Research Article

Investigating Determinants of Multiple Sclerosis in Longitunal Studies: A Bayesian Approach

Clelia Di Serio[1] and Claudia Lamina[2]

[1] *University Centre of Statistics in the Biomedical Sciences, (CUSSB), Vita-Salute University, Milan, Italy*
[2] *Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology,
 Innsbruck Medical University, Innsbruck, Austria*

Correspondence should be addressed to Clelia Di Serio, diserio.clelia@hsr.it

Modelling data from Multiple Sclerosis longitudinal studies is a challenging topic since the phenotype of interest is typically ordinal; time intervals between two consecutive measurements are nonconstant and they can vary among individuals. Due to these unobservable sources of heterogeneity statistical models for analysis of Multiple Sclerosis severity evolve as a difficult feature. A few proposals have been provided in the biostatistical literature (Heijtan (1991); Albert, (1994)) to address the issue of investigating Multiple Sclerosis course. In this paper Bayesian P-Splines (Brezger and Lang, (2006); Fahrmeir and Lang (2001)) are indicated as an appropriate tool since they account for nonlinear smooth effects of covariates on the change in Multiple Sclerosis disability. By means of Bayesian P-Spline model we investigate both the randomness affecting Multiple Sclerosis data as well as the ordinal nature of the response variable.

1. Introduction

Multiple Sclerosis (MS) is a progressive neurological disorder classified among complex diseases. Investigating MS causes and potential triggers is a difficult task since the clinical manifestations and course vary considerably. Therefore longitudinal studies, both clinical trial as well as natural history studies, become crucial to assess the disease evolution over time. How to measure MS phenotype has been a major problem [1, 2] due to the multifactorial nature of the disease. No unique criteria are provided in the literature for classifying MS patients with respect to different MS courses. Estimating the MS incidence rate is also a difficult feature due to the variability of MS symptoms and the potentially long duration of the latent disease period [3]. These considerations, among others, offer an insight about the numerous problems arising in measuring MS. The disease markers that are used in the MS literature are typically related either to impairments of functional status

(FS) or to dissemination of lesions. This latter measurement, though is not the object of our analysis, is becoming crucial to measure early disease activity. This class of measures includes magnetic resonance imaging (MRI), cerebrospinal fluid (CSF), and visual lesions. MRI data present the advantage to be countinuos variable, unlike clinical measurements referred to disability which are typically ordinal. Thus, whereas MRI data can still be modelled within a time series approach (see Albert [4]), disability measurements cannot. Indeed, to describe MS impairments data classical statistical models may fail due to (i) the difficult definition of the outcome variable, (ii) a large number of individual observations at a small number of time points, (iii) the high inter-individual variability. This paper is based on a multicenter database built within a worldwide research program and collecting untreated patients drawn from natural history studies as well as placebo patients sampled from major therapeutic studies conducted either by academic research groups or by the pharmaceutical industry(The program has been established at the Sylvia Lawry Centre for Multiple Sclerosis in 2001 supported by MSIF). The data set used includes 897 patients selected from 17 placebo controlled clinical trials and it is mainly aimed at the better understanding of the determinants of MS course in order to improve the efficiency of therapies for MS patients. The patients were included according to the McDonald diagnostic criteria [5]. In fact, data from clinical trials allow for a good monitoring and comparable time spans (1 to 4 months) between two subsequent observations; whereas natural history data do not. However these data are heterogenous in their structure at different levels: (i) time intervals between consecutive measurements as well as number of observations can vary considerably among individuals.

This paper proposes a new statistical perspective to model both the longitudinal nature of the data as well as the individual heterogeneity. The novelty is due both to the statistical framework chosen to model these data and to the random variable used to describe MS evolution (see Section 2). We use generalized additive mixed effect models (GAMMs) [6–9]. In general terms these models join the mixed effect models principles [10–13] together with the generalized additive models theory (GAMs) [14]. The basic idea in mixed effect models is that you want to incorporate not only population-specific effects that are constant among all individuals, the so-called fixed effects, but also subject-specific characteristics through random component. Thus, unobservable heterogeneity among individuals is included by means of random effects. Covariates effect on the responses is modelled by particular nonlinear smooth functions, a Bayesian P-Splines. This approach is proposed as a suitable tool to investigate the MS data structure and understand the role of prognostic factors in affecting the disease course.

In Section 2 some background on MS terminology and the related variables is provided. In Section 3 we describe the general features of generalized additive models. A Bayesian version of P-Splines [15] is introduced within the simplest case of random intercept model. The extension to a random slope model is illustrated within the applied MS framework (Section 4). We address the outcome variable both as a Gaussian as well as an ordinal response. In Section 4 these statistical tools are applied to the MS data set. Comparisons between the analyses performed with the different models and the discussion on the results are provided in Section 5.

2. EDSS Weighted Change as a Measure of MS Disability

Multiple Sclerosis (MS) is a chronic progressive disease that affects the brain and spinal cord (central nervous system). This disease is classified among the multifactorial genetic diseases

(or complex diseases); the causes and potential triggers of MS are thought to be based both on genetic predisposition and on biological and environmental patients characteristics. The variability of the MS symptoms and the potentially long duration of the latent period of the disease from onset make MS extremely difficult to measure. As mentioned above, the disease markers used in MS literature to measure disease activity are typically related either to impairments of functional status or to dissemination of lesions. This latter, which is not the object of our analysis, are becoming crucial to measure early disease activity. In this class of measures are included magnetic resonance imaging (MRI), cerebrospinal fluid (CSF), and visual lesions. In this paper we consider as outcome variable the degree of functional disability usually measured by the so-called *Kurtzke Expanded Disability Status Scale (EDSS)* [16].

The Kurtzke Expanded Disability Status Scale (EDSS) is a method of measuring disability in multiple sclerosis. This scale quantifies disability in eight Functional Systems (FSs) and allows neurologists to assign a Functional System Score (FSS) in each of these. (the Functional Systems are: pyramidal, cerebellar, brainstem, sensory, bowel and bladder, visual, cerebral, and other.) EDSS steps 1.0 to 4.5 refer to people with MS who are fully ambulatory. EDSS steps 5.0 to 9.5 are defined by the impairment to ambulation. The value 10 represents death due to MS. The EDSS has many shortcomings such as its nonlinearity and its discontinuity. Common ways to overcome data related problems mentioned above is to put MS data in survival analysis frameworks, modelling the time to a certain EDSS level (4.0 or 6.0), or time to worsening, defined by an increase of 1 point in EDSS. Dynamic approaches have been developed [17] to aim at early predictions of MS progression by means of dynamic MCMC. However these survival models may not be the best tool whenever the analysis focuses on modelling the EDSS course over time since the hazard function $h(t)$ is treated only at points where the failure occurs; a lot of available information is again lost, as measurements between the first observation and the reaching of the event are neglected. Thus, a survival framework does not really address the longitudinal nature of the data.

To investigate MS evolution we introduce a new variable "EDSS change" that we model over time. This is the ratio between two subsequent EDSS measurements. In addition, since higher EDSS values (such as EDSS = 5.5) are dominated by a serious loss in ambulation we have weighted the EDSS change to reflect the degree of severity in EDSS change. For instance the changes in EDSS values with EDSS more than 5.5 have been weighted twice as much than changes below this level. Thus, this weighted change "changew$_{EDSS}$" is a measure of severity in changes of disability but it is only conceptually related to the original EDSS values (it ranges from −3.5 to 9.5 and takes 25 values).

In the paper we model the "changew$_{EDSS}$" according to two different perspectives: (i) "changew$_{EDSS}$" is considered as a continuous variable. A Gaussian mixed effect model is therefore investigated; (ii) "changew$_{EDSS}$" values collapse in 5 categories, according to the severity of disease change (see Table 11 in appendix). An ordinal threshold model is applied.

3. The Model

In the statistical literature (Pinheiro and Bates (2000), [18–20]) it has been seen how mixed effects models provide a flexible and powerful tool for the analysis of repeated measures data. They are intuitively appealing in biomedical frameworks. Fixed effects are associated to an average population trend that is constant among all individuals; whereas random effects
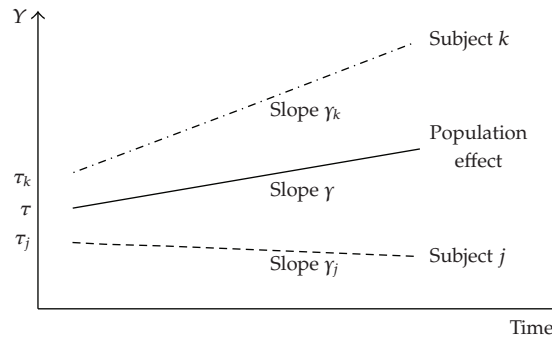
Figure 1: The random slopes model.

account for how the individual randomly deviates from the population trend. Therefore, a primary goal of this modelling is to investigate how large is the variance component associated to random effects in comparison to the residual variance [21]. In mixed effect modelling any number of random effects can be specified. Though, identifiably problems and computational complications may arise when introducing too many random components. The type and number of random effects are clearly related to the focus of the analysis to the extent that they are chosen to model the most important sources of unobserved heterogeneity. In the previously presented MS setting a high portion of unexplained variation is commonly thought to depend on the initial EDSS level and on the intensity of progression. Therefore it is reasonable in our modelling to allow for both the intercept as well as the slope of evolution profiles of each patient to vary randomly. In practice, a *random-intercepts model* is achieved by assigning to each subject a random effect. In clinical terms this modelling can be restrictive because it supports the hypothesis that the initial MS severity affects the MS course with a random impact; whereas patients are thought to have the same profile as regarding the MS progression. This means that these models require the slope coefficients to be equal for each subject.

A random slope model is needed to allow the intensity of evolution to vary among subjects since the coefficient of one or more explanatory variable varies randomly across higher-level units. Thus, in a longitudinal setting, the evolution profiles for each subject have specific intercepts and slopes (see Figure 1). In these models the between-subjects variance is a quadratic function of the covariates. The source of the increasing variability is within units rather than between. More details on variance structure for this kind of models are provided in the literature [13].

The GAMMs are here adopted to investigate through mixed effects modelling the MS data structure within a nonparametric Bayesian framework. This is done by modelling the dependence between the response variable Y and the explanatory variables $x_1, x_2, \ldots, x_p$ by means of a smooth function $f_j(x_j), j = 1, 2, \ldots, p$. With these types of modelling we aim at exploring and discovering unknown trend in MS data. One advantage of using smooth functions is that the functional form is directly drawn from the data leading to an estimate of the trend which reduces the variability of Y. The shape of each covariate effect is datadriven. The results can then be used to suggest a parametric form for the effect of covariates when modelling is needed for prediction purposes. This approach is flexible enough to allow for investigating within the same class of modelling the behaviour

of "changew$_{\text{EDSS}}$" when taken as continuous as well as ordinal response. The linear predictor is assumed to be a sum of smooth functions and has the form

$$\eta = \nu + f_1(x_1) + \cdots + f_p(x_p).$$ (3.1)

Expression (3.1) highlights a basic concept of GAMM, that is, the assumption of additivity of effects. Starting from this assumption it is possible to retain the interpretability of the familiar linear model and to model some predictors with smooth functions $f(x)$ and others with constant parameters. In principle, any known smoother can be used to estimate $f_j(x_j)$, such as polynomial smoothing Splines or regression Splines. In general, Splines offer a compromise between an interpolation of the data and a global smooth by representing the fit as a piecewise defined function. The pieces on the interval $[\xi_0, \xi_m]$ are separated by a sequence of knots $\xi_0 < \xi_1 < \cdots < \xi_m$. The partial functions B_i's, called basis functions, are fitted to the data within the range of two subsequent knots. They are restricted to follow a set of smoothing conditions with the neighboring basis functions at the breakpoints.

3.1. P-Splines

In this paper we deal with a particular class of smooth functions out of the big set of Splines, the P-Splines [22]. These are based on the traditional assumption that the effect f of a covariate x on the response can be approximated by a linear combination of Basic Splines Curves (B-Splines) which are a popular choice for basis functions due to their numerical stable behavior. Let $\Psi = \{\xi_i\}$, $i \in Z$ be a set of knots with $\xi_i < \xi_{i+1}$, $\xi_i \to -\infty$ for $i \to -\infty$ and $\xi_i \to \infty$ for $i \to \infty$. B-Splines depend only on the degree k and the values of Ψ. They are nonzero functions in a defined interval and zero outside of this interval. The Splines curve $s \in S_k(\Psi)$ can then be described as a linear combination of the separate B-Splines and their coefficients β_i as

$$s(x) = \sum_{i=-1}^{m+k} \beta_i B_i(x), \quad x \in [\xi_0, \xi_m].$$ (3.2)

P-Splines are introduced to address a crucial problem in Splines theory, that is, the choice of the number and the position of knots. In fact, to allow for flexibility in capturing the variability of the data structure, a large number of knots are recommended. Nevertheless, this may lead to overfitting. To address this issue we start noticing that the coefficients $\beta_i's$ in (3.2) can be considered as a measure of the basis amplitude since they regulate the roughness of the curves. The higher the difference between adjacent $\beta_i's$ is, the rougher the curve is. The approach of Eilers and Marx [23] relies on this idea. The authors proposed to overfit the data with a relatively large number of knots but to restrict the high variation of the curve by using a difference penalty on coefficients of adjacent B-Splines. Consider the regression of N data points (x_j, y_j) on a set of $r = m + k - 1$ B-Splines B_i. As a penalty we use the integral of the squared second derivative of the form

$$P(\lambda) = \lambda \int_{x_{\min}}^{x_{\max}} \left[\sum_{i=1}^{r} \beta_i B_j''(x) \right]^2 dx.$$ (3.3)

The parameter λ controls for the smoothness of the function continuously, therefore represents the smoothing parameter. Since minimization with this term is numerically complicated, it is approximated by a simple difference penalty based on the lth differences (Δ^l) of adjacent B-Spline-coefficients $\lambda \sum_{i=l+1}^{r} (\Delta^l \beta_i)^2$. This procedure leads to minimization of the term $\sum_{i=1}^{N}(y_i - \sum_{i=1}^{r} \beta_i B_i^k(x))^2 + \lambda \sum_{i=l+1}^{r} (\Delta^l \beta_i)^2$. The substitution of the integrated square of the lth derivative with the corresponding difference reduces the dimensionality of this problem from the number of observations N to the number of B-Splines r. This approach allows to combine the opposite requisites of the modelling, that is, enough flexibility without a large overfitting, a relatively large number of equally distant knots are suggested. The high variation of the curves is then reduced by penalizing the likelihood with a l difference penalty term (Δ^l) on adjacent B-Splines coefficients given by

$$P(\lambda) = \lambda \sum_{i=l+1}^{m+k} \left(\Delta^l \beta_i\right)^2. \tag{3.4}$$

The Fisher-Scoring Algorithm is used to conduct the maximization on the Penalized Likelihood with respect to the unknown regression coefficients. The smoothness of the function is regulated by the smoothing parameters $\lambda_j, j = 1, \ldots, p$. The method recommended by Eilers and Marx is to minimize the Akaike information criterion (AIC). Details about this criterion can be found in Hastie and Tibshirani [14]. The computation of AICs for many values of λ is time consuming and becomes quite impracticable in higher dimensions. Furthermore, the function AIC(λ) does not need to have a global minimum and it has been proved to show often several local minima, which makes it difficult to decide on one optimal λ value. It has been shown [15], that even in cases of a unique minimum, the choice of λ is not optimal, to the extent that it produces a curve which describes a poor approximation of the phenomenon of interest. Alternatives to AIC are cross-validation methods [18].

In the next subsections we use a Bayesian version of P-Splines in the case of a Gaussian as well as ordinal responses [24]. Bayesian P-Splines represent an alternative model to elude the problem of choosing an optimal λ value. Indeed, within this approach the assumption of a constant smoothing parameter is not needed, which can be inappropriate in complex situations where f is highly oscillating as well as rapidly changing.

3.2. Bayesian P-Splines and Mixed Effects Models

A Bayesian version for P-Splines has the advantage of allowing for simultaneous estimation of smooth functions and smoothing parameters. It can easily be extended to complex formulations like mixed effect models. This is a flexible way to use P-Spline since no constant smoothing parameters are assumed and they are locally adaptive. This can be very useful in MS context, where the smooth function may change curvature. Inference is fully Bayesian using MCMC simulation technique to draw sample from the posterior. In the Bayesian approach the unknown P-splines parameters $\beta_i's$ are model at another level in the hierarchy of the overall Bayesian model through a distributional assumption. Previous knowledge of the parameters, if available, can be used to define this prior distribution and to estimate them simultaneously together with the other model parameters. In many situations however, whenever no previous knowledge about the parameters is provided a *diffuse prior* distribution is assigned; $\beta \propto$ const indicates that each value of the parameter has the same probability.
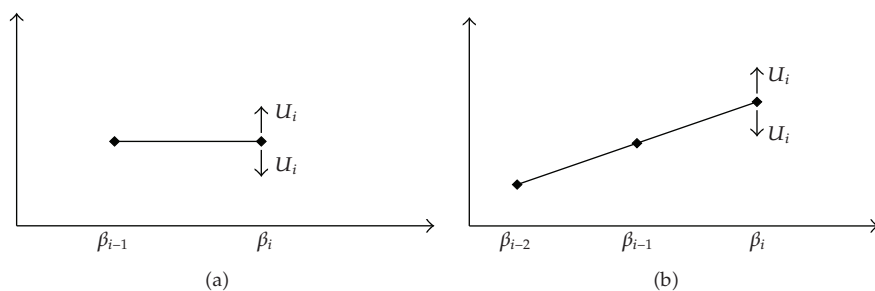
Figure 2: Prior distribution for RW(1) (a) and RW(2) (b).

Classical P-Splines are based on the lth differences (Δ^l) of adjacent B-Spline-coefficients $\lambda \sum_{i=l+1}^{r} (\Delta^l \beta_i)^2$. The unknown parameters $\beta_i, i = 1,\ldots,r$ are now considered as random variables and therefore, it is necessary to elicit prior distributions. Following Fahrmeir and Lang [7, 8] the difference penalty term $P(\lambda)$ is now replaced by their stochastic analogues: a *random walks*. For instance, first and second differences penalty terms correspond to first- and second-order random walks, given respectively by

$$\text{RW}(1) : \beta_i = \beta_{i-1} + u_i,$$

$$u_i \sim N\left(0, \tau^2\right), \ \beta_1 \propto \text{const}$$

$$\text{RW}(2) : \beta_i = 2 * \beta_{i-1} - \beta_{i-2} + u_i,$$

$$u_i \sim N\left(0, \tau^2\right), \ \beta_1 \propto \text{const}, \ \beta_2 \propto \text{const}.$$

$$(3.5)$$

Note that priors in (3.5) could be defined equivalently by specifying the conditional distributions of a particular parameter β_{kj}, given the left and right neighbours. Then the conditional mean can be interpreted as locally linear or quadratic fits at the kjth knot position.

This concept is intuitively illustrated in Figure 2. The coefficient β_i is restricted to deviate at most by u_i from the preceding coefficient β_{i-1}, or alternatively from the interpolating line between β_{i-2} and β_{i-1}, in the case of a second-order random walk.

In this case the joint distribution of the prior is given by $\beta \propto \exp(-(1/2\tau^2)\beta' \Gamma \beta)$ with Γ being the symmetric penalty matrix. In addition to the coefficients, the variance parameter τ regulating the smoothness of the function has to be supplemented with a prior distribution as well and it is estimated by means of the single-component Metropolis-Hastings Algorithm. The advantage of this procedure is that the problem of choosing a smoothing parameter is avoided. The variance parameter τ corresponds to the smoothing parameter λ in the classical approach of P-Splines, but in this Bayesian procedure it is not *datadriven* and therefore more reliable than λ. For this last variance parameter we make use of a weakly informative inverse gamma prior $\tau^2 \sim \text{IG}(a,b)$ that is $p(\tau^2) \propto (1/(\tau^2)^{a+1}) \exp(b/\tau^2)$ with small hyperparameters $a = b$ (e.g., with values 0.005), (see, e.g., Lang and Brezger [15]). Note that the asymptotic scenario implies that the number of components in τ^2 increases with growing sample size. This is, apparently, a nonstandard setting in the full Bayesian model as the parameter space changes with the sample size.

To relate the full Bayesian setting to these results a set of coherent convergence assumptions is needed on τ^2 [25]. Notice that other priors can be chosen. The presented choices are referred to a context where no a priori knowledge is assumed and are taken also for the sake of mathematical tractability. We next describe how fixed as well as random effects of the covariates can be easily included in a GAMM within a Bayesian perspective.

3.2.1. Bayesian P-Splines with a Gaussian Response

Suppose that repeated measurements have been taken on n individuals and a mixed effects model is used. Bayesian P-Splines are considered to model the nonparametric effect of the covariates on a Gaussian response. Fixed effects are included in the model additively with respect to the random effect and the P-Splines components. Bayesian P-Splines within a *random intercept model* are given by

$$y_i = b_i + \sum_{j=1}^{p} f_j(x_{ij}) + \sum_{k=1}^{K} \alpha_k w_{ik} + \varepsilon_i, \tag{3.6}$$

where $b_i, i = 1, 2, \ldots, N$, is the random intercept. $f_j's, j = 1, 2, \ldots, p$, denote the Bayesian P-Splines and model the nonparametric effect of p individual metric covariates $x_{ij}'s$ on the response y_i. The $\alpha_k's$ are the fixed effects parameters of the K individual population-specific covariates $w_i's$.

In a Bayesian context, in addition to the above discussed variance component for the random walk regulating the smoothness of the P-Splines, prior distributions are assigned to all the parameters in expression (3.6). They are commonly chosen as follows

(1) *Residual Variance Component.* $\varepsilon_i \sim N(\mu_i, \sigma^2)$ with σ^2 being the scale parameter. An inverse gamma distribution is commonly assigned as $\sigma^2 \sim \text{IG}(a_\sigma, b_\sigma)$. Setting a_σ to 1 and b_σ to small values we obtain a weakly informative distribution, as afore mentioned.

(2) *Variance Component for the Random Effects.* $b_i's, i = 1, \ldots, n$, are generally assumed to be *i.i.d.* Gaussian, $b_i \sim N(0, \tau_{ra}^2)$. Similar to the hyperparameter in the random walk approach, the variance parameter $\text{Var}(b_i) = \tau_{ra}^2$ is assumed to be random. Again these are usually inverse gamma distributed, so that $\tau_{ra}^2 \sim \text{IG}(a_{ra}, b_{ra})$ with $a_{ra} = 1$ and $b_{ra} = 0.005$.

(3) *Fixed Effects. diffuse priors* are chosen to express no prior knowledge about the fixed effects parameters.

We remark that in this framework two assumptions are required: (i) conditional independence of $y_i's$ given the covariates; (ii) mutual independence of the prior distributions for variance components and fixed effects. Inference procedures are based on Bayesian techniques to estimate the *posterior* distribution functions. Commonly *posteriors* are intractable and MCMC methods [26] are required to draw random samples from the posterior distribution. However the aforementioned elicitation of priors allows to overcome these computational problems since the full conditional distributions of α, b are multivariate

Gaussian; whereas the full conditionals of τ^2, τ_{ra}^2, and σ^2 are all inverse gamma distributions. Since all distributions are known, a simple Gibbs sampler can be used to update the parameters of the model either in single component steps or blockwise. A detailed updating algorithm and mean and variance parameters of the full conditionals can be found in Lang and Brezger [15].

3.2.2. Bayesian P-Splines with an Ordinal Response

The threshold model is based on the idea that the observable variable Y is merely a categorized version of a latent continuous variable $\tilde{Y}$ explained by the regressors in the linear form $\tilde{Y} = -X'\delta + \varepsilon$ with nuisance parameter $E(\varepsilon) = 0$. The relationship between Y and $\tilde{Y}$ is then expressed by

$$Y = r \iff \theta_{r-1} < \tilde{Y} \le \theta_r \tag{3.7}$$

with $-\infty = \theta_0 < \theta_1 < \cdots < \theta_k = \infty$ for $r = 1, \ldots, k$ categories. That means when the latent variable lies between the boundaries θ_{r-1} and θ_r the observable variable takes the value r. The distribution function F of the nuisance parameter ε naturally influences the appearance of the model. Common choices are the logistic or the normal distribution (*ordered probit model*). A detailed description of the cumulative threshold model can be found in Tutz [27].

The additive mixed effect model with an ordinal response does not differ from (3.6) except for the meaning of the latent response variable $\tilde{Y}$. The full conditional distribution of the latent variable is a truncated standard normal distribution, with truncation points determined by the thresholds as

$$P\left(Y \mid \tilde{Y}\right) = \sum_{r=1}^{k} I\left(\theta_{r-1} < \tilde{Y} \le \theta_r\right) P(Y = r) \tag{3.8}$$

with I the indicator function for the latent variable being between two subsequent categories.

Drawing out a truncated normal distribution evolves as numerically difficult and almost not solvable together with random effects. Thus, reparametrization strategies are used to overcome the numerical problems [18]. Furthermore, when we assume $\tilde{Y}$ to be the underlying latent variable of the ordinal response Y with thresholds $-\infty = \theta_0 < \theta_1 < \cdots < \theta_k = \infty$, then $k - 1$ parameters are to be estimated in addition to the unknown coefficient parameters. In a Bayesian approach the thresholds $\theta = (\theta_1, \ldots, \theta_{k-1})'$ are considered as random and supplemented with diffuse priors as well as the fixed effects.

4. P-Splines Model to Investigate MS Clinical Prognostic Factors

In this section we show how Bayesian P-Splines with mixed effects constitute an efficient method to model the heterogeneity in MS clinical data and to state the role of covariates in determining the severity of the disease. The covariates included are chosen among the most important prognostic factors in MS (Table 10 in appendix) according to the diagnostic criteria provided by McDonald et al. [5]. The following models are investigated

(1) P-Splines random intercept model:

      (i) with a Gaussian response,

     (ii) with an ordinal response.

(2) P-Splines random slope model:

      (i) with a Gaussian response,

     (ii) with an ordinal response.

Results from the latter modelling are here omitted since they do not really add additional information for interpreting the covariates role.

4.1. P-Splines Random Intercept Model with a Gaussian Response

The influence of the covariates on changew_{EDSS} is estimated with the Bayesian techniques. For the metric variables, a Bayesian P-Splines of degree 3 and a second-order random walk penalty were considered. For the benefit of estimating a smooth function for time, a random slope term has been left out. Thus, possible nonlinear effects of time may be detected. The further introduction of a random slope will require a linear term for time.

Let the response variable be normally distributed. The prior distribution functions for the parameters are those chosen according to the previous section.

The model can be specified by the formula

$$
\begin{aligned}
\text{change w}_{it} = {} & f_1(t_i) + f_2\big(\text{age}_i\big) + f_3(\text{edss}_i) + f_4(\text{dur}_i) + \alpha_1 * \text{course}_i^{(1)} \\
& + \alpha_2 * \text{course}_i^{(2)} + \alpha_3 * \text{gender}_i + b_i + \varepsilon_{it},
\end{aligned}
\tag{4.1}
$$

where $b_i, i = 1, \ldots, N$, is the random intercept. The functions $f_i, i = 1, \ldots, 4$, denote the P-Splines defined in Section 3.1 with their Bayesian extension from Section 3.2. For the benefit of estimating a smooth function for time, a random slope term has been left out. Thus, possible nonlinear effects of time may be detected. The introduction of a random slope would require a linear term for time. The model could be set in the framework of a random intercept model.

The prior distributions were chosen in the usual way, that is, diffuse priors for the fixed effects α_1, α_2, and α_3 and inverse gamma distribution for the variance component of the random effect and the residual variance with $a = 1$ and $b = 0.005$ each. For the P-Splines, 20 equidistant knots were chosen. The prior for the hyperparameter that regulates the smoothness is also inverse gamma distributed with $\tau^2 \sim \text{IG}(1, 0.005)$. Bayesian P-Splines functions are used to model nonparametrically the impact of four covariates. Fixed effects are acting additively. In our modelling estimates for fixed effects have been calculated by Maximum Likelihood (ML) or Restricted Maximum Likelihoo (REML) functions are used to model nonparametrically methods and then, empirical Bayes methods were used to get estimates for the random effects. In general, classical frequentist approach assumes parameters to be unknown but fixed; whereas in the Bayesian context, all parameters are specified as random variables with a prior distribution. The estimates are performed with the software BayesX (http://www.stat.uni-muenchen.de/~lang/bayesx/bayesx.html). BayesX allows for estimation of regression models such as generalized additive models (GAMs), generalized additive mixed models (GAMMs) within a unifying framework (see
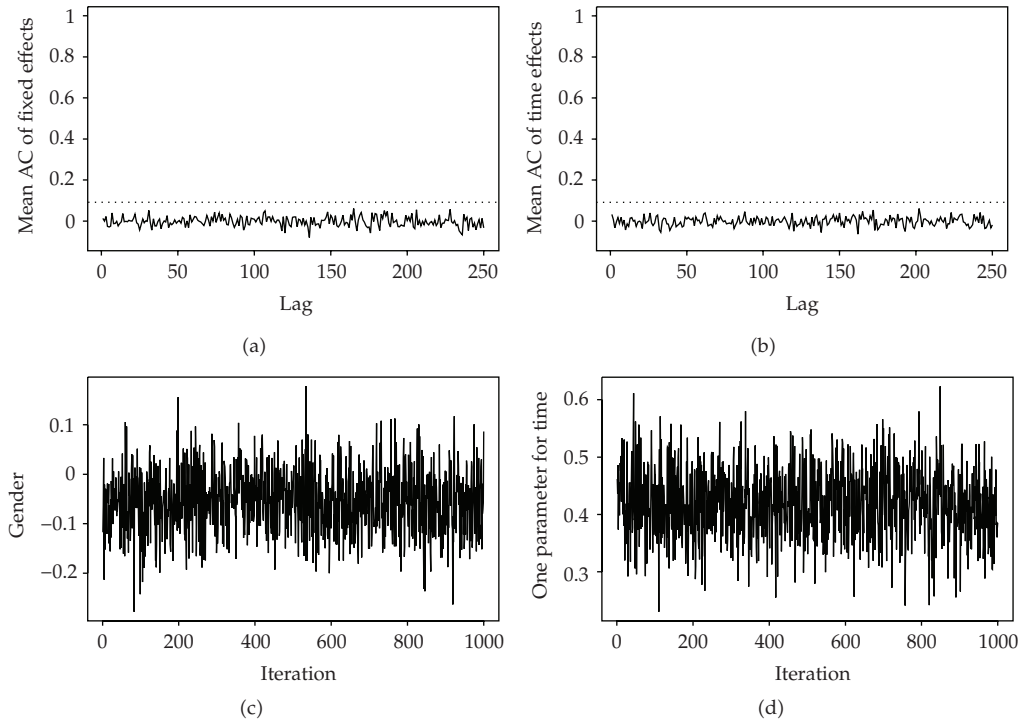
Figure 3: Autocorrelations of fixed effects and parameters for time (a,b) and mixing behavior of the estimate for gender and one time parameter (c,d).

Table 1: Estimates of variance components.

source of variation	Mean	Std. dev.	10% qu.	50% qu.	90% qu.
Within patients	0.593373	0.009933	0.580887	0.593407	0.606655
Between patients	0.536405	0.000929	0.498304	0.535726	0.573733

Brezger et al. [28]). The advantage of using this software is that it supports nonstandard regression situations such as regression for categorical responses. Inferential procedures are based on two different inferential concepts: (i) mixed model methodology corresponding to penalised likelihood or empirical Bayes inference (implemented in remlreg objects); (ii) Markov chain Monte Carlo simulation techniques corresponding to full Bayesian inference (implemented in bayesreg objects). Since the calculation of the posterior distribution obtained by using Bayes' theorem is computationally infeasible in most cases, Markov chain Monte Carlo (MCMC) methods are used for simulation. The posterior mean plots illustrate the impact of the risk factors on the EDSS weighted change. Before looking at the parameter estimates, the convergence and mixing behavior of the MCMC procedure is of interest. Test runs with a small number of iterations suggested to take a burn-in period of 20 000 and step width 500. The number of iterations was therefore set to 520 000, so that 1000 samples were stored. With these parameters, a good behavior of the chain was obtained. Figure 3 shows the sampling and autocorrelation plots of the fixed effects and gender, and of one parameter for the time effect. All other autocorrelation and sampling plots are comparable to the examples showed.

Table 2: Estimates of constant effects.

Variable	Mean	Std. var.	10% qu.	50% qu.	90% qu.
gender	−0.053879	0.065302	−0.140182	−0.054003	0.031402
course$^{(1)}$	0.323480	0.104238	0.185648	0.321651	0.460414
course$^{(2)}$	0.339837	0.098472	0.210783	0.339599	0.469657

We notice in Table 1 that the two variance components have similar magnitude. This suggests that the unobservable heterogeneity between patients explains a portion of total variation similar to that explained by the observable covariates. The prognostic factors included depict an average patient profile which is not representative of the population since it does not capture a big part of its variability.

Table 2 reports results concerning the fixed effects. The gender of the patients (female are reference category) has a negligible impact on the changew$_{\text{EDSS}}$. Actually, an effect of the variable "course" is revealed; patients entering the study in a progressive phase (course$^{(1)}$, course$^{(2)}$) show a higher risk of worsening than those who enter in a relapsing-remitting phase (reference category). Notice that, based on this result, the variable can be interpreted as a short-term predictor only. In fact, the course may change over time. Thus, interactions with time can be investigated.

Posterior means are plotted in Figures 6(a), 6(b), 6(c), and 6(d). These plots show an increasing linear effect of "time" on EDSS change (Figure 6(a)), thus suggesting that a higher worsening is observed in the patients included in longer studies. The variables "age at onset" (Figure 6(b)) and "duration" (Figure 6(d)) have a negligible impact on EDDS change (credible interval of the posterior mean includes zero value) suggesting that they do not affect the trend over time. This indicates that these covariates, which are commonly considered increasing risk factors for MS, affect the initial level of MS severity only whereas no significant impact on the intensity of progression of MS (described by changew$_{\text{EDSS}}$) can be revealed. More informative is the effect of the variable "baseline EDSS." In Figure 6(c) a constant trend is detected between levels 2 and 6 of baseline EDSS. This value is commonly reported for patients who are relatively stable regarding ambulation disability. Actually, a remarkable trend is attributable at the lowest and highest baseline EDSS levels. Patients with baseline EDSS at 0 or 1 increase more in their level of disability than those between 2 and 6 as well as patients entering with EDSS larger than 6.5. The direction of the trend depends on how ambulation and other functional status are weighted in the EDSS computation. Patients with high initial EDSS are likely to have their general functional status deteriorated rapidly (at these levels EDSS is also computed for 0.5 steps amplitude). Credible intervals often happen to widen in correspondence of Splines tails. In these dataset a low number of patients present extreme values for the analysed covariates according to the inclusion criteria.

Overall, it has to be noted that not all included effects influence the response variable significantly. This is also affected by the modelling approches. The plot of the population residuals (Figure 4) shows a skewed distribution with negative outliers. That is, the fixed part of the predictor highly underestimates the observed outcome variable for many patients. The introduction of random effects causes a shrinkage towards zero. In particular in Figure 5 a systematic trend in the residuals is revealed. Negative values of the response are overestimated by the predictor; whereas positive values are underestimated. In general, the fitted values tend to be more conservative and estimates are shifted towards less change in EDSS. This again supports the idea that a high amount of variation is explained by the random effects.
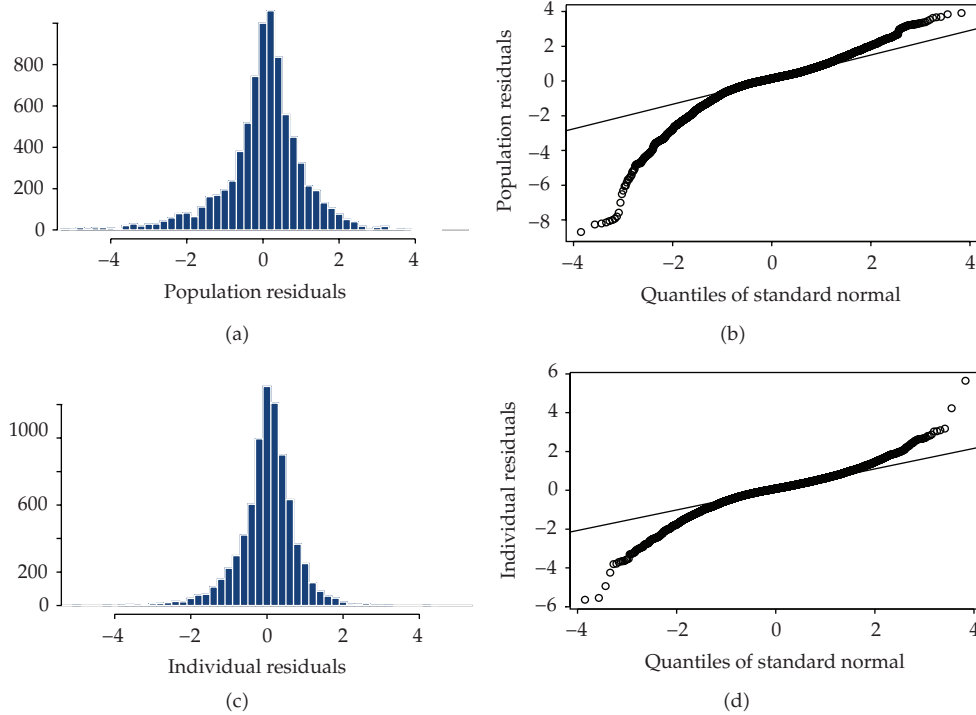
(a)

(b)

(c)

(d)

Figure 4: Histograms and normal-quantile plots for population residuals (a,b) and individual residuals (c,d).
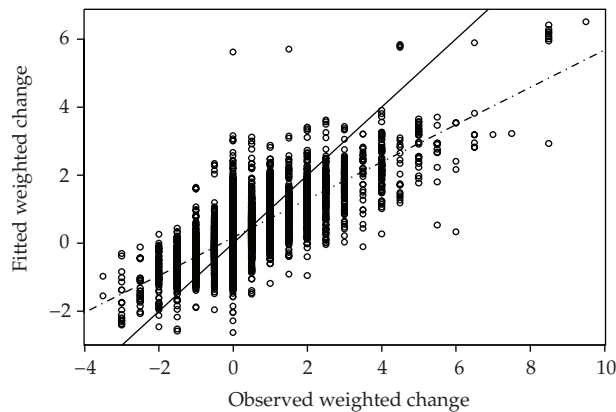

Figure 5: Plot of observed against fitted values (dashed line: linear regression line of the scatter plot; full line: the diagonal).

4.2. P-Splines Random Intercept Model with an Ordinal Response

Let now include in the P-Splines analysis the ordinal nature of the variable changew_{EDSS} which has been discharged in the previous modelling. In fact, a comparison between the two modelling aims at verifying whether investigating the response as a Gaussian rather than an ordinal variable leads to different evaluations of the role of MS prognostic factors.
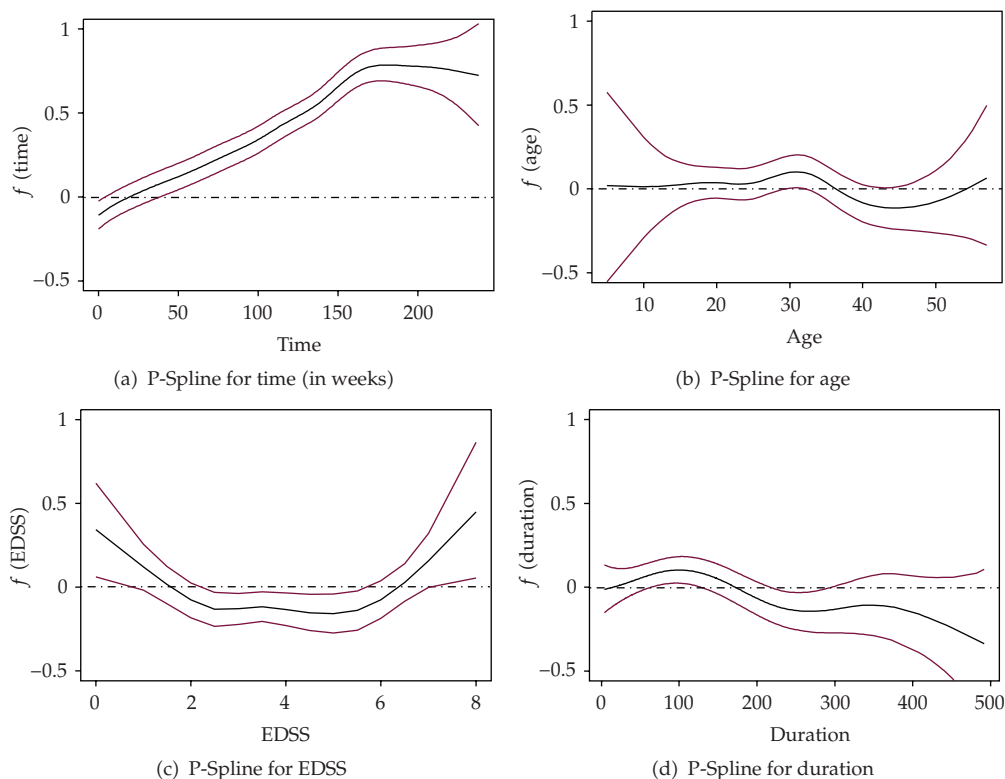
(a) P-Spline for time (in weeks)

(b) P-Spline for age

(c) P-Spline for EDSS

(d) P-Spline for duration

Figure 6: P-Spline posteriors for random intercept models with Gaussian response. Posterior means and confindence interval are plotted.

In Section 3 we have briefly described how in an ordinal thresholds model the posterior mean estimate depends on the thresholds parameter vector $\theta = (\theta_1, \ldots, \theta_{k-1})'$.

As mentioned in the Introduction, to focus on the severity of the disease change values of the variable "changew$_{EDSS}$" were grouped in 5 categories (see Table 11 in appendix). An ordinal threshold model is performed. Each category includes at least a change of 1.0 on the EDSS score to be confident, according with MS literature [1], that a real change in disability occurred. The new response variable "changeord" ranges, with 5 categories, from "big decrease" to "big increase" over a "stable" phase.

In accordance with the Gaussian response P-Splines model in (4.1) a general ordinal threshold model can be now written as

$$
\begin{aligned}
\text{changeord}_{it} = {} & f_1(t_i) + f_2(\text{age}_i) + f_3(\text{edss}_i) + f_4(\text{dur}_i) + \alpha_1 * \text{course}_i^{(1)} \\
& + \alpha_2 * \text{course}_i^{(2)} + \alpha_3 * \text{gender}_i + b_i + \varepsilon_{it}.
\end{aligned}
\tag{4.2}
$$

Thresholds for ordinal response variable are described in Table 3 where the thresholds parameters $\theta's$ have to be estimated from the data and interpreted according to Table 11 in appendix. The prior distributional assumptions are the same described for model (3.6). In addition, a diffuse non-informative prior $p(\theta) \propto \text{const}$ was chosen.

Table 3

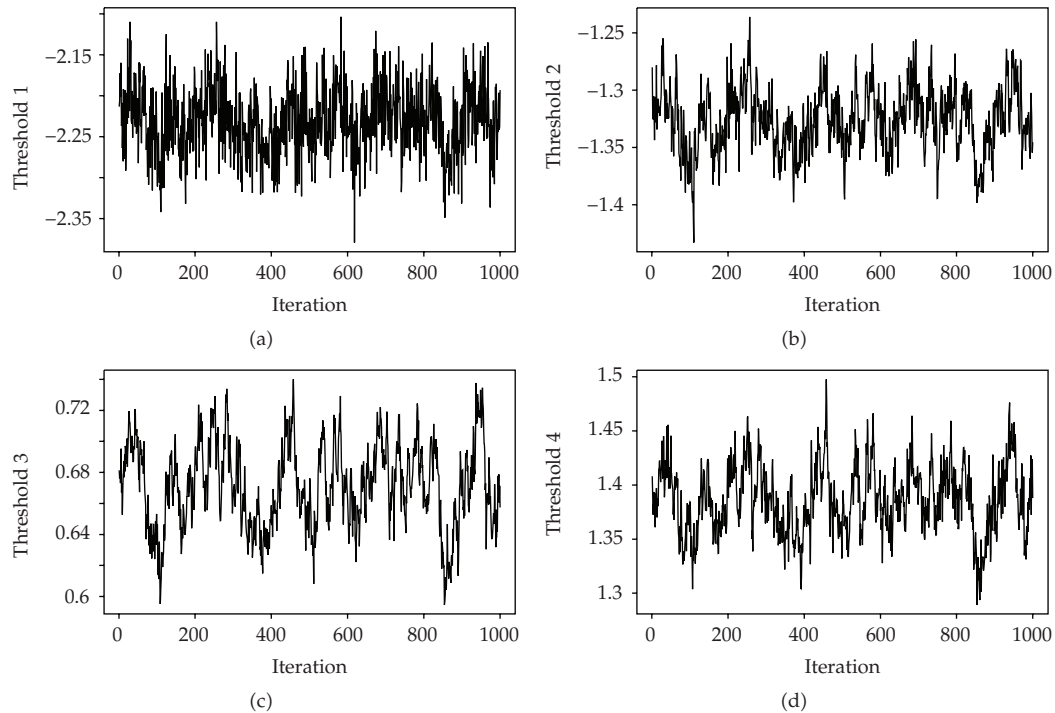| Thresholds	Changeord	
$\leq \theta_1$	Big decrease	$(\ll)$
$(\theta_1; \theta_2)$	Small decrease	$(<)$
$(\theta_2; \theta_3)$	Stable	$(=)$
$(\theta_3; \theta_4)$	Small increase	$(>)$
$\geq \theta_4$	Big increase	$(\gg)$

(a)

(b)

(c)

(d)

Figure 7: Sampling plots of threshold parameters.

The ordinal mixed effect model results are obtained by a combination of Bayesian and classical estimation procedures. First, Bayesian estimates for the fixed effects are derived. These estimates constitute the basis for the marginal likelihood estimation of the random effect, as implemented by the software MIXOR (http://www.uic.edu/~hedeker/mix.html) [29]. In the first Bayesian step, as in the Gaussian response model, P-Splines of degree 3 with second-order random walk penalty were considered. Convergence and mixing behavior of the MCMC procedure [26] show a much larger number of iterations needed. Actually, to guarantee an almost ideal behavior for the samples and autocorrelation plots of all P-Splines parameters, a burn-in period of 500 000 and a step width of 1000 were chosen. The diagnosis plots of the fixed effects showed a satisfying mixing behavior and a negligible autocorrelation. However, the trace plots of the threshold parameter samples (Figure 7) illustrate a bad mixing behavior. Positive and negative correlations seem to alternate. Hence, the estimation of threshold parameters is not stable.

Table 4: Estimates of threshold parameters.

Threshold	Estimator	Std. error	z-value	P-value
θ_1	0	—	—	—
θ_2	1.46894	5.66944	44.35629	<.0001
θ_3	4.59204	0.03922	117.09799	<.0001
θ_4	5.66944	0.04102	138.22608	<.0001

Table 5: Estimates of constant effects.

	Estimate	Std. error	z-value	P-value
gender	−0.08919	0.08839	−1.00902	.31297
course$^{(1)}$	0.54872	0.17777	3.08677	.00202
course$^{(2)}$	0.62386	0.12768	4.88632	.00000

Table 6: Crosstab of observed and fitted response.

		Fitted category					
		≪	<	=	>	≫	Total
	≪	**10**	61	47	0	0	118
	<	16	**120**	556	4	0	696
Observed category	=	85	128	**4684**	278	6	5181
	>	2	15	704	**661**	17	1399
	≫	0	6	154	453	**261**	874
	Total	113	330	6145	1396	284	8268

Bayesian estimates of the fixed effects provide information to reduce the number of parameters and to construct an appropriate ordinal regression model with smooth functions chosen as polynomial. Let now present the final estimation results obtained by this mixing two-step procedure.

By this mixing two-step procedure, the estimated threshold parameters are given in Table 4.

The fixed effect estimates are reported in Table 5. The estimates of both courses are significant and can be so interpreted. The estimated effect of "course$^{(2)}$" is about 0.62, thus it lies between θ_3 and θ_4 which corresponds to a "small increase" in the EDSS change. This is resonable since the secondary progressive course is more aggressive.

In Figure 8 the Splines for each variable in ordinal modelling with and without random effects are compared.

Preserving the ordinal nature of the EDSS weighted change did not provide evidence of a change in the interpretation of the estimated parameters when compared to the Gaussian model. Results (Figure 8), that appear different in the first sight, like the regression Splines for "duration" and "age at onset," do not show such a discrepancy, when looked at closer. Due to different outcome variables, the estimates cannot be compared directly. Furthermore, the variance of the estimations has to be taken into account. The P-Splines credible intervals in step I of the estimation process as well as the rough plots in step II can serve as indicators for the accuracy of the estimations.

As in the Gaussian model, the model fit is analyzed by comparing the fitted values against the observed values. Table 6 also indicates a systematic error. Only 69.4% of all observations are classified correctly. A good fit is only achieved in the category that defines
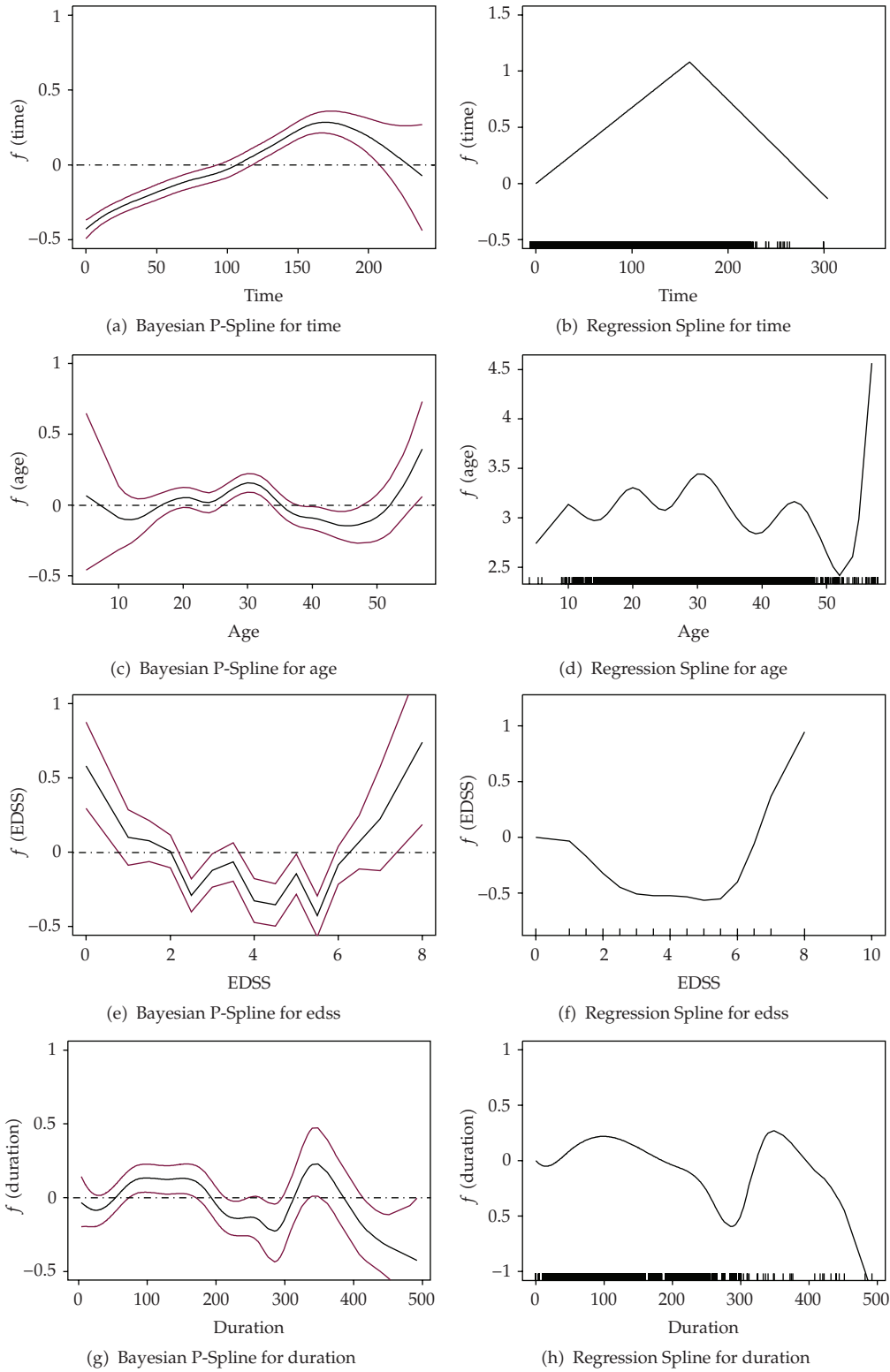
(a) Bayesian P-Spline for time

(b) Regression Spline for time

(c) Bayesian P-Spline for age

(d) Regression Spline for age

(e) Bayesian P-Spline for edss

(f) Regression Spline for edss

(g) Bayesian P-Spline for duration

(h) Regression Spline for duration

Figure 8: Estimation results for Gaussian model with no random effect (left) and with random effect (right).

a "stable" disease progression. All other fitted values are shifted towards this same category. That is, observations on both extreme ends of disease progression cannot be explained well by the ordinal model as well as by the Gaussian model. Moreover, it has to be noted that many computational problems occurred during the estimation of the ordinal model. The autocorrelation and trace plots of the threshold samples (Figure 7) showed a bad mixing and convergence behavior, although random effects have not been included in this stage of modelling. Analyzing the mixed-effects ordinal regression model in MIXOR also led to computational difficulties. Adjustments were needed to improve the chances of convergence. Thus, a Gaussian model should be preferred. Using the Gaussian model also seems to be justified, as the results of both approaches do not differ substantially.

The fitted values plotted against observed values revealed a systematic bias in both random intercept models. The analysis of residuals also suggested that additional random components should be included in the analysis. We next investigate a random slopes model as last step of our modelling.

4.3. P-Splines Random Slopes Model

Heterogeneity in individual MS progression is observed as regarding both the magnitude and the speed. The disability may rise fast in some patients in the beginning and then stabilize; whereas for other patients it rises steadily but slows thereafter. The random intercept models proposed before may be debatable for fitting repeated measures of weighted change in EDSS, since they underestimate the change for patients, whose disability greatly decreased or increased within the time frame of a clinical study. This could cause the bias in the fitted values seen previously within the random intercept models. Of course, the introduction of a random slope alone is not sufficient, as it assumes a constant slope over the whole range of the time frame. By adding a quadratic random effect, the curvature in the progression of disability can be reflected. The splines for the time effect in the previous model (Figures 5(a) and 5(b)) also showed a quadratic trend. Hence, a quadratic random slopes model seems to be appropriate to account for the effect of time detected before as well as the increased heterogeneity in disability. To ensure interpretability, fixed effects for the intercept, linear slope and quadratic slope were also included. Random effects can then be interpreted as deviations from the population mean as above discussed.

In the random slopes model the MS patients are considered to differ from the average trend of the populations as regarding both the initial disability level (random intercept) and the intensity of the MS clinical progression (random slopes). The proposed model is therefore given by

$$
\begin{aligned}
\text{change } w_{it} = {} & f_1(\text{age}_i) + f_2(\text{edss}_i) + f_3(\text{dur}_i) \\
& + \alpha_1 * \text{course}_i^{(1)} + \alpha_2 * \text{course}_i^{(2)} \\
& + \alpha_3 * \text{gender}_i + (b_{i0} + \alpha_0) \\
& + (b_{i1} + \alpha_4) * t_i + (b_{i2} + \alpha_5) * t_i^2 + \varepsilon_{it},
\end{aligned}
\tag{4.3}
$$

where b_{i0} is the random intercept, b_{i1} the random slope, and b_{i2} the quadratic random slope parameter. The fixed intercept is denoted by α_0, the fixed time by α_4, and the fixed quadratic time effect by α_5. As in the Gaussian model, since no previous information on
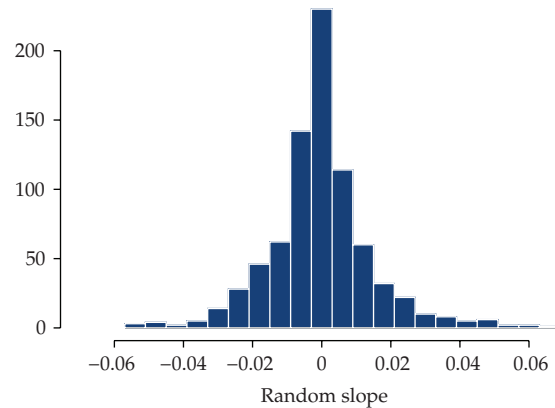
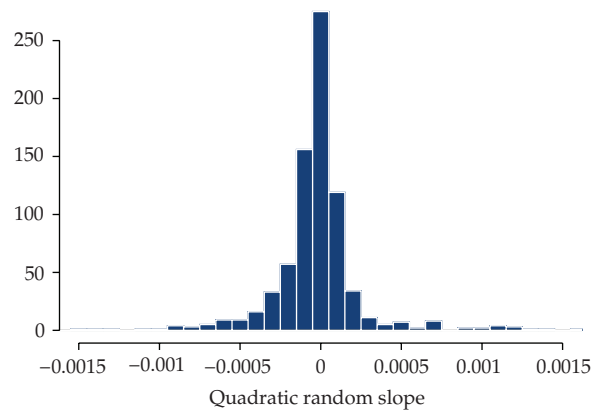**Figure 9:** Histogram of random slope estimates.

Figure 10: Histogram of quadratic random slope estimates.

the parameters is available, a Bayesian approach is justifiable, when using a diffuse prior distribution assumption for all fixed effects, whereas all random effects are assumed to be normally distributed.

With the introduction of a random slope component we notice that the variance components explained by the random intercept and slope shown in Table 7 are reduced by a significant amount in comparison to the Table 1, where observable and unobervable source of variation weighted similarly. Indeed the variance components attributable to the random effects are much smaller than the scale parameter. This is due to the higher number of parameters in the model, especially the random slope and quadratic random slope effects. Histograms of the random effects (Figures 9 and 10) illustrate the corresponding posterior distributions. The kurtosis is higher than a normal distribution and a lot of outliers can be detected. Although the distributions are almost symmetric and bellshaped, a deviation from a normal distribution is likely. The histogram in Figure 10 suggests presence of negative outliers. The mean and median are also slightly negative. Furthermore, patients, that are observed over a longer time generally deviate from the population effect in the negative direction. Thus, the flat and almost negative trend at the upper tail of the time distribution, that was detected before, is reflected in the random slope estimates.

Table 7: Estimates of variance components.

Source of variation	Mean	Std. dev.	10% qu.	50% qu.	90% qu.
Scale	0.360485	0.006067	0.352562	0.360414	0.368427
Intercept	0.024117	0.009096	0.012335	0.024135	0.036359
Linear slope	0.000449	$3.66{*}10^{-5}$	0.000404	0.000448	0.000495
Quadr. slope	$1.77{*}10^{-5}$	$9.30{*}10^{-7}$	$1.65{*}10^{-5}$	$1.76{*}10^{-5}$	$1.89{*}10^{-5}$

Table 8: Estimates of constant effects.

Variable	Mean	Std. var.	10% qu.	50% qu.	90% qu.
Intercept	0.03068427	0.017732	0.006067	0.0352562	0.0360414
Time	0.004538	0.001049	0.003242	0.004499	0.005897
Time^2	$1.192{*}10^{-4}$	0.000156	0.000191	0.000209	0.000218
Course$^{(1)}$	0.129323	0.049309	0.070234	0.128142	0.192123
Course$^{(2)}$	0.089652	0.042889	0.032993	0.089353	0.144362
Gender	-0.070236	0.030946	-0.11044	-0.06945	-0.110436

To ensure comparability of the random intercept and random slopes model, calculation in BayesX was performed with the same number of iterations, that is, a burn-in of 20 000 and a step width of 500. The convergence and mixing behavior were comparable to the ones obtained in the random intercepts model (Figure 3). Hence the models are also computationally equivalent.

Let next present the results of this modelling. In Table 7 we notice how the variance components of the random effects are significantly lower than the residual variance. This suggests a significant improvement in the modelling when compared to the random intercept models.

The estimation of the fixed effect is provided in Table 8. We notice that the fixed effects for time and quadratic time are positive, thus indicating that the disability averaged over the population is increasing over time. Furthermore the positive estimates of the progressive course are smaller than before. It seems that some amount of information, that was given by the course of disease in the previous model, is captured by the individual linear or quadratic slopes. To confirm this assumption, a closer look has been taken into the distribution of the random slope estimates within each group. Table 9 shows the mean values for the linear and quadratic random slope parameters. Slopes for progressive patients (sp, pp, or pr) are generally higher than for relapsing-remitting patients (rr). This suggests that the categorization into disease courses reflects the kind of progression over time, so that a time-dependent effect rather than a constant effect per group could be an alternative.

P-Splines curves plotted for the metric variables "age at onset," "baseline EDSS," and "duration" are here omitted since it did not change substantially the information obtained in the previous analyses, except for the credible intervals that resulted noticeably narrower than before.

Finally, by looking at the plot of fitted against observed values (Figure 11) we can still observe a systematic trend. But as the dashed line, indicating a linear regression fit, lies closer to the diagonal, model fit is improved significantly. However, outliers can still be crucially detected on both extreme ends of the weighted EDSS change.

Another question was whether it is justified to use the change in EDSS as a metric outcome variable. However it is still debatable whether a 1 point change, although weighted,

Table 9: Mean of random slope estimates, stratisfied for courses.

Course	Mean of random linear slope	Mean of random quadratic slope
pp or pr (course[(1)])	-0.000024	0.013340
sp (course[(2)])	0.000927	0.000060
rr (reference category)	-0.000561	-0.001539
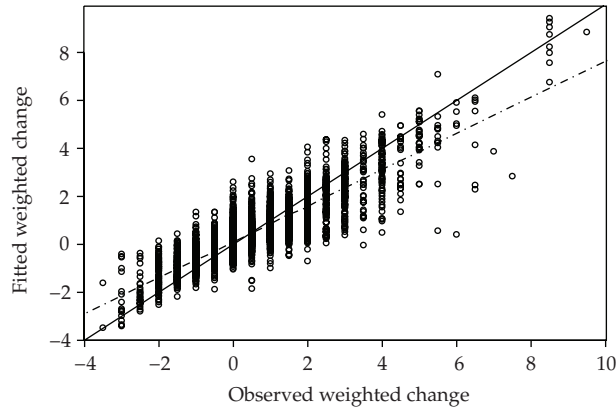
Figure 11: Plot of observed against fitted residuals in a random slope model.

does really have the same meaning over the whole EDSS range. Results on ordinal modelling are here omitted. Indeed, combining levels of the outcome variable to 5 ordered categories not only accounts for the ordinal structure in the response, but also ensures comparability of the responses. However, P-spline results are very similar and do not justify the use of such computationally demanding and time consuming procedure. Reducing the number of ordered categories to three could be an alternative. In this case, the levels of the response are reparametrized to obtain stable estimates and then, analysis can be carried out in BayesX.

5. Conclusions

In the biostatistic literature a few attempts of statistical modelling for investigating MS course are provided Heijtan (1991). In this paper the focus of the interest lies on modelling the unobserved heterogeneity in MS longitudinal data for the better understanding of the impact of prognostic factors on MS severity. A nonparametric approach is suggested to avoid restrictive assumptions about the analytical form of the relation between prognostic factors and outcome of interest. Furthermore, the introduction of random as well as fixed effects of the covariates addressed the issue of including both observed and unobserved heterogeneity. Hence, generalized additive mixed models (GAMMs) have been presented as a natural statistical tool to investigate nonparametrically, by means of Splines, the role of MS prognostic factors.

We have been mainly addressing two fundamental features.

(1) Most of the statistical modelling in MS consider EDSS as a metric variable, regardless of the ordinal nature of this measure. Does this assumption affect the estimation of the effect of the prognostic factors?

Table 10: Description of the covariates included in the analysis of SLCMSR data set (*pp: primary progressive; pr: progressive relapsing; sp: secondary progressive).

Changew	Weigthed change in EDSS from first observation
t	Time in weeks from first observation
Edss	EDSS at first observation
Age	Age at disease onset
Dur	Duration in months from onset to first observation
Gender	$= \begin{cases} 0 & \text{for female} \\ 1 & \text{for male} \end{cases}$
Course	$\text{Course}^{(1)} = \begin{cases} 1 & \text{if course = pp}^* \text{ or pr}^* \\ 0 & \text{otherwise} \end{cases}$ $\text{Course}^{(2)} = \begin{cases} 1 & \text{if course = sp}^* \\ 0 & \text{otherwise} \end{cases}$ Reference category is relapsing remitting

Table 11: Categorization of EDSS change.

weighted change	ordinal change	label	number of observations
≤ -2	big decrease	$\ll$	126
$-1.5; -1$	small decrease	$<$	721
$-0.5; 0; 0.5$	stable	$=$	5438
$1; 1.5$	small increase	$>$	1440
≥ 2	big increase	$\gg$	893

 (2) Unobserved sources of heterogeneity affect individual MS development. Does this source of heterogeneity create difference among patients as regarding how they enter the study or also how large and fast the progression is?

 An answer to the first question is provided by comparing results from Bayesian P-Splines models performed with a Gaussian as well as an ordinal response. A first conclusion is that the numerically demanding and time consuming estimation of an ordinal mixed effect model is not justified by a real gain in the results. Actually, the interpretation of the role of prognostic factors did not change dramatically. Thus, a Gaussian model is suggested. Overall, the "baseline EDSS" appears to have a strong influence on the weighted change in EDSS. Patients enrolled in a study with an EDSS lower than 2 or bigger than 5.5 are expected to experience a higher increase in disability than in patients with EDSS between 2 and 5.5. The effects of "duration," "age at onset," and "gender" are marginal and even negligible. Thus duration can be considered an offset variable rather than an explicative one for the change in EDSS over time. However, there is a general positive time trend during a clinical study. That is, the more time elapses are, the higher the change in EDSS is. This is in favor of the use of natural history data for addressing the issue of prediction in MS course. Again, the "course" of the disease emerged as a short-term predictive factor. This variable can be seen as a summary of the past disease progression. As soon as there are more variables available, that hopefully explain the previous disease course of a patient, a shrinkage of this effect can be expected. In general, the disability of patients, that are categorized in one of the progressive courses, increases more.

To address the second question the variance components in the random slope model and in the random intercept model are compared. The introduction of a random slope leads to a better estimate of the "within-patients" variability which is much higher than the "between-patients" variability; whereas this was not the case in the random intercept model. This implies that accounting for the variability in the progression of the individual disability allows for a much better classification of the patients. Furthermore, a comparison between the estimates of the fixed effect of time in random slopes model (Table 8) and the posterior mean estimate of time in random intercept models can be made. The stabilization of the effect of "time" after an increasing phase suggested by the random intercept model might be attributable to the random quadratic time effect. This is consistent with the hypothesis that unobserved heterogeneity plays a crucial role in evaluating the individual intensity of progression. Finally, the influence of the disease "course" is much smaller in the random slopes model. It has been shown that this is due to deviating random effects between the 3 groups of patients. Thus, it is advisable to think about time-dependent effects, that is, to estimate one slope for each group of patients.

In conclusion Bayesian approach to random slope models, based on MCMC algorithms, emerged as extremely flexible in the context of MS data as presented here. In particular BayesX is implemented to allow for smooth P-Splines for metric covariates. This is a noteworthy advantage with respect to common techniques based on simple linear models or on other strict assumptions on the functional form since nonlinear effects like the influence of the EDSS at first observation could not have been detected.

A fully Bayesian method for P-splines has been used according to Brezger (2000). This emerges as extremely flexible in situations where large data sets are used and a moderate number of smoothing parameters are to be estimated. Thus, Bayesian P-splines are recommended as a suitable and flexible tool in addressing complex data like MS frameworks, where the form of the smoothing function f is expected to be highly oscillating and rapidly changing.

Appendix

See Tables 10 and 11.

Acknowledgments

References

[1] J. H. Noseworthy, M. K. Vandervoort, C. J. Wong, and G. C. Ebers, "Interrater variability with the expanded disability status scale (EDSS) and functional systems (FS) in a multiple sclerosis clinical trial," *Neurology*, vol. 40, no. 6, pp. 971–975, 1990.

[2] M. Soilu-Hänninen, J. O. Koskinen, M. Laaksonen, A. Hänninen, E.-M. Lilius, and M. Waris, "High sensitivity measurement of CRP and disease progression in multiple sclerosis," *Neurology*, vol. 65, no. 1, pp. 153–155, 2005.

[3] S. Esbjerg, N. Keiding, and N. Koch-Henriksen, "Reporting delay and corrected incidence of multiple sclerosis," *Statistics in Medicine*, vol. 18, no. 13, pp. 1691–1706, 1999.

[4] P. S. Albert, H. F. McFarland, M. E. Smith, and J. A. Frank, "Repeated measures time series for modelling counts from a relapsing-remitting disease: application to modelling disease activity in multiple sclerosis," *Statistics in Medicine*, vol. 13, pp. 453–466, 1994.

[5] W. I. McDonald, A. Compston, G. Edan, et al., "Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis," *Annals of Neurology*, vol. 50, no. 1, pp. 121–127, 2001.

[6] T. Hastie and R. Tibshirani, "Bayesian backfitting," *Statistical Science*, vol. 15, no. 3, pp. 196–223, 2000.

[7] L. Fahrmeir and S. Lang, "Bayesian inference for generalized additive mixed models based on Markov random field priors," *Journal of the Royal Statistical Society C*, vol. 50, no. 2, pp. 201–220, 2001.

[8] L. Fahrmeir and S. Lang, "Bayesian semiparametric regression analysis of multicategorical time-space data," *Annals of the Institute of Statistical Mathematics*, vol. 53, no. 1, pp. 11–30, 2001.

[9] T. Kneib and L. Fahrmeir, "Structured additive regression for categorical space-time data: a mixed model approach," *Biometrics*, vol. 62, no. 1, pp. 109–118, 2006.

[10] M. J. Lindstrom and D. M. Bates, "Nonlinear mixed effects models for repeated measures data," *Biometrics*, vol. 46, no. 3, pp. 673–687, 1990.

[11] E. F. Vonesh and V. M. Chinchilli, *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, vol. 154 of *Statistics: Textbooks and Monographs*, Marcel Dekker, New York, NY, USA, 1997.

[12] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data*, Springer Series in Statistics, Springer, New York, NY, USA, 2000.

[13] P. J. Diggle, P. J. Heagerty, K.-Y. Liang, and S. L. Zeger, *Analysis of Longitudinal Data*, vol. 25 of *Oxford Statistical Science Series*, Oxford University Press, Oxford, UK, 2nd edition, 2002.

[14] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*, vol. 43 of *Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC, London, UK, 1990.

[15] S. Lang and A. Brezger, "Bayesian P-splines," *Journal of Computational and Graphical Statistics*, vol. 13, no. 1, pp. 183–212, 2004.

[16] J. F. Kurtzke, "Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS)," *Neurology*, vol. 33, pp. 1444–1452, 1983.

[17] R. Bergamaschi, A. Romani, S. Tonietti, A. Citterio, C. Berzuini, and V. Cosi, "Usefulness of Bayesian graphical models for early prediction of disease progression in multiple sclerosis," *Neurological Sciences*, vol. 21, no. 8, pp. S819–S823, 2000.

[18] L. Fahrmeir and G. Tutz, *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 2001.

[19] C. Ke and Y. Wang, "Semiparametric nonlinear mixed-effects models and their applications," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1272–1298, 2001.

[20] J. J. Faraway, *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2006.

[21] S. R. Searle, G. Casella, and C. E. McCulloch, *Variance Components*, John Wiley & Sons, New York, NY, USA, 1992.

[22] B. D. Marx and P. H. C. Eilers, "Direct generalized additive modeling with penalized likelihood," *Computational Statistics and Data Analysis*, vol. 28, no. 2, pp. 193–209, 1998.

[23] P. H. C. Eilers and B. D. Marx, "Flexible smoothing with *B*-splines and penalties," *Statistical Science*, vol. 11, no. 2, pp. 89–121, 1996.

[24] A. Brezger and S. Lang, "Generalized structured additive regression based on Bayesian P-splines," *Computational Statistics & Data Analysis*, vol. 50, no. 4, pp. 967–991, 2006.

[25] G. Kauermann, T. Krivobokova, and F. Ludwig, "Some asymptotic results on generalized penalized spline smoothing," Tech. Rep., Department of Decision Sciences and Information Management (KBI), Katholieke Universiteit, Leuven, Belgium, 2007.

[26] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Interdisciplinary Statistics, Chapman & Hall/CRC, Boca Raton, Fla, USA, 1996.

[27] G. Tutz, *Die Analyse Kategorialer Daten*, Oldenbourg Wissenschaftesverlag, Munich, Germany, 2000.

[28] A. Brezger, T. Kneib, and S. Lang, "BayesX: analysing Bayesian structured additive regression models," *Journal of Statistical Software*, vol. 14, p. 11, 2005.

[29] D. Hedeker and R. D. Gibbons, "A random-effects ordinal regression model for multilevel analysis," *Biometrics*, vol. 50, no. 4, pp. 933–944, 1994.