Research Article

Statistics in Medicine

Received 13 February 2015, Accepted 5 January 2017 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.7235

Bayesian hierarchical modeling of longitudinal glaucomatous visual fields using a two-stage approach

Susan R. Bryan,[a,b*†] Paul H.C. Eilers,[a] Joost van Rosmalen,[a] Dimitris Rizopoulos,[a] Koenraad A. Vermeer,[b] Hans G. Lemij[c] and Emmanuel M.E.H. Lesaffre[a,d]

The Bayesian approach has become increasingly popular because it allows to fit quite complex models to data via Markov chain Monte Carlo sampling. However, it is also recognized nowadays that Markov chain Monte Carlo sampling can become computationally prohibitive when applied to a large data set. We encountered serious computational difficulties when fitting an hierarchical model to longitudinal glaucoma data of patients who participate in an ongoing Dutch study. To overcome this problem, we applied and extended a recently proposed two-stage approach to model these data. 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 (VFs) 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. Because 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.

Keywords: Bayesian modeling; hierarchical structure; longitudinal data analysis; two-stage approach

1. Introduction

Since the introduction of Markov chain Monte Carlo (MCMC) sampling by Gelfand and Smith [1] and the development of the BUGS software [2], 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 [3] and STAN [4]. While this newly developed 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*. [5] 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 extend the approach with an additional sampling step to allow for the calculation of model selection and model evaluation criteria.

[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*
[*]*Correspondence to: Susan Bryan, Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands.*
[†]*E-mail: s.bryan@erasmusmc.nl*

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, glaucoma is one of the leading causes of irreversible blindness in the world [6]. Adequate treatment may slow down the disease, possibly even halting its progression. Evaluation of a longitudinal series of visual fields (VFs), as measured by standard automated perimetry, provides a way to detect early evidence of glaucoma and to determine functional deterioration. However, because of the subjective nature of this technique, standard automated perimetry 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 statistical methods may have existed to model such data, the difficulties in extracting it from the device have made this type of data rare and hence has prevented much research on the topic.

The response variable of interest is the sensitivity estimate describing the level of differential light sensitivity at different locations within each eye. The sensitivity estimates are left-censored because of a limitation of the device. Models, which take into account this type of censoring, such as the Tobit model, have been described in the literature [7]. 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 four levels, namely, (i) the individual, (ii) the eye, (iii) the hemifield, and (iv) the location. There is a vast amount of literature that addresses hierarchical mixed effects models, for both frequentist [8] and Bayesian [9, 10] 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*. [5] 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 comparisons. 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 that 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. Note that for the statistical analysis of the data, we excluded one patient (Section 6) leading to $n = 138$ patients. Follow-up data were collected at approximately six 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 were used in this analysis have been made available online at http://rod-rep.com.

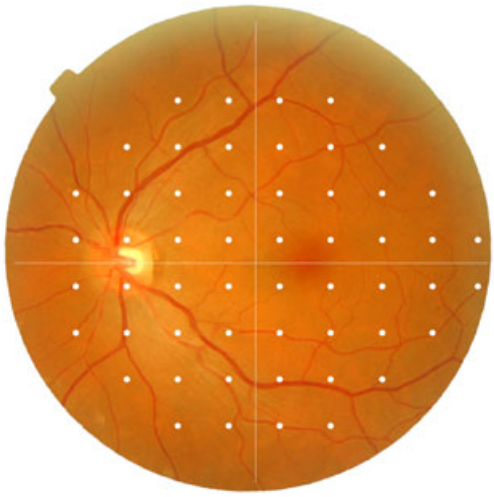

Figure 1. Fundus photo of a left human eye with the 54 test locations for the visual field test represented by white dots.

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 scale an attenuation x is defined as $s = 10\log_{10}(x)$, 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 [11].

2.2. *Previous research*

Parameters such as the mean deviation and visual field index summarize the 52 sensitivity estimates into single values that can be used by the clinicians when optimizing treatment strategies. Longitudinal modeling of these VF summary parameters has been performed before [12–15]. Modeling of individual test locations is potentially of greater interest, because it provides additional information such as the spatial nature of the fields that is otherwise lost in global parameters. In previous research, each location was analyzed as an independent sample [16–18]. 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 [8]. This was shown in the context of global VF measurements by Pathak *et al*. [19].

In glaucoma, variability is presumably related to fatigue effects and response errors, whereby sensitivity estimates decrease over time [20, 21]. Differences in fatigue effects, between the inferior and superior hemifields within an eye have been demonstrated [20]. Furthermore, this effect may differ between the first and second eye at the same visit. The number of false-negative answers has 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 [21, 22]. 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. 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 [15]. Although the effects of 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*. [23] found that this relationship is well represented by the function $\ln(SD) = A + B \times$

$\texttt{sensitivity}(dB)$, where A and B are 2.81 and -0.066 dB, respectively, for normal eyes and 3.62 and -0.098 dB for glaucomatous eyes. Russell *et al*. [22] 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 (because of glaucomatous damage), or relocation of the stimulus to the peripheral VF where there are fewer ganglion cells [24]. Zhu *et al*. [25] describe a method to detect change using an inferential statistical model that incorporates the nonstationary variability using a mixture of Weibull distributions.

Although there is a wide range of literature that discusses these aspects, the majority of previous work deals with the global indices or treats sensitivity estimates for each location in the VF 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, and 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 performed by building the model up sequentially, resulting in three different proposed models.

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 that defines the relationship between time and the latent, true sensitivity estimate is needed. The relationship between the observed sensitivity estimate y and the latent, true sensitivity estimate y^* is given by

$$y = \max(0, y^*).$$

3.2. Hierarchical model

We propose using a Bayesian hierarchical mixed effects model [9, 10] 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 four levels, namely, (i) the individual, (ii) the eye, (iii) the hemifield, and (iv) the location. The dependent variable in the model is y^*_{iehlt}, which denotes the latent sensitivity estimate for individual $i = 1, \ldots, N$; eye $e = 1, 2$; hemifield $h = 1, 2$; and location $l = 1, \ldots, 26$ at timepoint $t = 1, \ldots, T_i$. The independent variable in the model is $\texttt{time}_{it}$, which represents the time (in years) between tth measurement and the first measurement for each individual i, ranging from 0 to 10.5 years. Let $\mu_\alpha = (\mu_{\alpha_0}, \mu_{\alpha_1})^T$ correspond to the population-averaged intercept and slope of time. The individual-specific intercept and slope of time for individual i are represented by α_{0i} and α_{1i}, with deviations due to the eye given by γ_{0ie} and γ_{1ie}, due to the hemifield by η_{0ieh} and η_{1ieh}, and due to the location by λ_{0iehl} and λ_{1iehl}. We call this model 1.

Model 1:

$$\begin{aligned} y^*_{iehlt} &= \alpha_{0i} + \alpha_{1i}\texttt{time}_{it} + \gamma_{0ie} + \gamma_{1ie}\texttt{time}_{it} + \eta_{0ieh} + \eta_{1ieh}\texttt{time}_{it} + \lambda_{0iehl} + \\ &\quad \lambda_{1iehl}\texttt{time}_{it} + \epsilon_{iehlt} \\ &= \mu^{(1)}_{iehlt} + \epsilon_{iehlt}, \end{aligned} \tag{1}$$

where $\epsilon_{iehlt} \sim N(0, \sigma_\epsilon^2)$ and with priors:

$$\begin{aligned}
\alpha_i &= \begin{pmatrix} \alpha_{0i} \\ \alpha_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\alpha_0} \\ \mu_{\alpha_1} \end{pmatrix}, \Sigma_\alpha \right); \\
\gamma_{ie} &= \begin{pmatrix} \gamma_{0ie} \\ \gamma_{1ie} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_\gamma \right); \\
\eta_{ieh} &= \begin{pmatrix} \eta_{0ieh} \\ \eta_{1ieh} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_\eta \right); \\
\lambda_{iehl} &= \begin{pmatrix} \lambda_{0iehl} \\ \lambda_{1iehl} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_\lambda \right); \\
\mu_{\alpha_b} &\sim N(0, 10^9) \text{ for } b = 0, 1; \\
\sigma_\epsilon^2 &\sim \text{IG}(10^{-3}, 10^{-3}); \\
\Sigma_\alpha, \Sigma_\gamma, \Sigma_\eta, \text{ and } \Sigma_\lambda &\sim \text{IW}(\text{diag}(1, 1), 2).
\end{aligned}$$

3.3. *Global visit effect*

Montolio *et al*. [15] 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. Because 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 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. The GVE is an eye-specific visit effect. The term 'global' relates to the whole VF. 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 following model to capture the offset at the tth time point for eye e within individual i. This gives model 2.

Model 2:

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

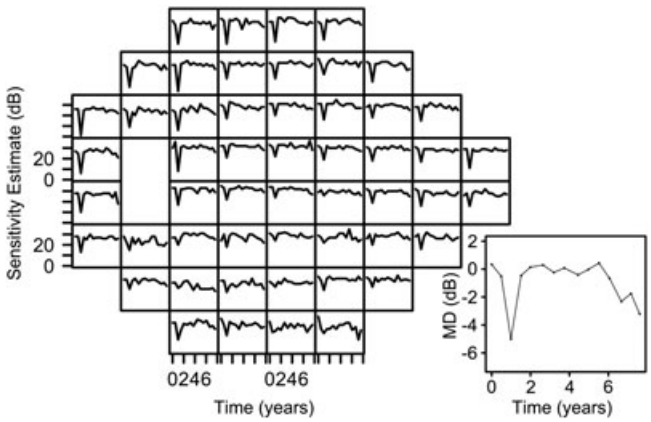

Figure 2. Retinal sensitivity estimates over time for each location of the visual field in the left eye of a single glaucoma patient. A decrease in the sensitivity estimates can be seen in all locations at around 1 year. The longitudinal profile of the mean deviation (MD) values over time is shown on the right. The visit-dependent decrease is also clear at around 1 year for the MD.

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 with the normal distribution and can handle heavy tails [26]. Hence, we let

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

where $t(\mu, \sigma^2, df)$ denotes the generalized t-distribution with mean μ, scale parameter σ^2, and df degrees of freedom. The prior for σ^2_{ϕ} was taken equal to IG(10^{-3}, 10^{-3}).

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, because 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 $\ln(SD) = A + B \times \texttt{sensitivity}(dB)$, where A and B are 2.60 and −0.06 dB, respectively. We extended model 2 to incorporate this relationship in model 3.

Model 3:

$$y^*_{iehlt} = \mu^{(2)}_{iehlt} + \epsilon_{iehlt},$$

with

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

and

$$\log(\sigma_{\epsilon,iehlt}) = \mu_{\zeta_0} + \mu_{\zeta_1}\mu^{(2)}_{iehlt}, \tag{3}$$

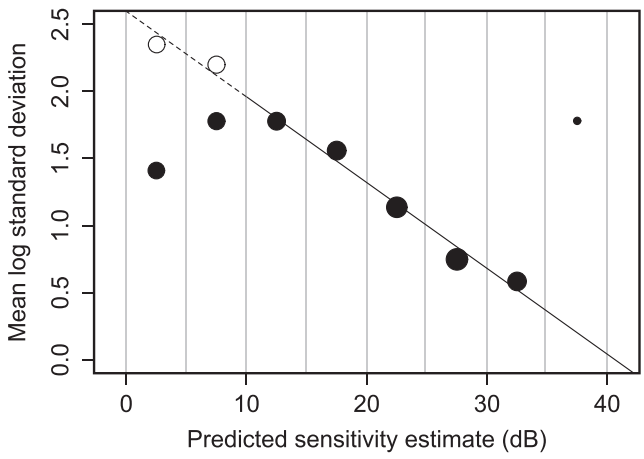

Figure 3. Bubble plot representing the mean logarithm of the standard deviation for different predicted sensitivity estimates determined using linear regression for each location. The predicted values were subdivided into groups with width 5 dB. The empty bubbles correspond to the hypothetical values, 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. Because of the small number of observations within the 35 to 40 dB range (0.02% of the total observations), we excluded these measurements in the linear regression.

Table I. Summary of parameters included cumulatively in each of the models.

Model	Parameter	Definition
1	y^*_{iehlt}	Latent sensitivity estimate
	μ_{α_0}	Population-averaged intercept
	μ_{α_1}	Population-averaged slope
	α_{0i}	Individual-specific intercept
	α_{1i}	Individual-specific slope
	γ_{0ie}	Deviation due to the eye (intercept)
	γ_{1ie}	Deviation due to the eye (slope)
	η_{0ieh}	Deviation due to the hemifield (intercept)
	η_{1ieh}	Deviation due to the hemifield (slope)
	λ_{0iehl}	Deviation due to the location (intercept)
	λ_{1iehl}	Deviation due to the location (slope)
	σ^2_ϵ	Variance of the measurement error
	Σ_α	Covariance matrix for the individual level
	Σ_γ	Covariance matrix for the eye level
	Σ_η	Covariance matrix for the hemifield level
	Σ_λ	Covariance matrix for the location level
2	ϕ_{iet}	Global visit effect
	σ^2_ϕ	Global visit effect variance
3	μ_{ζ_0}	Intercept in logarithm of the standard deviation
	μ_{ζ_1}	Slope in logarithm of the standard deviation
	$\sigma^2_{\epsilon,iehlt}$	Variance of the measurement error

where $\sigma_{\epsilon,iehlt}$ is the standard deviation for individual $i = 1, \ldots, N$; eye $e = 1, 2$; hemifield $h = 1, 2$; and location $l = 1, \ldots, 26$ at time $t = 1, \ldots, T_i$. The priors for μ_{ζ_0} and μ_{ζ_1} are both N(0, 10^9). For the other parameters in model 3, the priors are the same as those for models 1 and 2.

A summary of all parameters and their definitions for all the models is shown in Table I.

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 [5, 10]. 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 that needed to be estimated. As a consequence, we were unable to achieve convergence in a realistic time frame, and we experienced computer memory limitations when using WinBUGS or JAGS. For such situations, a computationally more tractable method is needed.

4.2. Two-stage approach

Lunn *et al*. [5] proposed two-stage Bayesian hierarchical modeling. 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*. [5] illustrated this method using models with two and three levels. The glaucoma data exhibit a similar hierarchical structure to that presented by Lunn *et al*. [5], but of a more complex nature with four levels. Figure 4 illustrates the hierarchical structure of the glaucoma data, as well as the cross-classified random effects, divided into two stages. 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.

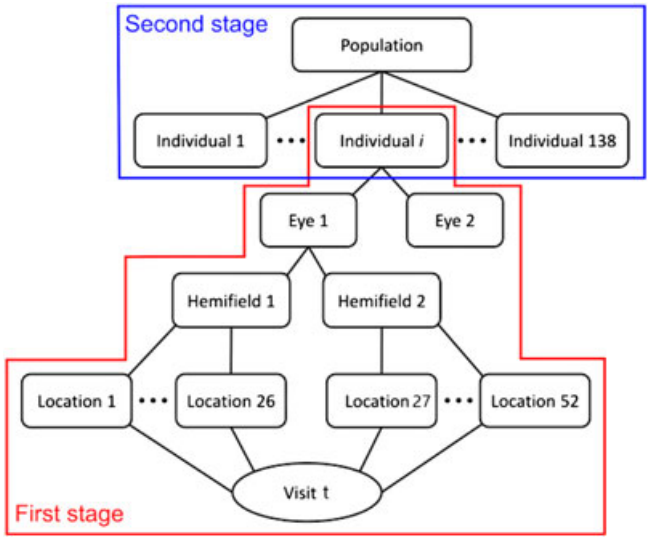

Figure 4. Illustration of the hierarchical structure of the data divided into the first and second stages as performed in the two-stage approach.

4.2.1. First stage. In the first stage, we analyzed each individual separately. This is now shown for model 3. The first stage model is given by:

$$\begin{aligned} y^*_{iehlt} = \alpha_{0i} + \alpha_{1i}\texttt{time}_{it} + \gamma_{0ie} + \gamma_{1ie}\texttt{time}_{it} + \eta_{0ieh} + \eta_{1ieh}\texttt{time}_{it} + \\ \lambda_{0iehl} + \lambda_{1iehl}\texttt{time}_{it} + \phi_{iet} + \epsilon_{iehlt}, \end{aligned} \tag{4}$$

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

$$\log(\sigma_{\epsilon,iehlt}) = \zeta_{0i} + \zeta_{1i}\mu^{(2)}_{iehlt},$$

and

$$\begin{aligned} \alpha_{bi} &\sim N(0, 10^9) \text{ for } b = 0, 1; \\ \gamma_{ie} &= \begin{pmatrix} \gamma_{0ie} \\ \gamma_{1ie} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{\gamma_i}\right); \\ \eta_{ieh} &= \begin{pmatrix} \eta_{0ieh} \\ \eta_{1ieh} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{\eta_i}\right); \\ \lambda_{iehl} &= \begin{pmatrix} \lambda_{0iehl} \\ \lambda_{1iehl} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{\lambda_i}\right); \\ \phi_{iet} &\sim t(0, \sigma^2_{\phi_i}, 3); \\ \sigma^2_{\phi_i} &\sim \text{IG}(10^{-3}, 10^{-3}); \\ \zeta_{bi} &\sim N(0, 10^9) \text{ for } b = 0, 1. \end{aligned}$$

Note that vague and independent priors are specified for α_{0i}, α_{1i}, because they are treated as fixed effects in the first stage.

One important detail about the two-stage approach is that it allows the individual variances to differ, that is, $\Sigma_{\gamma_i}, \Sigma_{\eta_i}, \Sigma_{\lambda_i}$, and $\sigma^2_{\phi_i}$, but also $\sigma^2_{\epsilon,iehlt}$ because the regression coefficients are now allowed to depend on the subject. In order to prevent the second-stage sampler from getting stuck near local posterior modes, large independent samples are needed from this first stage [5]. To achieve this, we ran 200,000 iterations with a burn-in of 150,000 iterations. Using a thinning factor of 10, this resulted in 5000 stored iterations for each parameter for each individual.

4.2.2. Second stage. At the end of the first stage, we have the following sampled values for the parameters split up according to whether they will be directly used in the second stage (parameters of interest) or not (nuisance parameters):

- Parameters of interest:
  $\{\alpha_{0i}, \alpha_{1i}, \zeta_{0i}, \zeta_{1i}, \Sigma_{\gamma_i}, \Sigma_{\eta_i}, \Sigma_{\lambda_i}, \sigma^2_{\phi_i}\}$ for $i = 1, \ldots, N$;
- Nuisance parameters:
  $\mathcal{N}_i = \{\gamma_{0i1}, \gamma_{1i1}, \gamma_{0i2}, \gamma_{1i2}, \eta_{0i1,1}, \ldots, \eta_{1i2,2}, \lambda_{0i1,1,1}, \ldots, \lambda_{1i2,2,26}, \phi_{i1,1}, \ldots, \phi_{i1,T_i}, \phi_{i2,1}, \ldots, \phi_{i2,T_i}\}$ for $i = 1, \ldots, N$.

The total set of nuisance parameters is then denoted as $\mathcal{N} = \{\mathcal{N}_1, \ldots, \mathcal{N}_{138}\}$. The split up into parameters of interest and nuisance parameters is motivated as follows. The $\alpha_i = (\alpha_{0i}, \alpha_{1i})^T$ and $\zeta_i = (\zeta_{0i}, \zeta_{1i})^T$ are transferred to the second stage, because we wish to estimate their population-averaged effects assuming a common (normal) distribution with means equal to $\mu_\alpha = (\mu_{\alpha_0}, \mu_{\alpha_1})^T$ and $\mu_\zeta = (\mu_{\zeta_0}, \mu_{\zeta_1})^T$, respectively. Because we decided to split the data at the individual level for computational reasons, all lower-level random effects are automatically considered to be nuisance. In the case of a meta-analysis, the random effects in the first stage may not be of direct interest. Hence, these terms can then be treated as nuisance parameters as performed by Lunn *et al*. [5]. However, for clinical applications such as ours, these may be important. So, we kept the covariance matrices of those lower-level random effects as input for the second stage to allow us to reestimate these random effects at a later stage. A Cholesky decomposition of the covariance matrices of the random effects (Σ_{γ_i}, Σ_{η_i}, and Σ_{λ_i}) was computed, because each of the entries in the covariance matrices were treated as separate parameters in the second stage. For example, for Σ_{γ_i}, the 2×2 full rank lower triangular matrix

$$C_{\gamma_i} = \begin{pmatrix} c_{1\gamma_i} & 0 \\ c_{2\gamma_i} & c_{3\gamma_i} \end{pmatrix},$$

with $c_{1\gamma_i}, c_{3\gamma_i}$ positive diagonal entries such that $C_{\gamma_i} C^T_{\gamma_i} = \Sigma_{\gamma_i}$. We denote the Cholesky decomposition factors for all of the covariance matrices as $\{c_{1\gamma_i}, c_{2\gamma_i}, c_{3\gamma_i}, c_{1\eta_i}, c_{2\eta_i}, c_{3\eta_i}, c_{1\lambda_i}, c_{2\lambda_i}, c_{3\lambda_i}\}$. Hence, the final set of parameters of interest that are directly processed in the second stage are $\{\alpha_i, \zeta_i, \sigma^2_{\phi_i}, c_{1\gamma_i}, c_{2\gamma_i}, c_{3\gamma_i}, c_{1\eta_i}, c_{2\eta_i}, c_{3\eta_i}, c_{1\lambda_i}, c_{2\lambda_i}, c_{3\lambda_i}\}$ for $i = 1, \ldots, N$.

In the second stage, we assume that the earlier parameters of interest share a common distribution. More specifically, we assumed here that $\alpha_i \sim N((\mu_{\alpha_0}, \mu_{\alpha_1})^T, \Sigma_\alpha)$, $\sigma^2_{\phi_i} \sim N(\sigma^2_\phi, \sigma^2_{\sigma^2_\phi})$. However, for ζ_{0i}, ζ_{1i}, and all Cholesky decomposition parameters, we assumed independent normal distributions, because of computational and memory restrictions. We refer to Section A.2.3 for the priors. The posterior distributions of the aforementioned parameters were then obtained by using their sampled values from the first stage as proposal distributions in a Metropolis–Hastings step. Further details about the two-stage approach are found in [5].

The total set of parameters involved in the second stage is denoted as $\mathcal{P}$ and is given by:

$$\mathcal{P} = \{\alpha_{01}, \alpha_{11}, \ldots, \alpha_{0N}, \alpha_{1N}, \mu_{\alpha_0}, \mu_{\alpha_1}, \Sigma_\alpha, \zeta_{01}, \zeta_{11}, \ldots, \zeta_{0N}, \zeta_{1N}, \mu_{\zeta_0}, \mu_{\zeta_1}, \sigma^2_{\zeta_0}, \sigma^2_{\zeta_1}, \sigma^2_\phi, \sigma^2_{\sigma^2_\phi}, \mu_{c_{1\gamma}}, \mu_{c_{2\gamma}}, \mu_{c_{3\gamma}},$$
$$\mu_{c_{1\eta}}, \mu_{c_{2\eta}}, \mu_{c_{3\eta}}, \mu_{c_{1\lambda}}, \mu_{c_{2\lambda}}, \mu_{c_{3\lambda}}, \sigma^2_{c_{1\gamma}}, \sigma^2_{c_{2\gamma}}, \sigma^2_{c_{3\gamma}}, \sigma^2_{c_{1\eta}}, \sigma^2_{c_{2\eta}}, \sigma^2_{c_{3\eta}}, \sigma^2_{c_{1\lambda}}, \sigma^2_{c_{2\lambda}}, \sigma^2_{c_{3\lambda}}\},$$

where $\{\mu_{c_{1\gamma}}, \mu_{c_{2\gamma}}, \mu_{c_{3\gamma}}, \mu_{c_{1\eta}}, \mu_{c_{2\eta}}, \mu_{c_{3\eta}}, \mu_{c_{1\lambda}}, \mu_{c_{2\lambda}}, \mu_{c_{3\lambda}}\}$ are defined in Section A.2.3.

Three chains were initialized with different starting values for all models determined from the first-stage samples. This was performed using the minimum, mean, and maximum sampled value for each parameter for every individual. Upon convergence, we computed the posterior mean, median, and standard deviation with the equal-tailed 95% credible interval for all parameters of interest.

5. Model evaluation

Classical approaches to model evaluation are applicable to the results of the first stage, because this stage represents a standard analysis. Hence, we evaluated the models at this stage for each individual separately using posterior predictive checks (PPCs), such as the χ^2 test statistic. We further compared the models after the second stage using the deviance information criterion (DIC) to determine the overall best model.

5.1. Deviance information criterion

In a Bayesian framework, a common tool for model evaluation is the DIC proposed by Spiegelhalter *et al*. [27]. The DIC is based on a contrast between the deviance at the posterior mean of the parameters

and the posterior mean deviance. DIC is defined here as

$$DIC = D(\bar{\mathcal{P}}, \bar{\mathcal{N}}) + 2p_D = \overline{D(\mathcal{P}, \mathcal{N})} + p_D, \tag{5}$$

with

$$p_D = \overline{D(\mathcal{P}, \mathcal{N})} - D(\bar{\mathcal{P}}, \bar{\mathcal{N}}),$$

where the deviance is denoted as $D(\mathcal{P}, \mathcal{N})$, the posterior expectation of the deviance is denoted as $\overline{D(\mathcal{P}, \mathcal{N})}$, and the posterior means of the parameters $\mathcal{P}$ and $\mathcal{N}$ as $\bar{\mathcal{P}}$ and $\bar{\mathcal{N}}$, respectively.

The definition of the DIC includes the parameters of interest $\mathcal{P}$, which we obtained from the second stage, as well as the random effects $\mathcal{N}$, which are treated as nuisance parameters in the two-stage approach. In practice, DIC is usually computed conditional on the random effects (so-called conditional DIC, see, e.g., [10]). However, the two-stage approach does not automatically provide estimates for all random effects under the one-stage model.

To address this problem, we propose to complement the two-stage approach with an additional step based on the method of composition [10] in combination with a Metropolis-within-Gibbs technique. To illustrate our proposed method, we refer again to model 3. Note that for the calculation of DICs, we need the posterior distribution $p(\mathcal{N} \mid y)$, with y the total set of observed VF data. The required posterior of the random effects can be obtained from the posterior distribution $p(\mathcal{P} \mid y)$ and $p(\mathcal{N} \mid \mathcal{P}, y)$ as

$$p(\mathcal{N} \mid y) = \int p(\mathcal{N} \mid \mathcal{P}, y) p(\mathcal{P} \mid y) d\mathcal{P}. \tag{6}$$

In practice, the integral is replaced by a Monte Carlo estimate, by sampling each element of $\mathcal{N}$ given y and the sampled value of $\mathcal{P}$ and then ignoring the sampled value $\mathcal{P}$. This is in fact an example of the method of composition.

Let the vector of all responses for the individual i and eye e be denoted by y_{ie}. Furthermore, let $\mathcal{N}_{ie}$ represent the subset of nuisance parameters in $\mathcal{N}$, for individual i and eye e. For each eye e of each individual i, the computations of our proposed method for sampling the random effects in $\mathcal{N}_{ie}$ consist of the following steps:

Step 1: We use the posterior results of $K = 500$ consecutive iterations of the second stage of the two-stage approach. For iteration k (with $k = 1, \ldots, K$), let $\tilde{\mathcal{P}}^{(k)}$ denote the sampled values of the parameters of interest ($\mathcal{P}$) estimated in the second stage. Conditional on $\tilde{\mathcal{P}}^{(k)}$, we sample the random effects in $\mathcal{N}_{ie}$ (for the eye, hemifield, location, and GVE), which were treated as nuisance parameters in the second stage, that is,

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

and we denote the sampled values as $\tilde{\mathcal{N}}_{ie}^{(k)}$. For each iteration k, the sampling of the random effects is performed with the Metropolis-within-Gibbs technique as described in steps 2A and 2B in the following.
Step 2A: For $k = 1$, we need to choose initial values for the random effects in $\tilde{\mathcal{N}}_{ie}^{(1)}$. To decrease computational time, the posterior modes of all parameters in $\mathcal{N}_{ie}$ were determined using an optimization routine, and these were used as initial values for $\tilde{\mathcal{N}}_{ie}^{(1)}$ in step 2B. The optimization was performed using the `optim()` function in R with a quasi-Newton algorithm. The function, which was maximized with respect to $\mathcal{N}_{ie}$, was the logarithm of $p(\mathcal{N}_{ie} \mid \bar{\mathcal{P}}, y_{ie}) \propto p(y_{ie} \mid \bar{\mathcal{P}}, \mathcal{N}_{ie})\, p(\mathcal{N}_{ie} \mid \bar{\mathcal{P}})$. Note that the optimization procedure is only performed for $k = 1$.
Step 2B: A random-walk Metropolis–Hastings algorithm is used to sample the random effects iteratively for each level, to take into account the correlation between the levels, yielding a Metropolis-within-Gibbs sampler. For example, first, a sampled value for the eye is obtained given the initial values for the levels lower than the eye (i.e., the hemifield, the location, and the GVE). This sampled value is then used in combination with the initial values for the levels lower than the hemifield (i.e., the location and the GVE) to obtain an estimate for the hemifield random effects and so forth. This Metropolis-within-Gibbs sampler is run for $s = 1, \ldots, S$ iterations, with $S = 200$, and the random effects obtained in the last iteration S are then used as the sampled random effects $\tilde{\mathcal{N}}_{ie}^{(k)}$ in step 1. The first $S - 1$ iterations thus

only serve as burn-in samples. For $k = 1$, the estimates obtained in step 2A are used as initial values (for $s = 1$). For $k = 2, \ldots, K$, the sampled values of the previous iteration of step 1, that is, $\tilde{\mathcal{N}}_{ie}^{(k-1)}$ are used as initial values.

In the Metropolis-within-Gibbs sampler, the random effects are sampled for each iteration s (with $s = 1, \ldots, S$) as follows:

1. The random effects for the eye level ($\gamma_{ie}^{(s)}$) are sampled using the Metropolis–Hastings algorithm with target density proportional to

$$\prod_h \prod_l \prod_t p(y_{iehlt} \mid \gamma_{ie}, \eta_{ieh}^{(s-1)}, \lambda_{iehl}^{(s-1)}, \phi_{iet}^{(s-1)}, \tilde{\mathcal{P}}^{(k)}) \times p(\gamma_{ie} \mid \tilde{\mathcal{P}}^{(k)}).$$

2. For hemifield $h = 1, 2$, the random effects for the hemifield level ($\eta_{ieh}^{(s)}$) are sampled using the Metropolis–Hastings algorithm with target density proportional to

$$\prod_l \prod_t p(y_{iehlt} \mid \gamma_{ie}^{(s)}, \eta_{ieh}, \lambda_{iehl}^{(s-1)}, \phi_{iet}^{(s-1)}, \tilde{\mathcal{P}}^{(k)}) \times p(\eta_{ieh} \mid \tilde{\mathcal{P}}^{(k)}).$$

3. For hemifield $h = 1, 2$ and location $l = 1, \ldots, 26$, the random effects for the location level ($\lambda_{iehl}^{(s)}$) are sampled using the Metropolis–Hastings algorithm with target density proportional to

$$\prod_t p(y_{iehlt} \mid \gamma_{ie}^{(s)}, \eta_{ieh}^{(s)}, \lambda_{iehl}, \phi_{iet}^{(s-1)}, \tilde{\mathcal{P}}^{(k)}) \times p(\lambda_{iehl} \mid \tilde{\mathcal{P}}^{(k)}).$$

4. For visit $t = 1, \ldots, T_i$, the random effects for the GVE ($\phi_{iet}^{(s)}$) are sampled using the Metropolis–Hastings algorithm with target density proportional to

$$\prod_h \prod_l p(y_{iehlt} \mid \gamma_{ie}^{(s)}, \eta_{ieh}^{(s)}, \lambda_{iehl}^{(s)}, \phi_{iet}, \tilde{\mathcal{P}}^{(k)}) \times p(\phi_{iet} \mid \tilde{\mathcal{P}}^{(k)}).$$

We let $\tilde{\mathcal{N}}^{(k)}$ denote all of the sampled values of $\mathcal{N}$, that is, the combined sampled values, $\tilde{\mathcal{N}}_{ie}^{(k)}$, for both eyes from all individuals. Combining the resulting samples of $\mathcal{N}$ with the samples of $\mathcal{P}$ from the second stage yields a set of $K = 500$ posterior samples of the parameters and random effects. These posterior samples are then used to calculate the deviance in iteration k as

$$D(\mathcal{P}, \mathcal{N})^{(k)} = -2 \sum_{i=1}^{N} \sum_{e=1}^{2} \sum_{h=1}^{2} \sum_{l=1}^{26} \sum_{t=1}^{T_i} \log(p(y_{iehlt} \mid \tilde{\mathcal{N}}^{(k)}, \tilde{\mathcal{P}}^{(k)})),$$

and finally, the DIC is calculated from these K deviance values.

5.2. *Posterior predictive check*

Let the vector of all responses for the ith individual be denoted by y_i. Furthermore, we denote all parameters for individual i from the first stage by ψ_i. Then suppose that $\psi_i^1, \ldots, \psi_i^K$ are samples from the converged Markov chain from $p(\psi_i \mid y_i)$. The posterior predictive p-value (PPP) for a discrepancy measure $D_G(y_i \mid \psi_i^k)$ is obtained by contrasting this value to $D_G(\tilde{y}_i^k, \psi_i^k)$ for k in $\{1, \ldots, K\}$, whereby the replicated data $\tilde{y}_i^k$ are sampled from $p(y_i \mid \psi_i^k)$. The following p-value can then be calculated for the ith subject

$$PPP = \frac{1}{K} \sum_{k=1}^{K} I[D_G(\tilde{y}_i^k, \psi_i^k) \leqslant D_G(y_i, \psi_i^k)]. \tag{7}$$

Here, the Gelman χ^2-test statistic [28] was used as the discrepancy measure to calculate the PPC for each individual. This is defined as

$$D_G(y_i, \psi_i^k) = \sum_{e=1}^{2} \sum_{h=1}^{2} \sum_{l=1}^{26} \sum_{t=1}^{T_i} \frac{[y_{iehlt} - E(y_{iehlt} \mid \psi_i^k)]^2}{\mathtt{var}(y_{iehlt} \mid \psi_i^k)}, \tag{8}$$

where a small value indicates a good model fit. When the PPP is small, say smaller than 0.05, the replicated data fit the assumed model importantly better than the observed data, and hence, we conclude in that case that our model does not fit the data well. In this hierarchical structure, such a p-value can be computed for each individual. In case of a good model fit, we expect that the distribution of p-values has a uniform distribution. Now, because the data is used twice: once for model fit and once for model evaluation (see, e.g., [10]), we will not obtain a uniform distribution for a good model fit, but rather a distribution more clustered around 0.5. For the same reason, the approach does not provide an absolute guarantee of a good model fit. But all improvements of the basic approach, see, for example, [10], are computationally too time-consuming here.

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, 276 eyes, 552 hemifields, and 14,352 locations remained. This included 4760 VFs, resulting in a data set consisting of 247,520 location-specific sensitivity estimates. All analyses were performed taking into account censoring.

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. An example of the model fits for one location is shown in Figure 5. The posterior summary statistics from the second stage are listed in Table II for each of the models. A lower DIC indicates a preferable model. Using DIC, model 2 (DIC = 1,173,872) performed better than model 1 (DIC = 1,211,653), with model 3 (DIC = 1,075,212 performing the best overall. Using the results from model 3, the population intercept (μ_{α_0}) was 19.82 dB with an average slope (μ_{α_1}) of −0.31 dB/year. The intercept (μ_{ζ_0}) and slope (μ_{ζ_1}) for the logarithm of the standard deviation was 2.82 and −0.08 dB, respectively. This corresponds to the 2.60 and −0.06 dB that was found in the exploratory analysis shown in Figure 3.

To evaluate the fit of the chosen model(s), we can use the PPC technique. More specifically, we could compute the PPP-value for each individual and then graphically evaluate the distribution of p-values. Figure 6 shows the ordered PPP-values for each of the models. Model 1 has a mean $PPP = 0.30$, model 2 a mean $PPP = 0.30$, and model 3 a mean $PPP = 0.50$. From this, it appears that model 3 has the best fit. That the distribution of p-values should deviate from the uniform distribution for a good model was explained earlier. In fact, we indeed see that for model 3, the p-values are more clustered around 0.5 than what would be expected if the true p-value had a uniform distribution. For the other two models, the distribution of p-values is systematically shifted to the left.

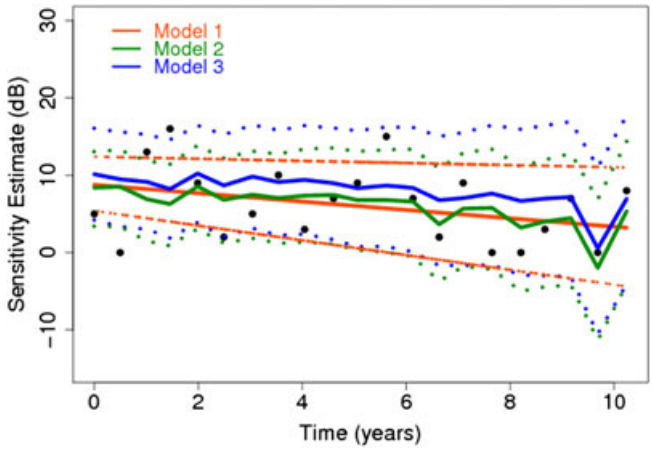

Figure 5. Scatter plot representing the retinal sensitivity estimates over time for one location of the VF. The lines represent the model fits for each of the three models with 95% credible intervals.

Table II. 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
μ_{α_0}	18.97	0.73(17.54 ; 20.39)	20.30	0.74(18.86 ; 21.76)	19.89	0.77(18.37 ; 21.39)
μ_{α_1}	−0.23	0.05(−0.33 ; −0.12)	−0.20	0.05(−0.30 ; −0.11)	−0.30	0.05(−0.40 ; 0.20)
μ_{ζ_0}					2.82	0.06 (2.69 ; 2.95)
μ_{ζ_1}					-0.08	0.00(−0.08 ; −0.07)
$\sigma^2_{\epsilon,iehlt}$	13.46	0.66(12.20 ; 14.81)	11.53	0.57(10.45 ; 12.70)		
σ^2_{ϕ}			1.86	0.15 (1.58 ; 2.18)	5.64	0.12 (5.29 ; 5.91)
Σ_{α}	$\begin{pmatrix} 63.79 & 0.06 \\ 0.06 & 0.20 \end{pmatrix}$		$\begin{pmatrix} 71.11 & -1.73 \\ -1.73 & 0.16 \end{pmatrix}$		$\begin{pmatrix} 76.10 & -0.14 \\ -0.14 & 0.17 \end{pmatrix}$	
Σ_{γ}	$\begin{pmatrix} 0.44 & -0.02 \\ -0.02 & 0.23 \end{pmatrix}$		$\begin{pmatrix} 0.45 & -0.01 \\ -0.01 & 0.27 \end{pmatrix}$		$\begin{pmatrix} 0.00 & 0.01 \\ 0.01 & 0.96 \end{pmatrix}$	
Σ_{η}	$\begin{pmatrix} 26.85 & 0.36 \\ 0.36 & 0.23 \end{pmatrix}$		$\begin{pmatrix} 24.73 & 0.19 \\ 0.19 & 0.23 \end{pmatrix}$		$\begin{pmatrix} 25.50 & 0.29 \\ 0.29 & 0.23 \end{pmatrix}$	
Σ_{λ}	$\begin{pmatrix} 41.41 & 0.40 \\ 0.40 & 0.13 \end{pmatrix}$		$\begin{pmatrix} 40.33 & -0.19 \\ -0.19 & 0.14 \end{pmatrix}$		$\begin{pmatrix} 40.13 & 0.11 \\ 0.11 & 0.09 \end{pmatrix}$	
DIC	1,211,653		1,173,872		1,075,212	

CI, credible interval; DIC, deviance information criterion.

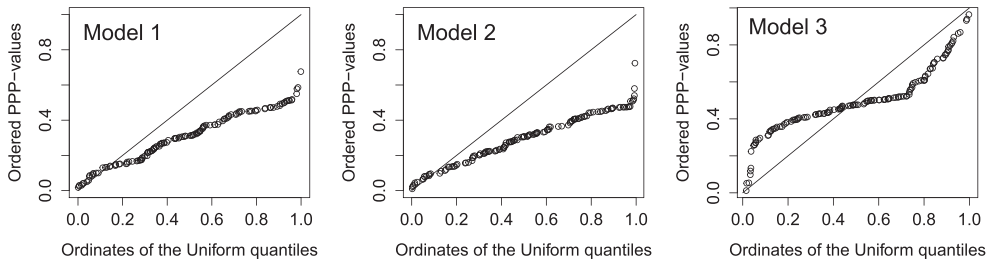

Figure 6. Posterior predictive check for each of the models across all individuals.

6.2. Comparing the models: some reflections

Our model takes into account the hierarchical structure of the data using a Bayesian hierarchical model. This allows us to take into account the correlation at each of the four levels. The estimated covariance matrices for each level were similar for all models. A large proportion of intercept variability was explained by the individual level random effects (model 3: $\Sigma_{\alpha 11} = 76.10$). A small amount of this intercept variability was explained by the eye level random effects (model 3: $\Sigma_{\gamma 11} = 0.0007$). This is because that each individual only has two eyes and with this approach, we assume that only the eyes belonging to the same individual share a common variance distribution. This variability increases for the four hemifields per individual (model 3: $\Sigma_{\eta 11} = 25.50$) and for the 104 locations per individual (model 3: $\Sigma_{\lambda 11} = 40.13$). The variance of the slope for the hemifield model 3: $\Sigma_{\eta 22} = 0.23$) seems to be large compared with the variance of the slope for the location (model 3: $\Sigma_{\lambda 22} = 0.09$). This shows that locations within a hemifield are correlated with respect to their change. This is expected, because the nerves within each hemifield belong to the same nerve bundle. Hence, it is likely that there is some correlation between nerves, and hence locations, which are affected by glaucomatous damage. It is important to note that all individuals received therapy throughout the study. The range of slopes may be reduced with therapy, because the intraocular pressure (IOP) range is reduced between eyes to somewhat similar levels. Without therapy, the individual's IOP would be between 15 and 40 mmHg. With therapy, this could be reduced to between 12 and 24 mmHg. Because all individuals in this study are being treated, the reduction in the IOP range could explain the small range of slopes of progression.

We proposed to model measurement errors that affect the point-wise sensitivity estimates within the same VF as GVEs. By correcting for the GVEs, we account for measurable factors, such as season, time of day, and the reliability indices, as well as those factors that cannot be measured such as fatigue and delayed reaction time. Including the GVE showed a considerable 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. Implementing progression models that incorporate the

GVE in clinical care may therefore improve the clinical management of glaucoma. A further evaluation of the GVE by determining the improvement of the model fit due to incorporating the GVE in the model and the effect of including the GVE on estimating the rate of progression has been performed in previous work. Furthermore, the improvement of point-wise predictions for future measurements accounting for the GVE has been shown [29].

Including the relationship between variability and sensitivity showed a further improvement in the model fit. The function, which describes this relationship, was consistent to that found by Henson *et al*. [23]; however, it was not shown previously how to include this relationship in a model, or whether including it would improve the model fit. Other studies have emphasized the importance of this relationship and questioned the validity of assuming a constant variance over the entire range of sensitivity estimates [22]. Gardiner *et al*. [30] showed that clinical VF testing may be unreliable when VF locations have sensitivity estimates below approximately 15–19 dB. By modeling this relationship, we are able to account for the lower sensitivity estimates being less reliable and hence allow the precision throughout the range of sensitivity estimates to differ accordingly.

In clinical practice, it is often difficult to determine which patients progress rapidly towards severe visual impairment and which ones are relatively stable, because of measurement variability. With our current approach, the true rate of progression is much better determined, so that clinicians can specifically target those with rapid progression, thereby significantly reducing the risk of severe visual impairment or blindness.

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 simplified 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 [31], WinBUGS, or OpenBUGS [5] software and the second-stage using OpenBUGS software [5]. However, for the second stage, an add-on program is needed. For more details on setting up OpenBUGS for performing the two-stage analyses, we refer to [5]. More information regarding the computations performed 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 allowed 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-stage 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 DIC. In order to calculate the DIC and compare different competing models for our data fitted using the two-stage approach, we proposed 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 fibers in the eye [32].

Appendix A. Model specifications

In this section, the model specifications for model 3 are given for the one-stage and two-stage approaches.

A.1. One-stage approach

A.1.1. Full model.

$$\begin{aligned} y^{*}_{iehlt} = {} & \alpha_{0i} + \alpha_{1i}\texttt{time}_{it} + \gamma_{0ie} + \gamma_{1ie}\texttt{time}_{it} + \\ & \eta_{0ieh} + \eta_{1ieh}\texttt{time}_{it} + \lambda_{0iehl} + \lambda_{1iehl}\texttt{time}_{ij} + \phi_{iet} + \epsilon_{iehlt}, \end{aligned} \tag{A.1}$$

where $\epsilon_{iehlt} \sim N(0, \sigma^2_{\epsilon,iehlt})$ and $\log(\sigma_{\epsilon,iehlt}) = \mu_{\zeta_0} + \mu_{\zeta_1}\mu^{(2)}_{iehlt}$.

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.

$$
\begin{aligned}
\alpha_i = \begin{pmatrix} \alpha_{0i} \\ \alpha_{1i} \end{pmatrix} &\sim N\left(\begin{pmatrix} \mu_{\alpha_0} \\ \mu_{\alpha_1} \end{pmatrix}, \Sigma_\alpha \right) \\
\gamma_{ie} = \begin{pmatrix} \gamma_{0ie} \\ \gamma_{1ie} \end{pmatrix} &\sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_\gamma) \\
\eta_{ieh} = \begin{pmatrix} \eta_{0ieh} \\ \eta_{1ieh} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_\eta \right) \\
\lambda_{iehl} = \begin{pmatrix} \lambda_{0iehl} \\ \lambda_{1iehl} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_\lambda \right) \\
\phi_{iet} &\sim t(0, \sigma^2_\phi, 3) \\
\mu_{\zeta_b} &\sim N(0, 10^9) \text{ for } b = 0, 1 \\
\mu_{\alpha_b} &\sim N(0, 10^9) \text{ for } b = 0, 1 \\
\sigma^2_\phi &\sim IG(10^{-3}, 10^{-3}) \text{ and} \\
\Sigma_\alpha, \Sigma_\gamma, \Sigma_\eta \text{ and } \Sigma_\lambda &\sim \text{IW}(\text{diag}(1, 1), 2).
\end{aligned}
$$

A.2. Two-stage approach

The full model is the same for both the one-stage and two-stage approaches. Priors for μ_α and α_i are also the same for both approaches. However, the covariance matrices for γ_{ie}, η_{ieh}, and λ_{iehl}, the scale parameter for ϕ_{iet} as well as ζ_i are subject-specific in the two-stage approach.

A.2.1. Full model priors.

$$
\begin{aligned}
\alpha_i = \begin{pmatrix} \alpha_{0i} \\ \alpha_{1i} \end{pmatrix} &\sim N\left(\begin{pmatrix} \mu_{\alpha_0} \\ \mu_{\alpha_1} \end{pmatrix}, \Sigma_\alpha \right) \\
\gamma_{ie} = \begin{pmatrix} \gamma_{0ie} \\ \gamma_{1ie} \end{pmatrix} &\sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{\gamma_i}) \\
\eta_{ieh} = \begin{pmatrix} \eta_{0ieh} \\ \eta_{1ieh} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{\eta_i} \right) \\
\lambda_{iehl} = \begin{pmatrix} \lambda_{0iehl} \\ \lambda_{1iehl} \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_{\lambda_i} \right) \\
\phi_{iet} &\sim t(0, \sigma^2_{\phi_i}, 3) \\
\zeta_{bi} &\sim N(\mu_{\zeta_b}, \sigma^2_{\zeta_b}) \text{ for } b = 0, 1 \\
\mu_{\alpha_b} &\sim N(0, 10^9) \text{ for } b = 0, 1 \\
\sigma^2_{\phi_i} &\sim N(\sigma^2_\phi, \sigma^2_{\sigma^2_\phi}) \\
\sigma^2_\phi &\sim \text{IG}(10^{-3}, 10^{-3}) \\
\sigma^2_{\sigma^2_\phi} &\sim \text{IG}(10^{-3}, 10^{-3}) \\
\mu_{\zeta_b} &\sim N(0, 10^9) \text{ for } b = 0, 1 \\
\sigma^2_{\zeta_b} &\sim \text{IG}(10^{-3}, 10^{-3}) \text{ for } b = 0, 1 \text{ and} \\
\Sigma_{\alpha_i}, \Sigma_{\gamma_i}, \Sigma_{\eta_i} \text{ and } \Sigma_{\lambda_i} &\sim \text{IW}(\text{diag}(1, 1), 2).
\end{aligned}
$$

A.2.2. First-stage priors.

$$\begin{aligned}
\alpha_{bi} &\sim N(0, 10^9) \text{ for } b = 0, 1 \\
\sigma^2_{\phi_i} &\sim \text{IG}(10^{-3}, 10^{-3}) \\
\zeta_{bi} &\sim N(0, 10^9) \text{ for } b = 0, 1 \\
\Sigma_{\alpha_i}, \Sigma_{\gamma_i}, \Sigma_{\eta_i} \text{ and } \Sigma_{\lambda_i} &\sim \text{IW}(\text{diag}(1, 1), 2).
\end{aligned}$$

A.2.3. Second-stage priors.

$$\begin{aligned}
\alpha_i = \begin{pmatrix} \alpha_{0i} \\ \alpha_{1i} \end{pmatrix} &\sim N(\begin{pmatrix} \mu_{\alpha_0} \\ \mu_{\alpha_1} \end{pmatrix}, \Sigma_\alpha) \\
\sigma^2_{\phi_i} &\sim N(\sigma^2_\phi, \sigma^2_{\sigma^2_\phi}) \\
\zeta_{bi} &\sim N(\mu_{\zeta_b}, \sigma^2_{\zeta_b}) \text{ for } b = 0, 1 \\
\mu_{\alpha_b} &\sim N(0, 10^8) \text{ for } b = 0, 1 \\
\Sigma_\alpha &\sim \text{IW}(\text{diag}(1, 1), 2) \\
\sigma^2_\phi &\sim \text{IG}(10^{-3}, 10^{-3}) \\
\sigma^2_{\sigma^2_\phi} &\sim \text{IG}(10^{-3}, 10^{-3}) \\
\mu_{\zeta_b} &\sim N(0, 10^9) \text{ for } b = 0, 1 \\
\sigma^2_{\zeta_b} &\sim \text{IG}(10^{-3}, 10^{-3}) \text{ for } b = 0, 1.
\end{aligned}$$

Using Cholesky decomposition, Σ_{γ_i}, Σ_{η_i}, and Σ_{λ_i} become

$$\begin{aligned}
c_{r\gamma_i} &\sim N(\mu_{cr\gamma}, \sigma^2_{cr\gamma}) \text{ for } r = 1, 2, 3 \\
c_{r\eta_i} &\sim N(\mu_{cr\eta}, \sigma^2_{cr\eta}) \text{ for } r = 1, 2, 3 \\
c_{r\lambda_i} &\sim N(\mu_{cr\lambda}, \sigma^2_{cr\lambda}) \text{ for } r = 1, 2, 3 \\
\mu_{cr\gamma}, \mu_{cr\eta}, \mu_{cr\lambda} &\sim N(0, 10^9) \text{ for } r = 1, 2, 3 \\
\sigma^2_{cr\gamma}, \sigma^2_{cr\eta}, \sigma^2_{cr\lambda} &\sim \text{IG}(10^{-3}, 10^{-3}) \text{ for } r = 1, 2, 3.
\end{aligned}$$

Appendix B. Methods

In this section, the different stages within the two-stage approach are described in detail. Again, we use model 3 for illustration purposes. Let

$$\begin{aligned}
\alpha_i &= \{\alpha_{01}, \alpha_{11}, \ldots, \alpha_{0,N}, \alpha_{1,N}\} \\
\zeta_i &= \{\zeta_{01}, \zeta_{11}, \ldots, \zeta_{0,N}, \zeta_{1,N}\} \\
C_{\gamma_i} &= \{c_{1\gamma_1}, c_{2\gamma_1}, c_{3\gamma_1}, \ldots, c_{1\gamma_N}, c_{2\gamma_N}, c_{3\gamma_N}\} \\
C_{\eta_i} &= \{c_{1\eta_1}, c_{2\eta_1}, c_{3\eta_1}, \ldots, c_{1\eta_N}, c_{2\eta_N}, c_{3\eta_N}\} \\
C_{\lambda_i} &= \{c_{1\lambda_1}, c_{2\lambda_1}, c_{3\lambda_1}, \ldots, c_{1\lambda_N}, c_{2\lambda_N}, c_{3\lambda_N}\}.
\end{aligned}$$

To simplify the notation, we let C_i denote all of the Cholesky decomposition parameters in C_{γ_i}, C_{η_i}, and C_{λ_i}. In the spirit of Lunn *et al.*, we are interested in the parameters of interest from the first stage, namely, $\alpha_i, \zeta_i, C_i, \sigma^2_{\phi_i}$. The population averages and covariance matrices for α_i are given by μ_α and Σ_α. The population averages and variances for $\zeta_i, C_i, \sigma^2_{\phi_i}$ are given by $\mu_\zeta, \mu_c, \sigma^2_\phi$ and $\sigma^2_\zeta, \sigma^2_c, \sigma^2_{\sigma^2_\phi}$, respectively, where

$$\begin{aligned}
\mu_\alpha &= \{\mu_{\alpha_{01}}, \mu_{\alpha_{02}}\} \\
\mu_\zeta &= \{\mu_{\zeta_{01}}, \mu_{\zeta_{02}}\} \\
\mu_c &= \{\mu_{c1\gamma}, \mu_{c2\gamma}, \mu_{c3\gamma}, \mu_{c1\eta}, \mu_{c2\eta}, \mu_{c3\eta}, \mu_{c1\lambda}, \mu_{c2\lambda}, \mu_{c3\lambda}\}
\end{aligned}$$

and

$$\sigma_\zeta^2 = \{\sigma_{\zeta_{01}}^2, \sigma_{\zeta_{02}}^2\}$$
$$\sigma_c^2 = \{\sigma_{c1\gamma}^2, \sigma_{c2\gamma}^2, \sigma_{c3\gamma}^2, \sigma_{c1\eta}^2, \sigma_{c2\eta}^2, \sigma_{c3\eta}^2, \sigma_{c1\lambda}^2, \sigma_{c2\lambda}^2, \sigma_{c3\lambda}^2\}.$$

B.1. Full model

The joint posterior distribution for model 3 using the two-stage approach is given by

$$\begin{aligned} &p(\mu_\alpha, \mu_\zeta, \mu_c, \sigma_\phi^2, \Sigma_\alpha, \sigma_\zeta^2, \sigma_c^2, \sigma_{\sigma_\phi^2}^2, \alpha_1, \ldots, \alpha_N, \zeta_1, \ldots, \zeta_N, C_1, \ldots, C_N, \sigma_{\phi_1}^2, \ldots, \sigma_{\phi_N}^2, \mathcal{N} \mid y) \propto \\ &p(\mu_\alpha)p(\mu_\zeta)p(\mu_c)p(\sigma_\phi^2)p(\Sigma_\alpha)p(\sigma_\zeta^2)p(\sigma_c^2)p(\sigma_{\sigma_\phi^2}^2)\prod_{i=1}^{N}\Big\{p(y_i \mid \alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2, \mathcal{N}_i)p(\alpha_i \mid \mu_\alpha, \Sigma_\alpha) \times \\ &p(\zeta_i \mid \mu_\zeta, \sigma_\zeta^2)\ p(C_i \mid \mu_c, \sigma_c^2)p(\sigma_{\phi_i}^2 \mid \sigma_\phi^2, \sigma_{\sigma_\phi^2}^2)p(\mathcal{N}_i)\Big\}. \end{aligned} \tag{B.1}$$

B.2. First stage

We analyze all individuals independently from the joint posterior distribution of each α_i, ζ_i, C_i conditional on y_i alone,

$$\begin{aligned} p(\alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2, \mathcal{N}_i \mid y_i) &\propto p(y_i \mid \alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2, \mathcal{N}_i)p(\alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2)p(\mathcal{N}_i) \\ &i = 1, \ldots, N. \end{aligned} \tag{B.2}$$

B.3. Second stage

From distribution (B.1), the posterior distributions for the second stage are given by

$$\begin{aligned} p(\mu_\alpha, \mu_\zeta, \mu_c, \sigma_\phi^2 \mid \Sigma_\alpha, \sigma_\zeta^2, \sigma_c^2, \sigma_{\sigma_\phi^2}^2, \alpha_1, \ldots, \alpha_N, \zeta_1, \ldots, \zeta_N, C_1, \ldots, C_N, \sigma_{\phi_1}^2, \ldots, \sigma_{\phi_N}^2, \mathcal{N}, y) \propto \\ p(\mu_\alpha, \mu_\zeta, \mu_c, \sigma_\phi^2)\prod_{i=1}^{N} p(\alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2 \mid \mu_\alpha, \mu_\zeta, \mu_c, \sigma_\phi^2, \Sigma_\alpha, \sigma_\zeta^2, \sigma_c^2, \sigma_{\sigma_\phi^2}^2) \end{aligned} \tag{B.3}$$

$$\begin{aligned} p(\Sigma_\alpha, \sigma_\zeta^2, \sigma_c^2, \sigma_{\sigma_\phi^2}^2 \mid \mu_\alpha, \mu_\zeta, \mu_c, \sigma_\phi^2, \alpha_1, \ldots, \alpha_N, \zeta_1, \ldots, \zeta_N, C_1, \ldots, C_N, \sigma_{\phi_1}^2, \ldots, \sigma_{\phi_N}^2, \mathcal{N}, y) \propto \\ p(\Sigma_\alpha, \sigma_\zeta^2, \sigma_c^2, \sigma_{\sigma_\phi^2}^2)\prod_{i=1}^{N} p(\alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2 \mid \mu_\alpha, \mu_\zeta, \mu_c, \sigma_\phi^2, \Sigma_\alpha, \sigma_\zeta^2, \sigma_c^2, \sigma_{\sigma_\phi^2}^2) \end{aligned} \tag{B.4}$$

$$\begin{aligned} p(\alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2, \mathcal{N}_i \mid \mu_\alpha, \mu_\zeta, \mu_c, \sigma_\phi^2, \Sigma_\alpha, \sigma_\zeta^2, \sigma_c^2, \sigma_{\sigma_\phi^2}^2, y) \propto \\ p(y_i \mid \alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2, \mathcal{N}_i)p(\alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2 \mid \mu_\alpha, \mu_\zeta, \mu_c, \sigma_\phi^2, \Sigma_\alpha, \sigma_\zeta^2, \sigma_C^2, \sigma_{\sigma_\phi^2}^2)p(\mathcal{N}_i) \\ i = 1, \ldots, N. \end{aligned} \tag{B.5}$$

The distributions from (B.3) and (B.4) are available in closed form, and 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 the Metropolis–Hastings algorithm, the target-to-proposal ratio, denoted as R, can be simplified to

$$\begin{aligned} R(\alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2, \mathcal{N}_i) &\propto \frac{p(y_i \mid \alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2, \mathcal{N}_i)}{p(y_i \mid \alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2, \mathcal{N}_i)} \times \\ &\quad \frac{p(\alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2 \mid \mu_\alpha, \mu_\zeta, \mu_c, \sigma_\phi^2, \Sigma_\alpha, \sigma_\zeta^2, \sigma_c^2, \sigma_{\sigma_\phi^2}^2)p(\mathcal{N}_i)}{p(\alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2)p(\mathcal{N}_i)} \\ &= \frac{p(\alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2 \mid \mu_\alpha, \mu_\zeta, \mu_c, \sigma_\phi^2, \Sigma_\alpha, \sigma_\zeta^2, \sigma_c^2, \sigma_{\sigma_\phi^2}^2)}{p(\alpha_i, \zeta_i, C_i, \sigma_{\phi_i}^2)} \\ &\quad i = 1, \ldots, N. \end{aligned} \tag{B.6}$$

For more details regarding the two-stage approach, we refer to [5].

Acknowledgements

The authors would like to thank Dr. Baoyue Li for his useful discussions that helped with the early stages of this work. The authors would also like to thank Stichting Wetenschappelijk Onderzoek het Oogziekenhuis, Stichting voor Ooglijders, and Stichting Glaucoomfonds for the financial support. Furthermore, the authors would like to thank the editor and the referees for their helpful comments and suggestions that drastically improved the quality of the manuscript.

References

1. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 1990; **85**(410):398–409.
2. Lunn D, Spiegerhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions. *Statistics in Medicine* 2009; **28**(25):3049–3067.
3. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2009; **71**(2):319–392.
4. Hoffman MD, Gelman A. The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 2014; **15**(1):1593–1623.
5. Lunn D, Barrett J, Sweeting M, Thompson S. Fully Bayesian hierarchical modelling in two stages, with application to meta-analysis. *Journal of the Royal Statistical Society. Series C, Applied Statistics* 2013; **62**(4):551–572.
6. Kingman S. Glaucoma is second leading cause of blindness globally. *Bulletin of the World Health Organization* 2004; **82**(11):887–888.
7. Tobin J. Estimation of relationships for limited dependent variables. *Econometrica* 1958; **26**(1):24–36.
8. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data* 1st edn., Springer Series in Statistics. Springer New York: New York, NY, 2000.
9. Ntzoufras I. *Bayesian Modeling Using WinBUGS* 1st edn. Wiley & Sons: Hoboken, N. J, 2009.
10. Lesaffre E, Lawson AB. *Bayesian Biostatistics* 1st edn. Wiley & Sons: Chichester, West Sussex, 2012.
11. Anderson DR, Patella VM. *Automated Static Perimetry* 2nd edn. Mosby: St Louis, 1999.
12. Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression. *American Journal of Ophthalmology* 2008; **145**(2):343–353.
13. Artes PH, Nicolela MT, LeBlanc RP, Chauhan BC. Visual field progression in glaucoma: total versus pattern deviation analyses. *Investigative Ophthalmology & Visual Science* 2005; **46**(12):4600–4606.
14. Kymes SM, Lambert DL, Lee PP, Musch DC, Siegfried CJ, Kotak SV, Stwalley DL, Fain J, Johnson C, Gordon MO. The development of a decision analytic model of changes in mean deviation in people with glaucoma: the COA model. *Ophthalmology* 2012; **119**(7):1367–1374.
15. Junoy Montolio FG, Wesselink C, Gordijn M, Jansonius NM. Factors that influence standard automated perimetry test results in glaucoma: test reliability, technician experience, time of day, and season. *Investigative Ophthalmology & Visual Science* 2012; **53**(11):7010–7017.
16. McNaught AI, Crabb DP, Fitzke FW, Hitchings RA. Modelling series of visual fields to detect progression in normal-tension glaucoma. *Graefe's Archive for Clinical and Experimental Ophthalmology* 1995; **233**(12):750–755.
17. Caprioli J, Mock D, Bitrian E, Afifi AA, Yu F, Nouri-Mahdavi K, Coleman A L. A method to measure and predict rates of regional visual field decay in glaucoma. *Investigative Ophthalmology & Visual Science* 2011; **52**(7):4765–4773.
18. Bryan SR, Vermeer KA, Eilers PHC, Lemij HG, Lesaffre EMEH. Robust and censored modeling and prediction of progression in glaucomatous visual fields. *Investigative Ophthalmology & Visual Science* 2013; **54**(10):6694–6700.
19. Pathak M, Demirel S, Gardiner SK. Nonlinear, multilevel mixed-effects approach for modeling longitudinal standard automated perimetry data in glaucoma. *Investigative Ophthalmology & Visual Science* Aug 2013; **54**(8):5505–5513.
20. Hudson C, Wild JM, O'Neill EC. Fatigue effects during a single session of automated static threshold perimetry. *Investigative Ophthalmology & Visual Science* 1994; **35**(1):268–280.
21. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Investigative Ophthalmology & Visual Science* 2000; **41**(8):2201–2204.
22. Russell RA, Crabb DP, Malik R, Garway-Heath DF. The relationship between variability and sensitivity in large-scale longitudinal visual field data. *Investigative Ophthalmology & Visual Science* 2012; **53**(10):5985–5990.
23. Henson DB, Chaudry S, Artes P H, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Investigative Ophthalmology & Visual Science* 2000; **41**(2): 417–421.
24. Swanson WH, Sun H, Lee BB, Cao D. Responses of primate retinal ganglion cells to perimetric stimuli. *Investigative Ophthalmology & Visual Science* 2011; **52**(2):764–771.
25. Zhu H, Russell RA, Saunders LJ, Ceccon S, Garway-Heath DF, Crabb DP. Detecting changes in retinal function: analysis with non-stationary Weibull error regression and spatial enhancement (ANSWERS). *PLoS ONE* 2014; **9**(1):e85654.
26. Lee KJ, Thompson SG. Flexible parametric models for random-effects distributions. *Statistics in Medicine* 2008; **27**(3):418–434.
27. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002; **64**(4):583–639.
28. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis* 3rd edn. Chapman and Hall/CRC: Boca Raton, 2013.
29. Bryan SR, Eilers PHC, Lesaffre EMEH, Lemij HG, Vermeer KA. Global visit effects in point-wise longitudinal modeling of glaucomatous visual fields. *Investigative Ophthalmology & Visual Science* 2015; **56**(8):4283–4289.

30. Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. *Ophthalmology* 2014; **121**(7):1359–1369.
31. Plummer M. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*, Vol. 124, 2003.
32. Erler NS, Bryan SR, Eilers PHC, Lesaffre EMEH, Lemij HG, Vermeer KA. Optimizing structure–function relationship by maximizing correspondence between glaucomatous visual fields and mathematical retinal nerve fiber models. *Investigative Ophthalmology & Visual Science* 2014; **55**(4):2350–2357.

Supporting information

Additional supporting information may be found online in the supporting information tab for this article.