4 Estimation Approach

4.1 One-stage approach

The Bayesian approach takes into account the uncertainty in all model parameters and allows for prior information to be incorporated. Furthermore, MCMC algorithms allow greater flexibility by relaxing the strong parametric assumptions commonly used in most frequentist hierarchical models [17, 18]. The classical Bayesian approach is one-stage hierarchical modeling, which has the advantage that subject-specific and overall parameters are estimated simultaneously. However, for a (relatively) large data set, this approach can be difficult or even impossible to implement for complex models with standard MCMC software. In our case, we had a total of 45,005 parameters which needed to be estimated. As a consequence, we were unable to achieve convergence in a realistic time frame and experienced computer memory limitations when using WinBUGS or JAGS. For such situations, a computationally more efficient method is needed.

4.2 Two-stage approach

Lunn et al. [18] proposed two-stage Bayesian hierarchical modeling. The glaucoma data also exhibit an hierarchical structure, but of a more complex nature. Figure 4 illustrates the hierarchical structure of the glaucoma data, as well as the cross-classified random

effects, divided into two stages. The two-stage approach allowed us to simplify the problem by splitting hierarchical models with M levels at level m*. Independent parameters of interest at level m* are obtained in stage 1 and used as proposal distributions for those parameters in stage 2. Lunn et al. illustrated this method using models with two and three levels. We applied this to a more complex model with four levels. In our case, we split the levels at the individual level, treating each individual as their own sample. These individuals were then analyzed independently before combining them to obtain population level estimates.
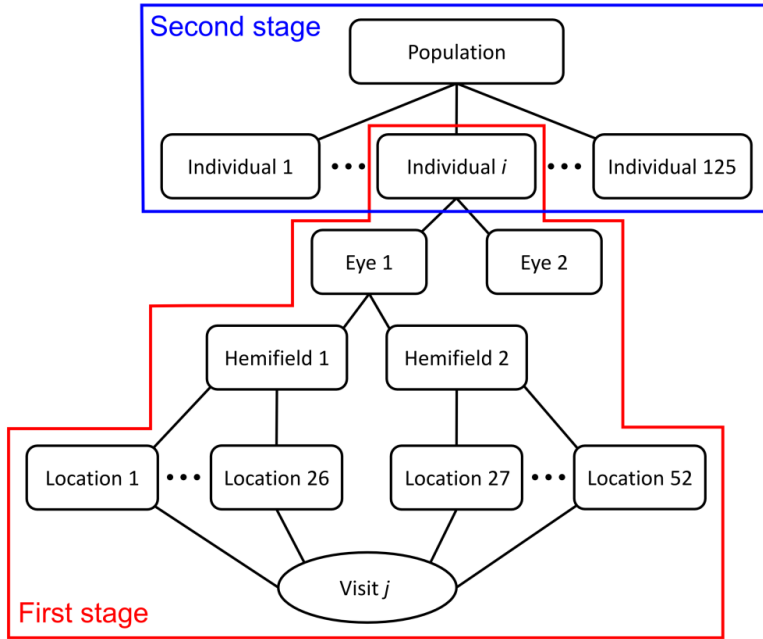

Figure 4: Illustration of the hierarchical structure of the data divided into the first and second stages as done in the two-stage approach.

4.2.1 First stage

In the first stage, we analyzed each individual separately. Without loss of generality, we only show this for Model 3. This becomes:

$$
\begin{aligned}
y_{iehlt} = \; & \alpha_{0i} + \alpha_{1i}\textbf{years}_{it} + \gamma_{0ie} + \gamma_{1ie}\textbf{years}_{it} + \eta_{0ieh} + \eta_{1ieh}\textbf{years}_{it} + \\
& \lambda_{0iehl} + \lambda_{1iehl}\textbf{years}_{it} + \phi_{iet} + \epsilon_{iehlt},
\end{aligned}
\tag{4}
$$

where

$$\gamma_{ie} = \binom{\gamma_{0ie}}{\gamma_{1ie}} \sim N\left(\binom{0}{0}, \Sigma_{\gamma i} = \left(\begin{smallmatrix} \Sigma_{\gamma 11i} & \Sigma_{\gamma 12i} \\ \Sigma_{\gamma 21i} & \Sigma_{\gamma 22i} \end{smallmatrix}\right)\right);$$

$$\eta_{ieh} = \binom{\eta_{0ieh}}{\eta_{1ieh}} \sim N\left(\binom{0}{0}, \Sigma_{\eta i} = \left(\begin{smallmatrix} \Sigma_{\eta 11i} & \Sigma_{\eta 12i} \\ \Sigma_{\eta 21i} & \Sigma_{\eta 22i} \end{smallmatrix}\right)\right);$$

$$\lambda_{iehl} = \binom{\lambda_{0iehl}}{\lambda_{1iehl}} \sim N\left(\binom{0}{0}, \Sigma_{\lambda i} = \left(\begin{smallmatrix} \Sigma_{\lambda 11i} & \Sigma_{\lambda 12i} \\ \Sigma_{\lambda 21i} & \Sigma_{\lambda 22i} \end{smallmatrix}\right)\right);$$

$$\phi_{iet} \sim t(0, \sigma_{\phi i}^2, 3) \text{ and}$$

$$\epsilon_{iehlt} \sim N(0, \sigma_{iehlt}^2),$$

and

$$\log(\sigma_{iehlt}) = \beta_{0i}^* + \beta_{1i}^* \mu_{iehlt}^{(2)}.$$

One important detail about the two-stage approach is that it allows the individual variances to differ, i.e. $\Sigma_{\gamma i}$, $\Sigma_{\eta i}$, $\Sigma_{\lambda i}$ and $\sigma_{\phi i}^2$, but also σ_{iehlt} since the regression coefficients are now allowed to depend on the subject. This is in contrast to the one-stage model which requires the variances to be the same, i.e. Σ_γ, Σ_η, Σ_λ and σ_ϕ^2. Hence, the two-stage approach is more flexible, as it can account for these differences if they are present in the data. In the Bayesian procedure prior distributions need to be assumed for all parameters. Explicit expressions for the priors are given in the Appendix. In order to prevent the second-stage sampler from becoming stuck near local posterior modes, large independent samples are needed from this first stage [18]. To achieve this, we ran 200,000 iterations with a burn-in of 150,000 and thinning of 10, resulting in 138 samples of 5,000 iterations each for each parameter.

4.2.2 Second stage

In the first stage, α_i and β_i^* were treated as fixed effects. These parameters need to be combined in the second stage to obtain the population-averaged effects, β and β^*. We denote them as θ_i. In the case of a meta analysis, the random effects in the first stage, which we denote as L_i, may not be of direct interest. Hence, these terms can be treated as nuisance parameters as done by Lunn et al. [18]. However, for clinical applications such as ours, these may be important. In order to avoid further computational problems, we took only the covariance matrices of the random effects in the first stage

to the second stage, which allowed us to re-estimate the random effects in an additional step. Since each of the elements in the matrices were treated as separate parameters in the second stage, Cholesky decomposition of the covariance matrices of the random effects, Σ_{γ_i}, Σ_{η_i} and Σ_{λ_i} respectively, was used. More specifically, for each of the above covariance matrices a full rank lower triangular matrix U with real and positive diagonal entries was generated, ensuring that UU^T is positive definite. We denote the Cholesky decomposition factors for all of the covariance matrices by C_i. Hence, we let $\{\theta_i, C_i\}$ represent parameters of interest from the first-stage. The samples of these parameters were then used as proposal distributions within a Metropolis-Hastings step in the second-stage to obtain $\{\theta, C\} = \{\theta_i, C_i, i = 1, \ldots, N\}$. Three chains were initialized with different starting values for all models determined from the first-stage samples. This was done using the minimum value, the mean value and the maximum value for each parameter for every individual. Upon convergence, we computed the posterior mean, median, standard deviation with the equal tail 95% credible interval (CI) for all parameters of interest.

5 Model Evaluation

Standard approaches are applicable to the results from the first stage since this stage represents a standard analysis. Hence, we evaluated the models at this stage for each individual separately using posterior predictive checks (PPC), such as the χ^2-test statistic. A further comparison of the models was done after the second stage using the Deviance Information Criterion (DIC) to determine the overall best model.

5.1 Posterior predictive check

We denote all parameters for individual i from the first stage, i.e. $\{\theta_i, C_i, L_i\}$, as ψ_i. Let $\psi_i^1, \ldots, \psi_i^K$ be the converged Markov chain from $p(\psi_i \mid y_i)$. Furthermore the vector of all responses for the ith individual is denoted by y_i. The posterior predictive P-value (PPP) for a discrepancy measure, $D(y_i \mid \psi_i^k)$ is then calculated and replicated data $\tilde{y}_i^k$ is sampled from $p(y_i \mid \psi_i^k)$. $D(\tilde{y}_i^k, \psi_i^k)$ can then be computed for k in $\{1, \ldots, K\}$, and p_D can be estimated by,

$$\bar{p}_D = \frac{1}{K} \sum_{k=1}^{K} I[D(\tilde{y}_i^k, \psi_i^k) \leq D(y_i, \psi_i^k)], \tag{5}$$

Here, the Gelman χ^2-test statistic [9] was used as the discrepancy measure to calculate the PPC for each individual. This is defined as:

$$D(y_i, \psi_i^k) = \sum_{e=1}^{E} \sum_{h=1}^{H} \sum_{l=1}^{L} \sum_{t=1}^{T_i} \frac{[y_{iehlt} - E(y_{iehlt} \mid \psi_i^k)]^2}{\texttt{var}(y_{iehlt} \mid \psi_i^k)}, \tag{6}$$

where $E(y_{iehlt})$ is defined as $\mu_{iehlt}^{(2)}$. A small value indicates a bad model fit. The above predictive P-values can then be contrasted against a uniform distribution to evaluate globally model fit for each individual. In general, if $\bar{p}_D$ is smaller than 0.05 or larger than 0.95, then this is an indication that the model might not fit the data well. Note that this procedure is, however, conservative because the data is used twice: one for model fit and one for model evaluation, see e.g. [17].

5.2   Deviance Information Criterion

In a Bayesian framework, a common tool for model evaluation is the Deviance Information Criterion (DIC) proposed by Spiegelhalter et al. (2002) [26]. The DIC is defined as:

$$DIC = D(\{\bar{\theta}, \bar{C}\}, \bar{L}) + 2p_D = \overline{D(\{\theta, C\}, L)} + p_D, \tag{7}$$

where

$$p_D = \overline{D(\{\theta, C\}L)} - D(\{\bar{\theta}, \bar{C}\}, \bar{L}).$$

In the definition of the DIC, we have the fixed effects parameters as well as the random effects which are treated as nuisance parameters in the two-stage approach. One of the disadvantages of the two-stage approach is not being able to directly obtain the random effects estimates. Since most model evaluation methods, such as the Deviance Information Criteria (DIC), require the random effect estimates for the computation, it is not clear

how to evaluate models using this approach. To overcome this limitation, we propose an extension of the two-stage approach, by including an additional step based on the Method of Composition in combination with a Metropolis-within-Gibbs technique. More specifically, for the calculation of the DIC we are required to obtain a sample of the random effects from their posterior distribution $p(L_i \mid y_i)$, which is written as:

$$p(L_i \mid y_i) = \int p(L_i \mid y_i, \Omega_i) p(\Omega_i \mid y_i) d\Omega_i, \tag{8}$$

where $\Omega_i = (\beta_0, \beta_1, \beta_0^*, \beta_1^*, \alpha_{0i}, \alpha_{1i}, \Sigma_\gamma, \Sigma_\eta, \Sigma_\lambda, \sigma_\phi^2)$ is the vector of all parameters of main interest; $\tilde{\Omega}_i$ represents the sampled values from the second stage. Identity (8) suggests that we can use the following simulation scheme. This is a sampling algorithm based on the Method of Composition in combination with a Metropolis-within-Gibbs technique, which is applied to each individual. For the ith individual the computations are done as follows:

Step 1: For iteration k, let the parameters estimated in the second stage of the two-stage approach be denoted by $\tilde{\Omega}_i^{(k)}$.

Step 2: Given $\tilde{\Omega}_i^{(k)}$, we sample:

$$\gamma_{0ie}, \gamma_{1ie}, \eta_{0ieh}, \eta_{1ieh}, \lambda_{0iehl}, \lambda_{1iehl} \text{ and } \phi_{iet},$$

which we denote as $L_i^{(k)}$. This is done using the Metropolis-within-Gibbs technique. Thus, for each sequence of generated parameters from the second stage, we apply MCMC sampling to obtain these estimates:

Step 2A: Initial values are determined using an optimization routine.

Step 2B: A random walk Metropolis Hastings algorithm is used for each of the levels. This is done iteratively, to take into account the correlation between the levels.

Step 3: After an inital burn-in, we save the last estimate for each parameter in $L_i^{(k)}$.

Step 4: This is repeated for K iterations, resulting in:

$$L_i^{(1)}, L_i^{(2)}, \ldots, L_i^{(K)}$$

Hence, in combination with the results from the second stage, we have now obtained

all the parameter estimates which are needed to compute the DIC.

6 Application to the Glaucoma study

For this analysis we included both eyes from the 139 individuals belonging to the Glaucoma study. After excluding VFs with unknown reliability as indicated by the instrument, 138 individuals and 276 eyes remained. This included 4,758 VFs, resulting in a data set consisting of 14,352 VFs and 247,520 location-specific sensitivity estimates. All analyses were done taking into account censoring, and hence using the latent sensitivity values, y_{iehlt}.

6.1 Results

The two-stage approach is advantageous, as it allows us to do exploratory analyses at the individual level in order to simplify the model before combining the samples in the second stage. In order to compare the models, we can evaluate the outcome of the PPC using graphical output. The PPP-values denoted by $\bar{p}_D$ were computed for every individual. Figure 5 shows the ordered PPP-values for each of the models. Model 1 has a mean $\bar{p}_D = 0.30$, Model 2 a mean $\bar{p}_D = 0.30$ and Model 3 a mean $\bar{p}_D = 0.50$. From this, it appears that Model 3 has the best fit. This approach gives a good indication of whether the models fit the data, specifically for each individual.
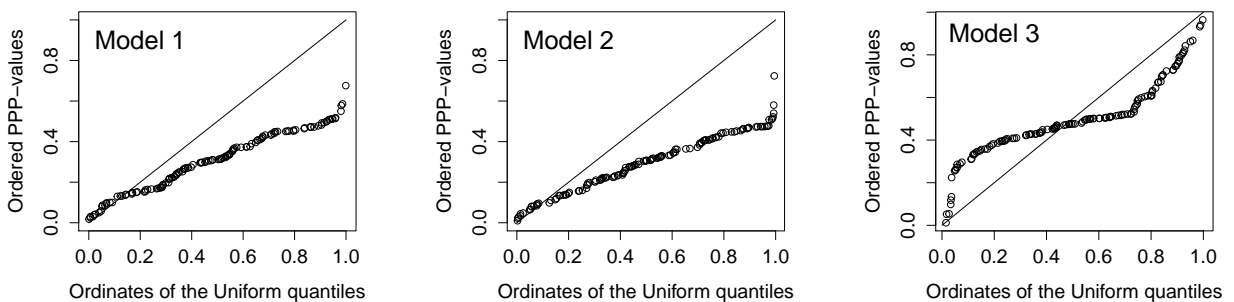

Figure 5: Posterior predictive check for each of the models across all individuals

An example of the model fits for 1 location is shown in Figure 6. The posterior summary statistics from the second stage are listed in Table 2 for each of the models. A difference in DIC of more than 10 indicates that the model with the lowest DIC has a

better fit. Using the DIC to compare the models, Model 2 (DIC=$-9.045e^{+07}$) performed better than Model 1 (DIC = $-8.827e^{+07}$), with Model 3 (DIC = $-1.792e^{+26}$) performing the best overall. Using the results from Model 3, the population intercept (β_0) was 19.82 dB with an average slope (β_1) of -0.31 dB per year. The intercept (β_0^*) and slope (β_1^*) for the logarithm of the standard deviation was 2.82 dB and -0.08 dB respectively. This corresponds to the 2.60 dB and -0.06 dB which was found in the exploratory analysis shown in Figure 3.
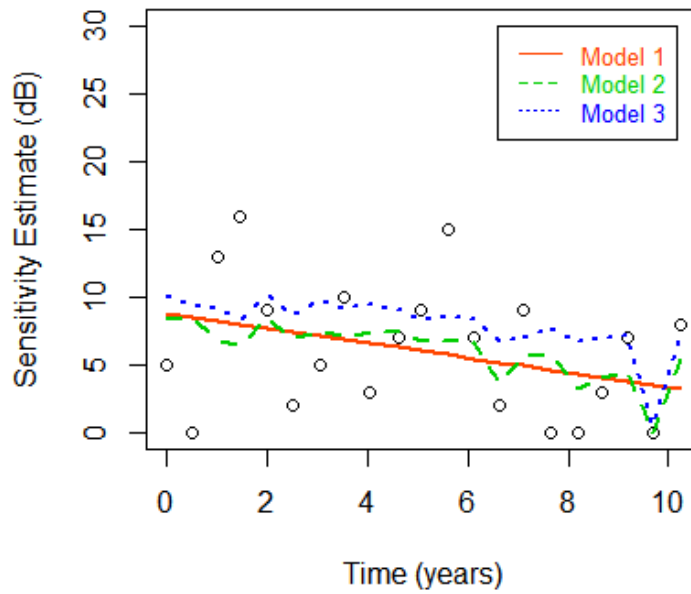

Figure 6: Scatter plot representing the retinal sensitivity estimates over time for 1 location of the VF. The lines represent the model fits for each of the 3 models

Table 2: Posterior summary statistics for the three models using the two-stage approach

Model	1			2			3		
Parameter	mean	sd	95% CI	mean	sd	95% CI	mean	sd	95% CI
β_0	18.95	0.72	(17.53 ; 20.36)	20.42	0.73	(18.95 ; 21.84)	19.89	0.77	(18.36 ; 21.37)
β_1	-0.22	0.0	(-0.33 ; -0.13)	-0.21	0.05	(-0.31 ; -0.11)	-0.31	0.05	(-0.41 ; 0.20)
β_0^*							2.82	0.06	(2.70 ; 2.95)
β_1^*							-0.08	0.08	(-0.08 ; -0.07)
σ^2	13.42	0.66	(12.17 ; 14.75)	11.51	0.56	(10.45 ; 12.68)			
σ_ϕ^2				0.62	0.05	(0.52 ; 0.73)	1.87	0.04	(1.81 ; 1.96)
DIC		$-8.827e^{+07}$			$-9.045e^{+07}$			$-1.792e^{+26}$	

6.2 Clinical Implications

With the GVE, we account for those factors, as well as those which can not be measured such as fatigue and delayed reaction time. Including the GVE showed a significant improvement in the model fit. Hence, by taking into account the GVE we were able to take into account a large part of the variability and obtain better estimates of the true rate of progression. By including the relationship between variability and sensitivity shows a further improvement in the model fit. The function which describes this relationship was consistent to that found by Henson et al [10], however it was not shown previously how to include this relationship in a model, or whether including it would improve the model fit. By including both of these aspects, we were able to improve the estimation of the true underlying progression and determine the real evolution of the sensitivity over time. Hence, these improved estimates could aid clinicians in optimizing treatment strategies.

7 Discussion

In this paper, we proposed a method to model point-wise VFs taking into account the complexity of psychophysical testing of visual function in glaucoma. The model is advantageous in dealing with the high measurement variability, and could be extended for the prediction of future VFs. Although it was possible to use the one-stage approach with simplied versions of the model or with smaller datasets, it was not possible to perform these analysis on the full data with a complex model as it was with the two-stage approach. The two-stage approach can be implemented in standard MCMC software. The relevant computations for the first-stage can be carried out in JAGS [23], WinBUGS or OpenBUGS [19] software and the second-stage using OpenBUGS software [18]. However, for the second stage an add-on program is needed. For more details on setting up Open-BUGS for performing the two-stage analyses, we refer to [18]. More information regarding the computations done in this paper can be obtained by emailing the first author. These computations can be easily tuned to adapt to other data sets by any practitioner.

The two-stage method is advantageous as it allows us to do exploratory analysis at an individual level. Hence, we are able to simplify and improve the model before combining it at a population level. Limited simulations showed that the one- and two-stage approaches gave similar results if the variances were the same for all individuals. The two-stage

approach assumes a more flexible method. However, there is the additional difficulty in constraining the parameters across individuals. One disadvantage of this approach is that it does not provide the required components to evaluate the fit and predictive ability of the model using the Deviance Information Criterion (DIC). In order to calculate the DIC and compare different competing models for our data fitted using the two-stage approach, we suggested a Monte Carlo scheme based on a Metropolis-within-Gibbs algorithm.

Other issues, which we see as future research directions, is to look at the optimal choice of the level where the data should be split. Extensions include exploiting the spatial nature of the data and capitalizing on the specific spatial organization of the nerve fibres in the eye [7].

References

1. Anderson, D. R. and Patella, V. M. (1999). *Automated static perimetry.* Mosby, 2 edition.

2. Artes, P. H., O'Leary, N., Hutchison, D. M., Heckler, L., Sharpe, G. P., Nicolela, M. T., and Chauhan, B. C. (2011). Properties of the statpac visual field index. *Investigative Ophthalmology & Visual Science*, 52(7):4030–4038.

3. Bengtsson, B. and Heijl, A. (2000). False-negative responses in glaucoma perimetry: Indicators of patient performance or test reliability? *Investigative Opthamology and Visual Science*, 41:2201–2204.

4. Bengtsson, B. and Heijl, A. (2008). A visual field index for calculation of glaucoma rate of progression. *American Journal of Ophthalmology*, 145(2):343–353.

5. Bryan, S. R., Vermeer, K. A., Eilers, P. H., Lemij, H. G., and Lesaffre, E. M. (2013). Robust and censored modeling and prediction of progression in glaucomatous visual fields. *Investigative Opthamology and Visual Science*, 54(10):6694–6700.

6. Caprioli, J., Mock, D., Bitrian, E., Afifi, A. A., Yu, F., Nouri-Mahdavi, K., and Coleman, A. L. (2011). A method to measure and predict rates of regional visual field decay in glaucoma. *Investigative Ophthalmology & Visual Science*, 52(7):4765–4773.

7. Erler, N. S., Bryan, S. R., Eilers, P. H., Lesaffre, E. M., Lemij, H. G., and Vermeer, K. A. (2013). Optimizing structure-function relationship by maximizing correspondence between glaucomatous visual fields and mathematical retinal nerve fiber models. *Investigative Opthamology and Visual Science*, 54(10):6694–6700.

8. Gelfand, A. and Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.

9. Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis.* Chapman and Hall/CRC, 3rd edition.

10. Henson, D. B., Chaudry, S., Artes, P. H., Faragher, E. B., and Ansons, A. (2000). Response variability in the visual field: Comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Investigative Opthamology and Visual Science*, 41(2):417–421.

11. Hoffman, M., , and Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15:1593–1623.

12. Hudson, C., Wild, John, and Eamon, O. (1994). Fatigue effects during a single session of automated static threshold perimetry. *Investigative Opthamology and Visual Science*, 35:268–280.

13. Junoy Montolio, F. G., Wesselink, C., Gordijn, M., and Jansonius, N. M. (2012). Factors that influence standard automated perimetry test results in glaucoma: Test reliability, technician experience, time of day, and season. *Investigative Ophthalmology & Visual Science*, 53(11):7010–7017.

14. Kingman, S. (2004). Glaucoma is second leading cause of blindness globally. *Bulletin of the World Health Organization*, 82(11):887–888.

15. Kymes, S. M., Lambert, D. L., Lee, P. P., Musch, D. C., Siegfried, C. J., Kotak, S. V., Stwalley, D. L., Fain, J., Johnson, C., and Gordon, M. O. (2012). The development of a decision analytic model of changes in mean deviation in people with glaucoma. *Ophthalmology*, 119(7):1367–1374.

16. Lee, K. and Thompson, S. (2008). Flexible parametric models for random-effects distributions. *Statistics in Medicine*, 27(10):418–434.

17. Lesaffre, E. and Lawson, A. (2012). *Bayesian Biostatistics*. Wiley & Sons, 1st edition.

18. Lunn, D., Barrett, J., Sweeting, M., and Thompson, S. (2013). Fully bayesian hierarchical modelling in two stages, with application to meta-analysis. *Journal of the Royal Society, Series C, Applied Sciences*, 62(4):551–572.

19. Lunn, D., Spiegerhalter, D., Thomas, A., and Best, N. (2009). The bugs project: Evolution, critque and future directions. *Statistics in Medicine*, 28(25):3049–3067.

20. McNaught, A. I., Hitchings, R. A., Crabb, D. P., and Fitzke, F. W. (1995). Modelling series of visual fields to detect progression in normal-tension glaucoma. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 233(12):750–755.

21. Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Wiley & Sons, 1st edition.

22. Pathak, M., Demirel, S., and Gariner, S. (2013). Nonlinear, multilevel mixed-effects approach for modeling longitudinal standard automated perimetry data in glaucoma. *Investigative Opthamology and Visual Science*, 54:5505–5513.

23. Plummer, M. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling.

24. Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations (with discussion). *Journal of the Royal Statistical Society: Series B*, 71(2):319–392.

25. Russell, R. A., Crabb, D. P., Malik, R., and Garway-Heath, D. F. (2012). The relationship between variability and sensitivity in large-scale longitudinal visual field data. *Investigative Ophthalmology & Visual Science*, 53(10):5985–5990.

26. Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (and discussion). *Journal of the Royal Statistical Society: Series B*, (64):583–639.

27. Swanson, W. H., Sun, H., Lee, B. B., and Cao, D. (2011). Responses of primate retinal ganglion cells to perimetric stimuli. *Investigative Ophthalmology & Visual Science*, 52(2):764–771.

28. Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 26(1):24.

29. Verbeke, G. and Molenberghs, G. (2009). *Linear Mixed Models for Longitudinal Data.* Springer, 1st edition.

30. Zhu, H., Russell, R. A., Saunders, L. J., Ceccon, S., Garway-Heath, D. F., and Crabb, D. P. (2014). Detecting changes in retinal function: Analysis with non-stationary weibull error regression and spatial enhancement (answers). *PLoS ONE*, 1(9).

A Appendix

A.1 One Stage Approach

A.1.1 Full Model

$$
\begin{aligned}
y_{iehlt} = {} & \beta_0 + \beta_1 \text{years}_{it} + \alpha_{0i} + \alpha_{1i}\text{years}_{it} + \gamma_{0ie} + \gamma_{1ie}\text{years}_{it} + \\
& \eta_{0ieh} + \eta_{1ieh}\text{years}_{it} + \lambda_{0iehl} + \lambda_{1iehl}\text{years}_{ij} + \phi_{iet} + \epsilon_{iehlt} \quad \text{(A.1)}
\end{aligned}
$$

where $\epsilon_{iehlt} \sim N(0, \sigma^2_{iehlt})$ and $log(\sigma_{iehlt}) = \beta_0^* + \beta_1^* \mu_{iehlt}^*$.

A.1.2 Priors

In the Bayesian procedure prior distributions need to be stipulated for all parameters. When no prior information is available then the prior distribution should reflect this. In this case a vague prior is a natural choice.

$$
\begin{aligned}
\beta_b &\sim N(0, 10^8) \text{ for } b = 0, 1; \\
\beta_q^* &\sim N(0, 10^8) \text{ for } q = 0, 1 \\
\alpha_i &= \left(\begin{smallmatrix} \alpha_{0i} \\ \alpha_{1i} \end{smallmatrix}\right) \sim N\left(\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right), \Sigma_\alpha = \left(\begin{smallmatrix} \Sigma_{\alpha 11} & \Sigma_{\alpha 12} \\ \Sigma_{\alpha 21} & \Sigma_{\alpha 22} \end{smallmatrix}\right)\right); \\
\gamma_{ie} &= \left(\begin{smallmatrix} \gamma_{0ie} \\ \gamma_{1ie} \end{smallmatrix}\right) \sim N\left(\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right), \Sigma_\gamma = \left(\begin{smallmatrix} \Sigma_{\gamma 11} & \Sigma_{\gamma 12} \\ \Sigma_{\gamma 21} & \Sigma_{\gamma 22} \end{smallmatrix}\right)\right); \\
\eta_{ieh} &= \left(\begin{smallmatrix} \eta_{0ieh} \\ \eta_{1ieh} \end{smallmatrix}\right) \sim N\left(\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right), \Sigma_\eta = \left(\begin{smallmatrix} \Sigma_{\eta 11} & \Sigma_{\eta 12} \\ \Sigma_{\eta 21} & \Sigma_{\eta 22} \end{smallmatrix}\right)\right); \\
\lambda_{iehl} &= \left(\begin{smallmatrix} \lambda_{0iehl} \\ \lambda_{1iehl} \end{smallmatrix}\right) \sim N\left(\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right), \Sigma_\lambda = \left(\begin{smallmatrix} \Sigma_{\lambda 11} & \Sigma_{\lambda 12} \\ \Sigma_{\lambda 21} & \Sigma_{\lambda 22} \end{smallmatrix}\right)\right); \\
\phi_{iet} &\sim t(0, \sigma_\phi^2, 3) \text{ and} \\
\epsilon_{iehlt} &\sim N(0, \sigma^2).
\end{aligned}
$$

The variance was given a vague inverse gamma prior. The covariance matrices of the random effects, i.e. Σ_γ, Σ_η and Σ_λ, were given a vague inverse Wishart distribution with degrees of freedom equal to the number of dimensions and small diagonal values for the scale matrix.

A.2 Two Stage Approach

A.2.1 Full Model

$$y_{iehlt} = \beta_0 + \beta_1 \mathbf{years}_{it} + \alpha_{0i} + \alpha_{1i}\mathbf{years}_{it} + \gamma_{0ie} + \gamma_{1ie}\mathbf{years}_{it} +$$

$$\eta_{0ieh} + \eta_{1ieh}\mathbf{years}_{it} + \lambda_{0iehl} + \lambda_{1iehl}\mathbf{years}_{it} + \phi_{iet} + \epsilon_{iehlt} \qquad \text{(A.2)}$$

where $\epsilon_{iehlt} \sim N(0, \sigma^2_{iehlt})$ and $log(\sigma_{iehlt}) = \beta^*_0 + \beta^*_1 \mu^*_{iehlt}$.

A.2.2 Priors

$$\phi_{iet} \sim t(0, \sigma^2_{\phi i}, 3); \sigma^2_{\phi_i} \sim N(\sigma^2_\phi, \Sigma_\phi)$$

$$\alpha_{gi} \sim N(0, 10^8) \text{ for } g = 0, 1;$$

$$\gamma_{ie} = \begin{pmatrix} \gamma_{0ie} \\ \gamma_{1ie} \end{pmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{\gamma i} = \begin{pmatrix} \Sigma_{\gamma 11i} & \Sigma_{\gamma 12i} \\ \Sigma_{\gamma 21i} & \Sigma_{\gamma 22i} \end{pmatrix});$$

$$\eta_{ieh} = \begin{pmatrix} \eta_{0ieh} \\ \eta_{1ieh} \end{pmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{\eta i} = \begin{pmatrix} \Sigma_{\eta 11i} & \Sigma_{\eta 12i} \\ \Sigma_{\eta 21i} & \Sigma_{\eta 22i} \end{pmatrix});$$

$$\lambda_{iehl} = \begin{pmatrix} \lambda_{0iehl} \\ \lambda_{1iehl} \end{pmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{\lambda i} = \begin{pmatrix} \Sigma_{\lambda 11i} & \Sigma_{\lambda 12i} \\ \Sigma_{\lambda 21i} & \Sigma_{\lambda 22i} \end{pmatrix}) \text{ and }$$

$$\epsilon_{iehlt} \sim N(0, \sigma^2_{iehlt}).$$

Using Cholesky decompostion, $\Sigma_{\gamma i}$, $\Sigma_{\eta i}$ and $\Sigma_{\lambda i}$ become,

$$C_{r\gamma_i} \sim N(C_{r\gamma}, 10^8) \text{ for } r = 1, 2, 3$$

$$C_{r\eta_i} \sim N(C_{r\eta}, 10^8) \text{ for } r = 1, 2, 3$$

$$C_{r\lambda_i} \sim N(C_{r\lambda}, 10^8) \text{ for } r = 1, 2, 3$$

A.2.3 First Stage Model

$$y_{iehlt} = \alpha_{0i} + \alpha_{1i}\mathbf{years}_{it} + \gamma_{0ie} + \gamma_{1ie}\mathbf{years}_{it} + \eta_{0ieh} + \eta_{1ieh}\mathbf{years}_{it} +$$

$$\lambda_{0iehl} + \lambda_{1iehl}\mathbf{years}_{it} + \phi_{iet} + \epsilon_{iehlt} \qquad \text{(A.3)}$$

where $\epsilon_{iehlt} \sim N(0, \sigma^2_{iehlt})$ and $log(\sigma_{iehlt}) = \beta^*_{0i} + \beta^*_{1i}\mu^*_{iehlt}$.

A.2.4   First Stage Priors

$$
\begin{aligned}
\beta^*_{qi} &\sim N(0, 10^8) \text{ for } q = 0, 1 \\
\alpha_{gi} &\sim N(0, 10^8) \text{ for } g = 0, 1 \\
\phi_{iet} &\sim t(0, \sigma^2_{\phi_i}, 3) \\
\gamma_{ie} &= (\gamma_{0ie}, \gamma_{1ie})^T \sim N_2((0, 0)^T, \Sigma_{\gamma_i}) \\
\eta_{ieh} &= (\eta_{0ieh}, \eta_{1ieh})^T \sim N_2((0, 0)^T, \Sigma_{\eta_i}) \\
\lambda_{iehl} &= (\lambda_{0iehl}, \lambda_{1iehl})^T \sim N_2((0, 0)^T, \Sigma_{\lambda_i}).
\end{aligned}
$$

The variance was given a vague inverse gamma prior. The covariance matrices of the random effects, i.e. Σ_{γ_i}, Σ_{η_i} and Σ_{λ_i}, were given a vague inverse Wishart distribution with degrees of freedom equal to the number of dimensions and small diagonal values for the scale matrix.

A.2.5   Second Stage Priors

$$
\begin{aligned}
\beta^*_{qi} &\sim N(\beta^*_q, 10^8) \text{ for } q = 0, 1 \\
\alpha_i &= (\alpha_{0i}, \alpha_{1i})^T \sim N_2((\beta_0, \beta_1)^T, \Sigma_\alpha) \\
\sigma^2_{\phi_i} &\sim N(\sigma^2_\phi, \Sigma_\phi) \\
C_{r\gamma_i} &\sim N(C_{r\gamma}, 10^8) \text{ for } r = 1, 2, 3 \\
C_{r\eta_i} &\sim N(C_{r\eta}, 10^8) \text{ for } r = 1, 2, 3 \\
C_{r\lambda_i} &\sim N(C_{r\lambda}, 10^8) \text{ for } r = 1, 2, 3
\end{aligned}
$$

B   Methods

Let,

$$\beta^* = \{(\beta_{0i}^*, \beta_{1i}^*), \text{ for } i = 1, \dots, N\}$$

$$\alpha = \{(\alpha_{0i}, \alpha_{1i}), \text{ for } i = 1, \dots, N\}$$

$$C_\gamma = \{(C_{1\gamma_i}, C_{2\gamma_i}, C_{3\gamma_i}), \text{ for } i = 1, \dots, N\}$$

$$C_\eta = \{(C_{1\eta_i}, C_{2\eta_i}, ..., C_{3\eta_i}), \text{ for } i = 1, \dots, N\}$$

$$C_\lambda = \{(C_{1\lambda_i}, C_{2\lambda_i}, ..., C_{3\lambda_i}), \text{ for } i = 1, \dots, N\}$$

Then, from Lunn et al. (2013), θ = parameters of interest: $\alpha, \beta^*, C_\gamma, C_\eta, C_\lambda$

λ = nuisance parameters: $\phi, \gamma, \eta, \lambda$

μ = mean of θ: $\beta^{(1)}, \beta^{(2)}, \bar{C}_\gamma, \bar{C}_\eta, \bar{C}_\lambda$

Σ = covariance matrix of θ: $\Sigma_{\beta(1)}, \Sigma_{\beta(2)}, \Sigma_{C_\gamma}, \Sigma_{C_\eta}, \Sigma_{C_\lambda}$

B.1   Full Model

The joint posterior distribution is given by,

$$p(\beta^{(1)}, \beta^{(2)}, \bar{C}_\gamma, \bar{C}_\eta, \bar{C}_\lambda, \Sigma_{\beta(1)}, \Sigma_{\beta(2)}, \Sigma_{C_\gamma}, \Sigma_{C_\eta}, \Sigma_{C_\lambda}, \alpha, \beta^*, C_\gamma, C_\eta, C_\lambda, \phi, \gamma, \eta, \lambda | y)$$

$$\propto p(\beta^{(1)}) p(\beta^{(2)}) p(C_\gamma) p(C_\eta) p(C_\lambda) p(\bar{C}_\gamma) p(\bar{C}_\eta) p(\bar{C}_\lambda) p(\Sigma_{\beta(1)}) p(\Sigma_{\beta(2)}) p(\Sigma_{C_\gamma}) p(\Sigma_{C_\eta}) p(\Sigma_{C_\lambda}) \times$$

$$\prod_{i=1}^{N} \Big\{ p(y_i | \alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}, \phi_i, \gamma_i, \eta_i, \lambda_i) p(\alpha_i | \beta^{(1)}, \Sigma_{\beta(1)}) \ p(\beta_i^* | \beta^{(2)}, \Sigma_{\beta(2)}) \times$$

$$p(C_{\gamma_i} | \bar{C}_\gamma, \Sigma_{C_\gamma}) p(C_{\eta_i} | \bar{C}_\eta, \Sigma_{C_\eta}) p(C_{\lambda_i} | \bar{C}_\lambda, \Sigma_{C_\lambda}) p(\phi, \gamma, \eta, \lambda) \Big\} \tag{B.1}$$

B.2   First Stage

We analyse all individuals independently from the joint posterior distribution of each $\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}$ conditional on y_i alone,

$$p(\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}, \phi_i, \gamma_i, \eta_i, \lambda_i | y_i) \propto \quad p(y_i | \alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}, \phi_i, \gamma_i, \eta_i, \lambda_i) \times$$

$$p(\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}) p(\phi_i, \gamma_i, \eta_i, \lambda_i) \tag{B.2}$$

B.3 Second Stage

From the distributions in (2.1) these are given by,

$$p(\beta^{(1)}, \beta^{(2)}, \bar{C}_\gamma, \bar{C}_\eta, \bar{C}_\lambda | \Sigma_{\beta(1)}, \Sigma_{\beta(2)}, \Sigma_{C_\gamma}, \Sigma_{C_\eta}, \Sigma_{C_\lambda}, \alpha, \beta^*, C_\gamma, C_\eta, C_\lambda, \phi, \gamma, \eta, \lambda, y)$$

$$\propto p(\beta^{(1)}, \beta^{(2)}, \bar{C}_\gamma, \bar{C}_\eta, \bar{C}_\lambda) \prod_{i=1}^{N} p(\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i} | \beta^{(1)}, \beta^{(2)}, \bar{C}_\gamma, \bar{C}_\eta, \bar{C}_\lambda,$$

$$\Sigma_{\beta(1)}, \Sigma_{\beta(2)}, \Sigma_{C_\gamma}, \Sigma_{C_\eta}, \Sigma_{C_\lambda}) \tag{B.3}$$

$$p(\Sigma_{\beta(1)}, \Sigma_{\beta(2)}, \Sigma_{C_\gamma}, \Sigma_{C_\eta}, \Sigma_{C_\lambda} | \beta^{(1)}, \beta^{(2)}, \bar{C}_\gamma, \bar{C}_\eta, \bar{C}_\lambda, \alpha, \beta^*_, C_\gamma, C_\eta, C_\lambda, \phi, \gamma, \eta, \lambda, y)$$

$$\propto p(\Sigma_{\beta(1)}, \Sigma_{\beta(2)}, \Sigma_{C_\gamma}, \Sigma_{C_\eta}, \Sigma_{C_\lambda}) \prod_{i=1}^{N} p(\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i} | \beta^{(1)}, \beta^{(2)}, \bar{C}_\gamma, \bar{C}_\eta, \bar{C}_\lambda,$$

$$\Sigma_{\beta(1)}, \Sigma_{\beta(2)}, \Sigma_{C_\gamma}, \Sigma_{C_\eta}, \Sigma_{C_\lambda}) \tag{B.4}$$

$$p(\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}, \phi_i, \gamma_i, \eta_i, \lambda_i | \beta^{(1)}, \beta^{(2)}, \bar{C}_\gamma, \bar{C}_\eta, \bar{C}_\lambda, \Sigma_{\beta(1)}, \Sigma_{\beta(2)}, \Sigma_{C_\gamma}, \Sigma_{C_\eta}, \Sigma_{C_\lambda}, y)$$

$$\propto p(y_i | \alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}, \phi_i, \gamma_i, \eta_i, \lambda_i, \phi_i, \gamma_i, \eta_i, \lambda_i) p(\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i} | \beta^{(1)}, \beta^{(2)},$$

$$\bar{C}_\gamma, \bar{C}_\eta, \bar{C}_\lambda, \Sigma_{\beta(1)}, \Sigma_{\beta(2)}, \Sigma_{C_\gamma}, \Sigma_{C_\eta}, \Sigma_{C_\lambda}) p(\phi_i, \gamma_i, \eta_i, \lambda_i)$$

$$i = 1, ..., N. \tag{B.5}$$

The distributions from (B.3) and (B.4) are available in closed form and can hence we can sample from them directly by using standard algorithms. For the distributions (B.5) we use the distributions in (B.2) as the proposal distributions within a Metropolis-Hastings step. For this, the target-to-proposal ratio can be simplified to,

$$C(\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}, \phi_i, \gamma_i, \eta_i, \lambda_i)$$

$$= \frac{p(y_i|\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}, \phi_i, \gamma_i, \eta_i, \lambda_i)}{p(y_i|\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}, \phi_i, \gamma_i, \eta_i, \lambda_i}) \times$$

$$\frac{p(\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}|\beta^{(1)}, \beta^{(2)}, \bar{C}_\gamma, \bar{C}_\eta, \bar{C}_\lambda, \Sigma_{\beta(1)}, \Sigma_{\beta(2)}, \Sigma_{C_\gamma}, \Sigma_{C_\eta}, \Sigma_{C_\lambda}) p(\phi_i, \gamma_i, \eta_i, \lambda_i)}{p(\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}) p(\phi_i, \gamma_i, \eta_i, \lambda_i)}$$

$$= \frac{p(\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i}|\beta^{(1)}, \beta^{(2)}, \bar{C}_\gamma, \bar{C}_\eta, \bar{C}_\lambda, \Sigma_{\beta(1)}, \Sigma_{\beta(2)}, \Sigma_{C_\gamma}, \Sigma_{C_\eta}, \Sigma_{C_\lambda})}{p(\alpha_i, \beta_i^*, C_{\gamma_i}, C_{\eta_i}, C_{\lambda_i})} \qquad \text{(B.6)}$$