Di Serio, Lamina:

Bayesian P-Splines to investigate the impact of covariates on Multiple Sclerosis clinical course

Sonderforschungsbereich 386, Paper 353 (2003)

Online unter: http://epub.ub.uni-muenchen.de/

BAYESIAN P-SPLINES TO INVESTIGATE THE IMPACT OF COVARIATES ON MULTIPLE SCLEROSIS CLINICAL COURSE

CLELIA DI SERIO AND CLAUDIA LAMINA

Abstract. This paper aims at proposing suitable statistical tools to address heterogeneity in repeated measures, within a Multiple Sclerosis (MS) longitudinal study. Indeed, due to unobservable sources of heterogeneity, modelling the effect of covariates on MS severity evolves as a very difficult feature.

Bayesian P-Splines are suggested for modelling non linear smooth effects of covariates within generalized additive models. Thus, based on a pooled MS data set, we show how extending bayesian P-splines (Lang and Brezger, 2001) to mixed effects models, represents an attractive statistical approach to investigate the role of prognostic factors in affecting individual change in disability.

1. Introduction.

Many clinical studies collect repeated measures data which allow for assessing the disease process over time when analysed longitudinally. Multiple Sclerosis (MS) is an example of a multifactorial genetic neurological disease where the analysis is commonly focused on the change of the disease status over time. Modelling this kind of chronic diseases is a difficult task 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. In fact, these data are heterogeneus in their structure as regarding different levels of heterogeneity: the time intervals between measurements can vary among individuals and not even the number of observations has to be the same due to skipped examinations or withdrawals from the study. Furthermore, the correlation between repeated measurements within subjects has to be included. These complications are difficult to account for in a conventional analysis focused on investigating and interpreting the role of covariates on the disease course.

Mixed effect models for repeated measures data have become popular in these frameworks because of their flexible covariance structure which allows for non-constant correlation among the observations. More, they have also a natural interpretation. Random effects are introduced to include an unobservable heterogeneity among individuals which represents another source of variation in addition to the residual variance. In practice, you not only want to incorporate subject-specific effects by random effects, but also population-specific effects that are constant among all individuals, the so called fixed effects.

.

In this paper the main interest lies on combining the mixed effect models theory with a non-parametric extension of generalized linear models: the generalized additive models (GAM). We aim at showing how this modelling represents a suitable tool to describe and to handle heterogeneity within a MS setting.

Let us give some notation about general linear models (GLM).Generalized linear models (McCullagh and Nelder, 1980) extend classical linear regression models to allow also for non-normally distributed response variables. As in linear regression, the effect of the covariates $X_1, ..., X_p$ on the response variable Y on covariates $X_1, ..., X_p$is explained by the linear predictor

$$\eta = \alpha + X_1\beta_1 + ... + X_p\beta_p \tag{1.1}$$

where $\beta = (\beta_1, ..., \beta_p)'$ is the unknown parameter vector and the distribution of the response belongs to an exponential family. By taking the normal distribution, the familiar linear regression is returned as a special case.

The second step of generalization is done by the so called link-function: The linear predictor η is related to the mean $\mu = E(Y)$ of the response variable Y of interest by a monotone one-to-one link-function g and its unique inverse function h by

$$g(\mu) = \eta \Longleftrightarrow \mu = h(\eta). \tag{1.2}$$

Thus, not the mean itself is modelled, but rather a transformation of the mean, introduced by the link-function and its inverse function.

Starting from the simplest type of GLM, the linear regression models, mixed effects are introduced in Section 3. First, two types of mixed effect models are illustrated within a simple linear regression framework: a random intercept and a random slope model. This type of modelling assumes a Gaussian distribution for the response variable. Given that the disability scale of interest has an ordinal nature an alternative ordinal threshold model is then presented. Section 4 extends these concepts to a nonparametric framework, thus introducing the generalized additive models (Hastie and Tibshiriani, 1990). Within this class of models, Bayesian P-Splines with mixed effect (Lang and Brezger, 2001; Fahrmeir and Lang, 2001) are presented in Section 5. Section 6 illustrates how these statistical tools apply to the MS data set, comparing results from three different mixed effect models. Discussion of the results is provided in Section 7.

In the next section we introduce some basics about Multiple Sclerosis and the related topics.

2. EDSS 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. The disease markers that are 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*). The EDSS measures MS-related impairments of functional systems (FS). The values of these systems, ranging between 0 - normal function - and 5 -severe disability (sometimes 0 to 6)- are used to define the categorical EDSS score. EDSS is therefore an ordinal discrete variable ranging from 0 (no MS symptoms) to 10 (death due to MS) in half point increments.

This paper is motivated by a worldwide research program that has been established at the Sylvia Lawry Centre for Multiple Sclerosis Research[1] (SLCMSR). The general aim of the research is the better understanding of the determinants of MS course in order to improve the efficiency of therapies for MS patients. Within this program a database has been created, that contains data on untreated patients from natural history studies and on placebo patients from major therapeutic studies conducted worldwide by academic research groups and pharmaceutical industry. Specifically, this paper is based on 897 patients randomly sampled from 17 placebo controlled clinical trials with 8716 repeated measures. In fact, data from clinical trials allow for a good monitoring and comparable time spans (1 to 4 months) between to subsequent observations) whereas natural history data do not.

The EDSS has many shortcomings, one of them being nonlinear and discontinuous. Thus, common alternative outcome measures in clinical trials have been time to reach a certain level (4.0 or 6.0) in EDSS, or time to worsening, defined by an increase of 1 point in EDSS. Actually these outcomes cannot be used whenever the interest of the analysis lies on the whole EDSS course during the clinical trial. Therefore we focused the analysis on the "EDSS change" over time.

The decision, to take the change in EDSS at all, may be criticized, but measures have been taken, to ensure that the variable "**change**" means the same thing for each patient. There is a broad consensus within experts, that, when higher values of EDSS are already reached, changes have higher clinical significance. Since higher EDSS-values are dominated by ambulation, a change of 1 from e.g. EDSS 7 to 8 is more severe than from 1 to 2, where only a slight increase in one of the functional scores is needed. The European Agency for the Evaluation of Medicinal Products stated that "Based on EDSS values, treatment failure or progression should be predefined e.g. as the achievement of a specified degree of disability or of a sustained worsening of relevant magnitude (1 point when EDSS scores ≤ 0.5; 0.5 points if baseline score is > 5.5)" (EMEA (2001)). According to this guidelines, changes in EDSS-values higher than 5.5 have been weighted twice as much than changes below this level. This weighted change ("**changew**") is a measure of severeness in changes of disability and cannot be related to the original EDSS-values any more. At a first attempt, this weighted change, although an ordinal variable, has been assumed to be metrical, since there are 25 ordered categories, ranging from -3.5 to 9.5. Furthermore, in a first analysis the variable "changew" has been assumed to be normally distributed. To understand whether it is justified to use the change in EDSS as a metric outcome variable an additional model as been performed by taking "changew" as ordinal variable.

3.

[1]SLCMSR was founded in 2001 at the Technical University of Munich with financial support of the Multiple Sclerosis International Foundation (MSIF).

4. Mixed effect models: a brief overview.

Mixed effect models, as mentioned in the introduction, are an appropriate statistical tool to describe relationships between a response variable and some covariates in data where fixed and random effects describe the population-specific and subject-specific parameters respectively. In the recent statistical literature (Pinheiro and Bates 2000; Fahrmeir and Tutz, 2001) it has been shown how they provide a flexible and powerful tool for the analysis of repeated measures and how they are intuitively appealing in complex settings such as biomedical frameworks. Actually, fixed effects are associated to an average population trend that is constant among all individuals, whereas random effects account for how the individual randomly deviates from the population trend. Thus, these models incorporates both observable and unobservable heterogeneity among patients. A fundamental issue in this framework is how large is the variance component associated to random effects when compared to the residual variance.

For the sake of simplicity let us introduce the basic notation and concepts for mixed effect models within a GLM where the linear predictor coincides with the link function: that is a a linear regression model.

Linear mixed effect (LME) models assume that the normal response y_i for the i-th subject depends linearly on population-specific effects β and subject-specific effects b_i :

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i \qquad \text{for } i = 1, ...N \text{ individuals} \tag{4.1}$$

The response vector $Y_i = (Y_{i1}, ..., Y_{iT_i})'$consists of repeated observations for subject i at time points $t = 1, ...T_i$. The p fixed effects β describe average trends, whereas the q subject-specific parameters b_i describe how the evolution of the i-th subject deviates from the average evolution in the population. X_i and Z_i are ($T_i \times p$)- and ($T_i \times q$) design matrices that contain the known covariates. Finally, the ($T_i \times 1$)-vector ε_i is the usual vector of residuals, mutually independent on $b_i's$, and it is assumed to be independent normally distributed, $\varepsilon_i \backsim N(0, \sigma^2 R_i)$, with R_i being the covariance matrix containing the temporal correlations between $\varepsilon_{it}{}'s$. Usually R_i is chosen to be equal to the identity matrix I_{T_i} of dimension T_i, so that $cov(\varepsilon_i) = \sigma^2 I_{T_i}$. In that case, the T_i responses on individual i are independent, conditional on b_i and β. In a random effects model the unobservable heterogeneity among subjects can be expressed by a random variable with a distribution function. This so called mixing distribution is often assumed to be a (multivariate) normal distribution, so that $b_i \sim N(0, Q)$, with Q being a positive semi-definite ($q \times q$)-matrix and it is usually assumed to be unstructured.

In mixed effect modelling any number of random effects can be specified. Though, identifiability problems and computational complications may arise when introducing too many random components. The type and number of random effects are crucially related to the focus of the analysis in the extent that they are chosen to model the most important sources of unobserved heterogeneity.

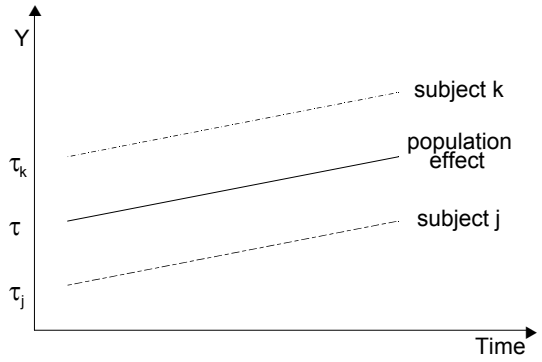


FIGURE 1. The random-intercepts model

4.1. **Random-intercept and random-slopes models.** In the MS 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. The resulting random-intercept and random slope models are next described.

Definition 1. *Random-intercepts model.*

Let a response variable y_i be affected by a cluster- or subject effect with constant slope coefficient γ, then a random-intercept model is described as

$$y_i = \tau_i + \gamma w_i + \varepsilon_i \quad \textit{with} \;\; \varepsilon_i \sim N(0, \sigma^2) \tag{4.2}$$

The intercepts are assumed to be iid $N(\tau, \sigma^2)$ distributed. That means graphically, that the effect for each subject is parallel to the population trend (see Fig. 1). With $\beta' = (\tau, \gamma')$, $X_i' = (1, w_i')$, $Z_i' = 1$ and $b_i = (\tau_i - \tau)$, $b_i \sim N(0, \sigma_\tau^2)$, a linear random effects model of the form (4.1) is obtained.

In practice, a random-intercepts model is achieved by taking the subject identification number as a random variable, as by Fig (1). In a MS setting this support the clinical hypothesis that the initial MS severity affects the MS course with a random impact whereas patients are thought to have an average profile as regarding the MS progression. Thus, random-intercepts models can be restrictive in that they require the slope coefficients to be equal for each subject. A random slope model allows the intensity of evolution to vary among subjects.

Definition 2. *Random-slopes model.*

Let a response variable y_i represent the evolution profiles for each subject that have specific intercepts and and slopes (see Fig. 2); then y_i is described by the following model

$$y_i = \tau_i + \gamma_i w_i + \varepsilon_i \quad \textit{with} \;\; \varepsilon_i \sim N(0, \sigma^2) \tag{4.3}$$

Suppose, that the regression coefficients $\beta_i' = (\tau_i, \gamma_i')$ vary independently across subjects according to a normal density with $\beta_i \sim N(\beta, Q)$,where $E(\beta_i) = \beta$ can

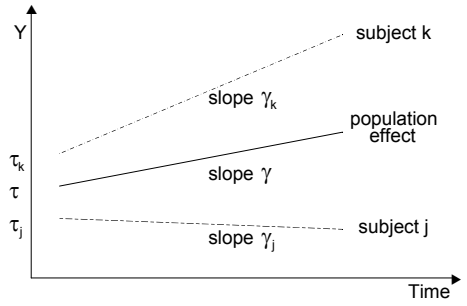


FIGURE 2. The random slopes model

be interpreted as the population effect. That means that the subject-specific effect can be written as $b_i = (\beta_i - \beta)$ *with* $b_i \sim N(0, Q)$*. Rewriting the regressors as* $X_i' = Z_i' = (1, w_i')$ *returns the linear random effects model of the form (4.1).*

In practice, one gets a random-slopes model by taking the subject identification number (id) as well as the interaction of this subject id with time (in a longitudinal setting) as random (Fig.2) . Obviously, choosing the appropriate fixed and random effects in a model is not an easy task.

Several methods have been used to estimate the parameters for the fixed effects. These methods are usually based on a marginal model, that is a model that does not explicitly assume the presence of random effects. Two are the most common estimation procedures: the maximum likelihood (ML) and restricted maximum likelihood (REML) methods. Detailed descriptions and comparisons of the various estimation methods used for LME models can be found, for example, in Searle et al. (1992), Vonesh and Chinchilli (1997) and Verbeke and Molenberghs (2000).

In longitudinal studies where an unmeasured source of heterogeneity is thought to affect the response, the interest often lies on comparisons between the variance components attributable to the random effects and the residual variance of the marginal model. This is a measure of deviation of subject-specific profiles from the population trend, thus allowing also for the detection of outlying subjects or even a group of outlying subjects. In estimating the parameters for random effects Bayesian techniques are commonly provided as efficient tools (Fahrmeir and Tutz (2001). Empirical Bayes estimates, $\widehat{b_i} = E(b_i|Y_i = y_i)$, are based on the posterior distribution of the parameters given the observed data, $f(b_i|y_i)$. These are computed conditionally on all the parameters of the marginal model, that is on the fixed effects estimated with ML and RML-methods. MCMC techniques are then commonly used as approximation procedures to handle the complex calculation of the postestior mean and covariance matrix structure. Detailed description and tools are provided in Gilks *et al.* (1996).

The previous mixed models assume a Gaussian response that is linearly affected by the fixed and random effect. This assumption might be restrictive whenever the response of interest is based on an ordinal variable, such as EDSS. Thus, we next introduce an alternative model that allow the response to have an ordinal structure: the ordinal threshold model with mixed effect.

4.2. **The ordinal threshold model with mixed effect.** The threshold model is based on the idea, that there is a latent non-observable metric variable and that the observed variable merely is a categorized version of this latent variable. The relationship between the latent variable $\widetilde{Y}$ and the observable variable Y can then be described as follows:

$$Y = r \Longleftrightarrow \theta_{r-1} < \widetilde{Y} \leq \theta_r \tag{4.4}$$

with $-\infty = \theta_0 < \theta_1 < ... < \theta_k = \infty$ for $r = 0, ...k$ categories.

That means, if the latent variable lies between the boundaries $\theta_{r-1}$and θ_r, the observable variable takes the value r. $\widetilde{Y}$ is explained by the regressor variables in the linear form

$$\widetilde{Y} = -X'\delta + \varepsilon \tag{4.5}$$

where the nuisance parameter has the distribution function F with $E(\varepsilon) = 0$. Thus, the conditional mean of $\widetilde{Y}$ can be written as $E(\widetilde{Y}|X) = -X'\delta$. These assumptions result in the cumulative model

$$P(Y \leq r|X) = F(\theta_r + X'\delta) \tag{4.6}$$

The cumulativity refers to the following property:

$$P(Y \leq r|X) = P(Y = 1|X) + ... + P(Y = r|X) \tag{4.7}$$

So far, no intercept has been included in the model formulation (4.6). Including an intercept would make identifiability of the thresholds $\theta_0 < ... < \theta_{r-1}$impossible, that is ensured by setting the intercept equal to 0.

The choice of the distribution function F naturally influences the appearance of the model. Common choices are the logistic or the normal distribution. In the following, the normal distribution will be used. A detailed description of the cumulative threshold model and more information on other modelling strategies within the framework of a threshold approach can be found in (Tutz, 2000).

Choosing the normal distribution gives the so called *ordinal probit model*:

$$P(Y \leq r|X) = \phi(\theta_r + X'\delta) \tag{4.8}$$

with

$$\phi(\nu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\nu} e^{-\frac{x^2}{2}} dx. \tag{4.9}$$

The model can have the following intuitive interpretation: By means of the relationship (4.4) between the observable and underlying variable the density is cut into k parts. The areas defined by the density curve and the thresholds θ_r and θ_{r+1}

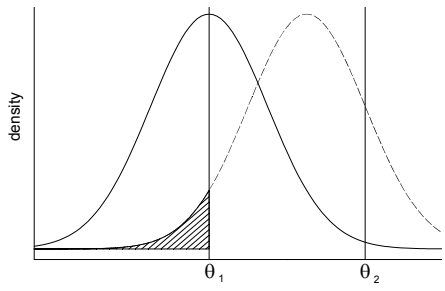


FIGURE 3. The densities of population 1 (dashed line) and population 2 (full line), cut into 3 parts by thresholds θ_1 and θ_2

can then be interpreted as the probability of being in category r. An example is provided for a three categories variable by Fig. (3).

The *ordinal probit model* can be included in the class of generalized linear models (GLM) that allows for non-normally distributed response variables. In this formulation the link function g_r depends on the number of categories and for the distribution in (4.9) is given by

$$g(\mu) = g_r(\pi_1, \pi_2, ..., \pi_k) = \theta_r + X'\delta, \tag{4.10}$$

$$r = 1, ..., k; \quad \pi_r = P(Y = r|X) \tag{4.11}$$

Let us now formulate an *ordinal probit model* where mixed effects are included for repeated observations ($t = 1, ...T_i$) with an ordinal response structure and a random intercept.

Definition 3. *Let x'_{it} be the design vector for the fixed effects, θ_r the rth threshold and $b_i \sim N(0, \sigma^2)$ the subject-specific effects. Then, the linear predictor has the form*

$$\eta_{itr} = \theta_r \ + b_i + x'_{it}\delta. \tag{4.12}$$

This can be interpreted as subject-specific shifting of the thresholds where

$$\theta_{ir} \ = \theta_r \ + b_i. \tag{4.13}$$

Thus, the conditional response probabilities are given by

$$\pi_{it1} = F(\theta_{i1} + x'_{it}\delta), \quad \pi_{itr} = F(\theta_{ir} + x'_{it}\delta) - \pi_{it,r-1}, \quad \textit{for } r = 2, ...k. \tag{4.14}$$

In the previous modelling the interest lies on estimating both the parameters of the ordinal model θ and δ as well as the parameters of the random effects σ. Details about the estimation procedures are provided by Hedeker and Gibbons (1994).

5. Generalized additive models.

To formulate the final models for our analysis we place the previously presented concepts in a non-parametric framework. Actually, this paper is aimed at combining mixed effects with a non-parametric relationship between the response (ordinal or Gaussian) and the predictors. This will be done within a Bayesian approach.

The proposed models are based on the following idea: when the dependence between the response variable Y and the explanatory variables $X_1, X_2, ..., X_p$ cannot be described by a linear predictor η as by (1.1) then it is modelled nonparametrically by a smooth function $f_j(X_j)$, $j = 1, 2, ..p$.

Notice that these models can be used for exploring the data structure and discovering unknown trend in the data rather than for prediction purposes. A big advantage of using smooth functions is that they let the data decide on the functional form producing an estimate of the trend that is less variable than Y itself. The shape of each of the covariate effects is data-driven. The analyst can then use the smooth to suggest parametric form for the term and then apply the appropriate transformation to predict the response.

In this so called generalized additive model (GAM), introduced by Hastie and Tibshirani [1990], the linear predictor is then assumed to be a sum of smooth functions and has the form

$$\eta = \alpha + f_1(X_1) + ... + f_p(X_p). \tag{5.1}$$

There are many different approaches for modelling the functions $f_1, ...f_p$. In principle, any known smoother can be used to estimate the function, such as polynomial smoothing Splines or regression Splines.

Note, that the generalized additive model consists of a sum of such smooth functions. That is, additivity of effects is assumed. This concept retains the interpretability of the familiar linear model and allows, that some predictors can be modelled with smooth functions $f(x)$, and others with constant estimators.

In this paper we deal with a particular class of smooth functions out of the big set of Splines: the P-Splines (Marx and Eilers 1998). These are based on the assumption that the effect f of a covariate x on the response can be approximated by a linear combination of basis functions. Obviously, the choice of appropriate basis functions is crucial to the final regression Splines. Basic Splines Curves (B-Splines) are a popular choice for basis functions due to their numerical stable behavior. They are defined as it follows.

Definition 4. *B-Splines*

Let $\Psi = \{\xi_i\}, i \in Z$ be a set of knots with $\xi_i < \xi_{i+1}, \xi_i \longrightarrow -\infty$ for $i \longrightarrow -\infty$ and $\xi_i \longrightarrow \infty$ for $i \longrightarrow \infty$. The set of all possible Splines of degree k to Ψ is called space of Splines $S_k(\Psi)$.

The B-Splines of degree k to the knot ξ_i of Ψ is defined recursively as:

$$k = 1: \qquad B_i^1(x) = \begin{cases} \frac{x-\xi_i}{\xi_{i+1}-\xi_i} & : \xi_i \leqq x < \xi_{i+1} \\ \frac{\xi_{i+2}-x}{\xi_{i+2}-\xi_{i+1}} & : \xi_{i+1} \leqq x < \xi_{i+2} \\ 0 & otherwise \end{cases}$$

$$k > 1: \qquad B_i^k(x) = \frac{x-\xi_i}{\xi_{i+1}-\xi_i} B_i^{k-1}(x) + \frac{\xi_{i+k+1}-x}{\xi_{i+k+1}-\xi_{i+1}} B_{i+1}^{k-1}(x)$$

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 :

$$s(x) = \sum_{i=-k+1}^{m-1} \beta_i B_i^k(x), x \in [\xi_1, \xi_m] \tag{5.2}$$

B-Splines depend only on the degree k and the values of Ψ. They are non-zero functions in a defined interval and zero outside of this interval. This then makes them numerically superior to other basis functions.

A crucial problem in Splines theory 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 is recommended. Nevertheless, this may lead to overfitting.

To address this problem we start noticing that the coefficients $\beta_i's$ can be considered as a measure of the basis amplitude in the extent that they regulate the roughness of the curves. The higher the difference between adjacent $\beta_i's$ is, the rougher the curve is. On this idea is based the approach of Eilers and Marx (1996). They introduce a penalization term in the maximum likelihood estimation. To combine the opposite requisites of the modelling, that is enough flexibility without a large overfitting, a relatively large number of equally distant knots is suggested. The high variation of the curves is then reduced by penalizing the likelihood with a l difference penalty term on adjacent B-Splines coefficients as it follows:

$$\lambda \sum_{i=l+1}^{r} \left(\Delta^l \beta_i\right)^2 \tag{5.3}$$

The Penalized Likelihood is then given by

$$PL = l(y, \beta) - \lambda \sum_{i=l+1}^{r} \left(\Delta^l \beta_i\right)^2. \tag{5.4}$$

Fisher-Scoring-Algorithm is used to conduct the maximization on the Penalized Likelihood with respect to the unknown regression coefficients.

Other features of Penalized B-Splines, also called P-Splines are:

- P-Splines can fit polynomial data exactly. If the response variable y is a polynomial in x of degree k, then P-Splines of degree k or higher will exactly fit the data for any λ.
- P-Splines conserve the moments of the data
- The limit of a P-Splines fit of degree k or higher with a large smoothing parameter is a polynomial of degree k-1

The smoothness of the function is now regulated by the smoothing parameters $\lambda_j, j = 1, ...p$. The method recommended by Eilers and Marx is to minimizing the Akaike information criterion (AIC). Details about this criterion can be found in Hastie and Tibshirani (1990). The computation of AIC's for many values of λ is very time-consuming and becomes very impracticable in higher dimensions. Furthermore, the function $AIC(\lambda)$ doesn't need to have a global minimum. Actually, it often has several local minima, which makes it difficult to decide on one optimal λ value. It has been shown (Brezger, 2000), that even in cases of a unique minimum, the choice of λ is not optimal, in the extent that it produces a curve, that is too rough. Alternatives to AIC are cross-validation methods (Fahrmeir and Tutz, 2001).

6. Bayesian P-Splines and mixed effects models.

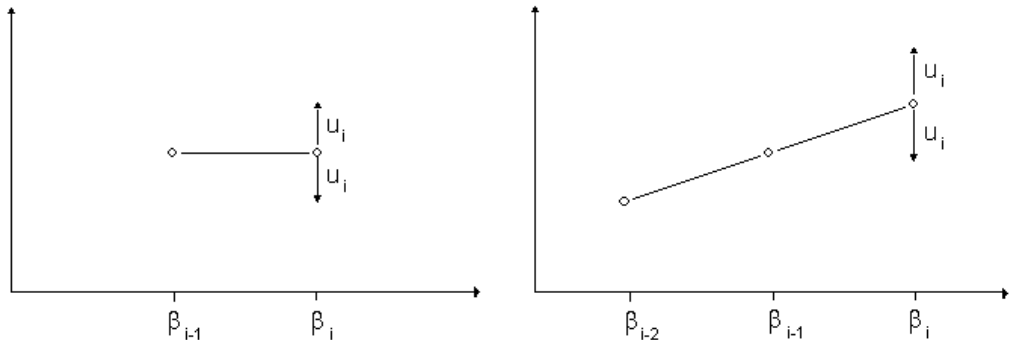


FIGURE 4. Prior distribution for RW(1) (left) and RW(2) (right)

To avoid the problem of the choice of an optimal λ value a Bayesian approach to P-Splines is suggested. Following Fahrmeir and Lang (2001) the Bayesian version of P-Splines allows for considering the smoothing parameters as random and for a simultaneous estimation together with the other model parameters. Starting from the general formulation in (5.2) the difference penalty term on adjacent B-Splines coefficients given by (5.3) is replaced by a Bayesian analogue: a random walk. For instance, first and second differences penalty terms correspond to first and second order random walks, given respectively by:

$$\text{RW(1)}: \quad \beta_i = \beta_{i-1} + u_i, \tag{6.1}$$

with $u_i \sim N(0, \tau^2)$ and $\beta_1 \propto const.$

$$\begin{aligned} \text{RW(2)} \quad &: \quad \beta_i = 2 * \beta_{i-1} + \beta_{j-2} + u_i, \\ u_i \quad &\sim \quad N(0, \tau^2),\ \beta_1 \propto const \text{ and } \beta_2 \propto const. \end{aligned} \tag{6.2}$$

Another possibility are second order random walks (RW(2)). Likewise, RW(2) corresponds to second differences in classical P-Splines. Note, that the priors for the initial values, β_1 for RW(1), and $\beta_1$and β_2 for RW(2) respectively, are diffuse, that is:

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

with $u_i \sim N(0, \tau^2)$, $\beta_1 \propto const$ and $\beta_2 \propto const$.

The illustration of this concept in Fig. (4) shows that u_i, or equivalently the variance parameter $Var(u_i) = \tau^2$, regulates the smoothness of the function. 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. Bayesian P-Splines

It can be shown (Brezger, 2000), that the joint distribution of the prior is given by

$$\beta \propto \exp(-\frac{1}{2\tau^2}\beta' K\beta)$$

with the symmetric penalty matrix K.

In addition to the coefficients, the variance parameter τ_i has to be supplemented with a prior distribution as well. Thus, this parameter is also assumed to be random and is estimated simultaneously with the coefficients 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_i corresponds to the smoothing parameter λ in the classical approach of P-Splines, but it is data-driven and therefore more reliable than λ. The priori for the variance parameter, also called hyperparameter, is often chosen within the family of inverse Gamma distributions, that is:

$$\tau^2 \sim IG(a, b).$$

By setting $a = 1$ and $b = 0.005$ a flat distribution is obtained, which resemble a situation of no prior knowledge on the entire parameter space, that is a Jeffrey's non-informativeness.

We next describe how random effects can be easily included in a Bayesian P-Splines setting, thus leading to a mixed effect model with non-parametric effect f of a covariates x on the response.

Two P-Splines models with mixed effects are proposed to handle gaussian and ordinal responses respectively.

6.1. **Unobserved heterogeneity 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 non parametric effect of the covariates on a Gaussian response. Fixed effect are very likely to be included in the model in addition to the random effect and the P-Splines components.

For the sake of simplicity let now present Bayesian P-Splines within a random intercept model as it follows:

$$y_i = b_i + \sum_{j=1}^{p} f_j(x_{ij}) + \sum_{k=1}^{K} \beta_k w_{ik} + \varepsilon_i \tag{6.3}$$

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

In a Bayesian context, in addition to the above discussed variance component for the random walk regulating the smoothness of the P-Splines, random components are also assigned. In our modelling the following priors are chosen:

1) residual variance component: $\varepsilon_i \sim N(\mu_i, \sigma^2)$ with σ^2being the scale parameter. An inverse Gamma distribution is commonly assigned as

$$\sigma^2 \sim IG(a_\sigma, b_\sigma) \tag{6.4}$$

Again, setting a_σ to 1 and b_σ to 0.005 we obtain an approximately non-informative distribution.

2) variance component for the random effects: $b_i's, i = 1, ..., n$ are generally assumed to be i.i.d. Gaussian, i.e.

$$b_i \sim N(0, \tau_{ra}^2) \tag{6.5}$$

Similarly to the hyperparameter in the random walk approach, the variance parameter $Var(b_i) = \tau_{ra}^2$ is assumed to be random. In this case, the Inverse Gamma distribution is taken as well, so that

$$\tau_{ra}^2 \sim IG(a_{ra}, b_{ra}) \text{ with } a_{ra} = 1 \text{ and } b_{ra} = 0.005. \tag{6.6}$$

3) fixed effects: in a Bayesian framework diffuse priors are chosen to express no prior knowledge about the parameters of the fixed effects; this means:

$$\beta_i \propto const$$

Notice that in these 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 in the models described in this section are based on Bayesian techniques to estimate the posterior distribution functions. It can be shown, that the full conditional distributions of β,and 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. It is reasonable to update the parameters β of function evaluations and the random effects jointly by block moves. A detailed updating algorithm and mean and variance parameters of the full conditionals can be found in Lang and Brezger (2001).

6.2. **Unobserved heterogeneity with an ordinal response.** Consider an ordinal response variable Y, that is assumed to be a categorized version of a latent variable $\widetilde{Y}$ with thresholds $-\infty = \theta_0 < \theta_1 < ... < \theta_k = \infty$. Thus, there are $k-1$ parameters to estimate in addition to the unknown coefficient parameters. The thresholds $\theta = (\theta_1, ..., \theta_{k-1})'$ are considered as random. Like fixed effects (see (**??**)), they are supplemented with diffuse priors, i.e.

$$p(\theta) \propto const. \tag{6.7}$$

The additive model with an ordinal response does not differ from the (6.3) except for the meaning of the response latent variable $\widetilde{Y}$.The full conditional distribution of the latent variable is a truncated standard normal distribution, with truncation points determined by the thresholds as

$$P(Y|\widetilde{Y}) = \sum_{r=1}^{k} I(\theta_{r-1} < \widetilde{Y} \leq \theta_r) I(Y = r) \tag{6.8}$$

Drawing out of a truncated normal distribution evolves as numerically difficult and almost not solvable together with random effects. Thus, estimation procedures based on separated steps are needed. For details on sampling schemes on ordinal data and reparametrization strategies see Fahrmeir and Lang (2001).

7. P-Splines with mixed effects to investigate MS clinical prognostic factors.

In this section we propose P-Splines with mixed effects as a suitable model to address the MS clinical data. As discussed in section 2 a crucial issue in MS research is to investigate the impact of covariates in determining the severity of the disease. We now apply the P-Splines mixed effects models to the SLCMSR data set above introduced to show how this modelling can afford the high level of heterogeneity of MS data and can provide important information about the role of the prognostic factors in the disease. The included covariates which are some of the most important prognostic factors for MS are described in Table A in Appendix I. The results are provided on the analysis of MS data with the following models:

1. P-Splines random intercept model
   - with a Gaussian response
   - with an ordinal response
2. P-Splines random slope model

7.1. **P-Splines random intercept model with a Gaussian response.** The influence of the covariates on the change in EDSS is estimated with Bayesian techniques.For the metric variables, 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 non-linear effects of time may be detected. The further introduction of a random slope will require a linear term for time.

Let now take the response variable as normally distributed. The prior distribution functions for the parameters are those chosen in the previous section as referred to the general model (6.3).

The model can be specified by the formula

$$\begin{aligned} changew_{it} &= f_1(t_i) + f_2(age_i) + f_3(edss_i) + f_4(dur_i) + \beta_1 * course_i^{(1)} \\ &\quad +\beta_2 * course_i^{(2)} + \beta_3 * gender_i + b_i + \varepsilon_{it}, \end{aligned} \tag{7.1}$$

where $b_i, i = 1, ...N$ is the random intercept, identified by an unique index variable. Bayesian P-Splines functions are modelling non-parametrically the impact of four covariates. Fixed effect are acting additively as above specified.

The estimates are perfomed with the software BayesX (www.stat.uni-muenchen.de/∼lang/bayesx/bayesx.html
The posterior mean plots illustrate the role of the risk factors on the EDSS 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 taking a burn-in period of 20000 and step width 500. The number of iterations was therefore set to 520000, so that 1000 samples were stored. With these parameters, a good behavior of the chain was obtained. Fig. 5 shows the sampling and autocorrelation plots of the constant effects and gender, and of one parameter for the time effect. All other autocorrelation and sampling plots are comparable to the examples showed.

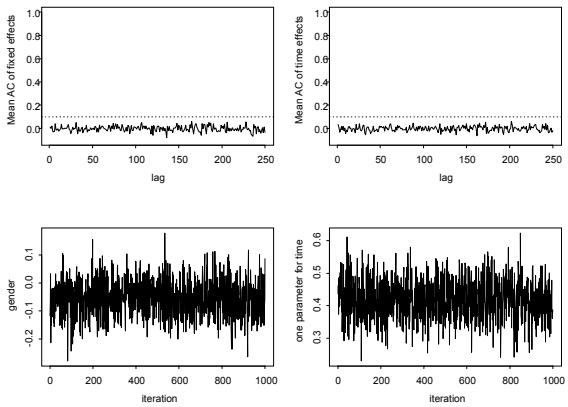


FIGURE 5. Autocorrelations of fixed effects and parameters for time (top) and mixing behavior of the estimate for gender and one time parameter (bottom)

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

TABLE 1. Estimates of variance components

The results of the variance components and fixed effects estimates are reported in Tables (1) and (2).

We notice 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. This leads to conclude that the prognostic factors included depict an average patient profile which is not representative of the population.

As concerning the fixed effects Table (2) shows how the gender of the patients (female are reference category) has a negligible impact on the EDSS change, although a slight increased risk is detected for men consistently with the MS literature. Actually, an effect of the variable "course" can 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 this result can be interpreted as a short term predictors only. In fact, it is very

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

TABLE 2. Estimates of constant effects

difficult to attribute any other role to this predictivity given that the course may change during the disease course.

Posterior means are plotted in Fig.s 7a, 7b, 7c, 7d.. These plots show an increasing linear effect of "time" on EDSS change (Fig. 7a), thus suggesting that longer studies allow to observe a higher worsening in the patients. The variable "age at onset" (Fig.7b) has a negligible impact on the EDSS change (credible interval of the posterior mean includes zero) although a slight negative trend after 30 years might suggest a decreasing risk impact of age on MS. This results seems to be inconsistent with MS literature where age is known to have an increasing risk effect on the MS severity. An explanation can be found to this apparent inconsistency. Actually, it can be shown (Di Serio, 2003) that age affects increasingly MS as regarding the initial level of the disease severity, but it doesn' t show a significative impact on the intensity of progression of MS. Thus, age effect is not detected whenever the response variable is the EDSS change over time representative of the MS trend over time.

More informative are the variables "EDSS baseline" and "duration". We can notice by Fig. 7c that patient with low initial EDSS present a lower probability of worsening, whereas this effect reverts for patients who enter the study with high initial EDSS ($EDSS = 5 \div 6$). Furthermore, a constant trend is detected between level 2 and 5 of EDSS baseline: these patients are in MS literature revealed to be relatively stable regarding ambulation disability. Indeed, a remarkable trend is attributable at the lowest and highest EDSS baseline levels. 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 worse their general functional status rapidly (at these levels EDSS is also computed for 0.5 steps amplitude). Finally we notice that the variable "duration" (Fig. 7d) shows also a surprising decreasing risk effect. Again this can be interpreted together with the above discussed effect of "age at onset" in the extent that these two variables might include the same information.

It can be noticed that the credible intervals at the Splines tails oft happen to widen. This can be explained by noticing that, due to standard inclusion criteria, in the clinical trials pooled for this analysis a lower number of patients presents extreme values for the analysed covariates.

Overall, it has to be noted, that not all included effects influence the response variable significantly. This is also affected by the modelling. The plot of the population residuals (Fig. **??**) 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 Fig. (7) shows a systematic trend in the residuals: 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 estimations are shifted towards less change in EDSS. This again support the idea that a high amount of variation is explained by the random effects.

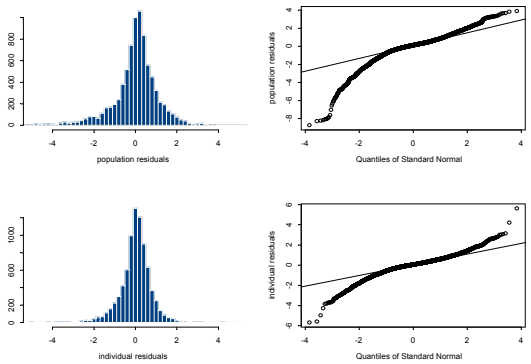


FIGURE 6. Histograms and normal-quantile plots for population residuals (top) and individual residuals (bottom).

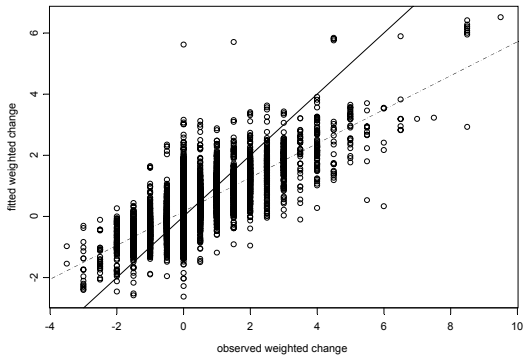


FIGURE 7. Plot of observed against fitted values (dashed line: linear regression line of the scatter plot; full line: the diagonal)

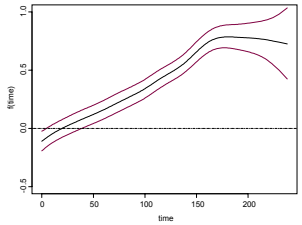


Figure 7a. P-Spline for time (in weeks)

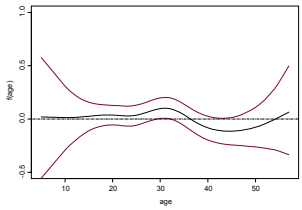


Figure 7b. P-Spline for age at onset

thresholds	changeord	
$\leq \theta_1$	big decrease	$(<<)$
$(\theta_1; \theta_2)$	small decrease	$(<)$
$(\theta_2; \theta_3)$	stable	$(=)$
$(\theta_3; \theta_4)$	small increase	$(>)$
$\geq \theta_4$	big increase	$(>>)$

TABLE 3. Boundaries of thresholds and their corresponding categories

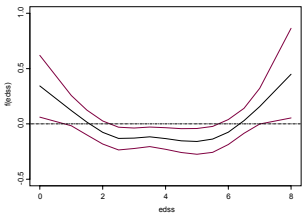


Figure 7c. P-Spline for baseline EDSS

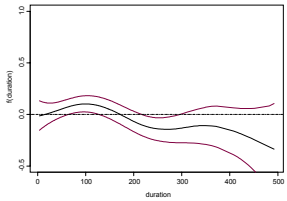


Figure 7d. P-spline for duration

7.2. **P-Splines random intercept model with an ordinal response.** In the previous modelling the ordinal nature of the weighted EDSS change was not taken into account. Let now include in the P-Splines analysis the ordinal structure of the data. Based on a comparison between the two modelling we aim at understanding whether different results arise concerning the effect of MS prognostic factors.

In an ordinal model, as above discussed, the posterior mean estimates depend also on the thresholds parameter vector $\theta = (\theta_1, ...\theta_{k-1})'$. For the sake of simplicity the 25 categories of EDSS change where reduced to 5 , as in Table B in Appendix I: Each category included at least a change of 1.0 on the EDSS-score to be confident, according with MS literature (Noseworthy et al., 1990), 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.

The general ordinal threshold model, according with the Gaussian response P-Splines model can be now written as

$$\begin{aligned} changeord_{it} &= f_1(t_i) + f_2(age_i) + f_3(edss_i) + f_4(dur_i) + \beta_1 * course_i^{(1)} \\ &\quad +\beta_2 * course_i^{(2)} + \beta_3 * gender_i + b_i + \varepsilon_{it}. \end{aligned} \tag{7.2}$$

The ordinal response variable is categorized as described in table (3). The prior distributional assumptions are the same described for model (6.3). In addition, a diffuse non-informative prior $p(\theta) \propto const$ was chosen.

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 as above. These estimates constitute the basis for the marginal likelihood estimation of the random effect, as performed by the software MIXOR (www.uic.edu/∼hedeker/mix.html). In the first Bayesian step, as in the Gaussian response model, P-Splines of degree 3 with second order random walk penalty where

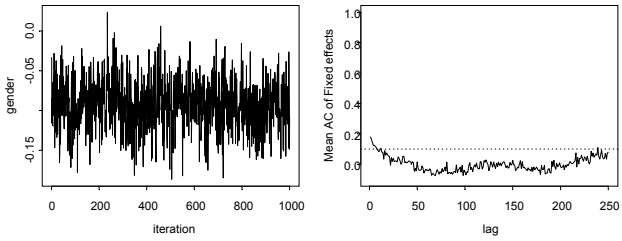


FIGURE 8. Sampling plot of gender (left) and autocorrelation plot of fixed effects (right)

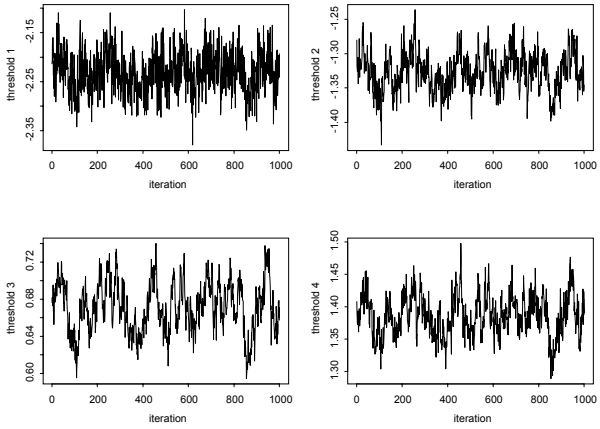


FIGURE 9. Sampling plots of threshold parameters

considered. Convergence and mixing behavior of the MCMC procedure 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 500000 and a step-width of 1000 was chosen.

The diagnosis plots of the constant effects, shown in Fig. 8, are also satisfying. However, the trace plots of the threshold parameter samples (Fig. 9) illustrate a bad mixing behavior. Positive and negative correlations seem to alternate. Hence, the estimation of threshold parameters is not very stable.

Fixed effects Bayesian estimates provide information to reduce the number of parameters and to construct an appropriate ordinal regression model where the smooth functions are chosen as polynomial. For details see Lamina (2002).

Let now present the final estimation results obtained by this mixing two-steps 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 showing progressive and can be so interpreted: the estimated effect of "course $^{(2)}$", for instance, is about 0.62, then it lies between θ_1 and θ_2 and the category "small decrease" in disability corresponds then to the predicted outcome, according to Table (4).

threshold	Estimator	std.error	z-value	p-value
θ_1	0	–	–	–
θ_2	1.46894	5.66944	44.35629	<0.0001
θ_3	4.59204	0.03922	117.09799	<0.0001
θ_4	5.66944	0.04102	138.22608	<0.0001

TABLE 4. Estimates of threshold parameters

	Estimate	Std.Error	z-value	p-value
gender	-0.08919	0.08839	-1.00902	0.31297
course$^{(1)}$	0.54872	0.17777	3.08677	0.00202
course$^{(2)}$	0.62386	0.12768	4.88632	0.00000

TABLE 5. Estimates of constant effects

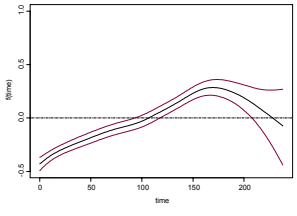
		fitted category					
		<<	<	=	>	>>	Total
observed category	<<	**10**	61	47	0	0	118
	<	16	**120**	556	4	0	696
	=	85	128	**4684**	278	6	5181
	>	2	15	704	**661**	17	1399
	>>	0	6	154	453	**261**	874
	Total	113	330	6145	1396	284	8268

TABLE 6. Crosstab of observed and fitted response

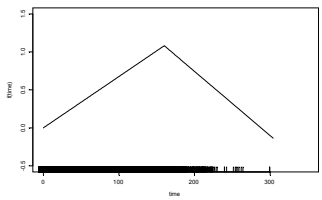
In the next plots the Splines obtained in each of the two estimation steps of the ordinal modelling are compared.

Fig. 9A. **STEP I** (estimation with no random effect)

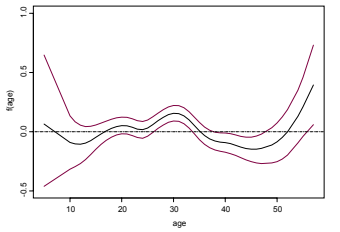
STEPII (estimation with mixed effect)



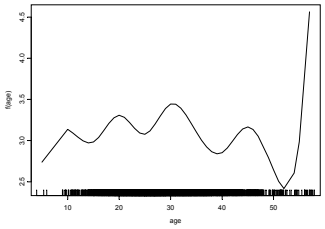
bayesian P-Spline for time



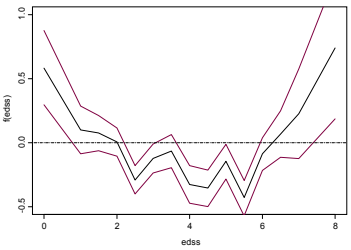
Regression spline for time



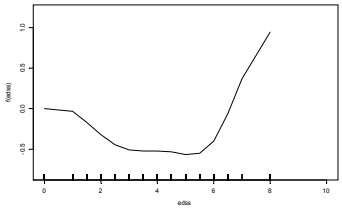
bayesian P-Spline for age



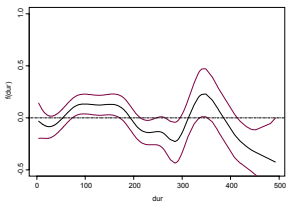
Regression spline for age



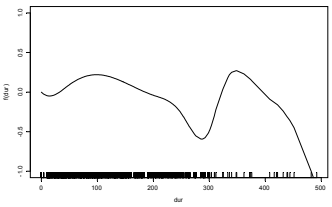
bayesian P-Spline for bas. EDSS



Regression spline for baseline EDSS



bayesian P-Spline for duration



Regression spline for duration

Preserving the ordinal nature of EDSS change did not provide evidence of a change in the interpretation of the estimated parameters with respect to the Gaussian model. Results (Fig. 9A), that appear very different in the first sight, like the regression Splines for duration and age at onset, don't show such a discrepancy, when looked at closer. Due to different outcome variables, the estimates cannot be compared directly. Yet the hypothesis that effects appear to be of bigger influence in the ordinal model, has been confirmed. Furthermore, the variance of the estimations has to be taken into account. Unfortunately, confidence bands couldn't be added in the plots of regression Splines obtained in MIXOR. But the P-Splines of the ordinal model in BayesX as well as the rough plots can serve as indicators, how accurate the estimation is.

As in the Gaussian model, the model fit is analyzed by comparing the fitted values against the observed values. The crosstab in 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 "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 very well by the ordinal model as well as the Gaussian model. Moreover it has to be noted, that many computational problems occured during the estimation of the ordinal model. The autocorrelation and trace plots of the threshold samples (Fig.s 9 ,8) showed a bad mixing and convergence behavior, although random effects havn't been included in this stage of modelling. Analyzing the mixed-effects ordinal regression model in

MIXOR also led to numerical difficulties. Adjustments had to be made 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 don't differ substantially.

From the reported analyses of the fitted values against observed values concerning both the random intercept models, a systematic bias has been revealed. The analysis of residuals also suggested that additional random components should be included in the analysis.

We next formulate, as last step of our modelling, a random slopes model.

7.3. **P-Splines random slopes model.** Heterogeneity in individual MS progression is observed as regarding both the magnitude and the speed: The disability of one patient may rise fast in the beginning and then stabilize, whereas it rises steadily, but slow for another patient. This might be the cause of the bias in the fitted values derived by the random intercept models: They underestimate the change for patients, whose disability greatly decreased or increased within the time frame of a clinical study. Thus, the introduction of a random slope together with a quadratic random effect should account for both the different magnitude of MS progression as well as the different curvature in the progression of disability. The choice of a quadratic random slopes model is also suggested by the nonparametric effect of time detected in the Splines.

We recall that in a random slopes model the patients in a clinical trial are considered to different 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

$$
\begin{aligned}
changew_{it} \;=\; & f_1(age_i) + f_2(edss_i) + f_3(dur_i) + \\
& \beta_1 * course_i^{(1)} + \beta_2 * course_i^{(2)} + \\
& \beta_3 * gender_i + (b_{i0} + \beta_0) + \\
& (b_{i1} + \beta_4) * t_i + (b_{i2} + \beta_5) * t_i^2 + \varepsilon_{it},
\end{aligned} \tag{7.3}
$$

where β_{i0} is the random intercept, β_{i1} the random slope and β_{i2} the quadratic random slope parameter. The fixed intercept is denoted by β_0, the fixed time by β_4 and the fixed quadratic time effect by β_5. As in the Gaussian model, prior distributions of all fixed effects are diffuse, whereas all random effects are assumed to be normally distributed.

To ensure comparability of the random intercept and random slopes model, calculation in BayesX was performed with the same number of iterations, i.e. a burn-in of 20000 and a step-width of 500. The convergence and mixing behavior was comparable to the ones obtained in the random intercepts model (Fig. 5). Hence, non of the Gaussian models is numerically superior.

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.

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 7. Estimates of variance components

The estimation of the constant effect are 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). Hence, 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.

Variable	Mean	Std.Var.	10% Qu.	50% Qu.	90% Qu.
intercept	0.017732	0.006067	0.352562	0.360414	0.368427
time	0.004538	0.001049	0.003242	0.004499	0.005897
$time^2$	$1.19*10^{-5}$	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

TABLE 8. Estimates of constant effects

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

TABLE 9. Mean of random slope estimates, stratisfied for courses

P-Splines curves plotted for the metric variables "age at onset", "baseline EDSS" and "duration" are here omitted given that it didn't change substantially the information obtained by the previous analysis, except for the credible intervals that resulted noticeably narrower than before.

Finally, by looking at the plot of fitted against observed values (Fig. **??**) 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.

8. Conclusions.

This paper was aimed at providing a suitable statistical tool to capture the heterogenous structure of longitudinal MS data. A non-parametric approach allowed to avoid restrictive assumptions about the analytical form of the relation between explicative covariates and prognostic factors. Therefore generalized additive models (GAM) have been suggested as a natural tool to investigate non-parametrically, by means of Splines, the role of MS prognostic factors. Furthermore, this modelling was combined with mixed effects theory, thus allowing for including both observed and unobserved heterogeneity.

In the presented analysis we addressed two main issues.

i) 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?

ii) 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 is the progression?

To provide an answer to the first question mixed effect Splines models have been compared for Gaussian and ordinal responses. Overall it was shown, that the numerically demanding and time-consuming estimation of an ordinal mixed effects

model is not justified by a real gain in the results. Actually, the interpretation of the role of prognostic factors didn't change dramatically. Thus, a Gaussian model is suggested.

Overall, the entry EDSS appears to be of big influence on the weighted change in EDSS. Patients enrolled in a study with an EDSS lower than 2 or bigger than 5.5 can be expected to experience a higher increase in disability as patients in between. The effects of duration of disease, age at onset and gender are marginal and even negligeable. However, there is a general "positive" time trend during a clinical study. That is, the more time elapses, the higher is the change in EDSS. The course of the disease also emerged as a 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.

Introducing random slopes models allow for investigating the second issue mentioned above. The random slopes model explains a "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 a random effects for the 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 was 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, i.e. to estimate one slope for each group of patients.

Bayesian methods, based on MCMC algorithms, emerged as extremely flexible. Models, that are too complex for classical maximum likelihood estimation, can be estimated in a Bayesian context. The software BayesX was used which incorporates of smooth P-Splines for metric covariates. Alternatively, a combination of Bayesian techniques and marginal likelihood estimation procedures were used for estimating ordinal threshold model .

References

[1] Brezger, A., "Bayesianische p-splines." Master's thesis, Department of Statistics, University of Munich, 2000.

[2] Di Serio, C., "Non-linear mixed effects methodology for modelling Multiple Sclerosis data", *MIMEO, 2003.*

[3] McCullagh, P., and Nelder, J., *Generalized Linear Models.* Chapman and Hall, 1989.

[4] Eilers, P., and Marx, B., "Flexible Smoothing using B-splnies and Penalized Likelihood". *Statistical Sc.*,11, 1996.

[5] Fahrmeir L., and Lang S., "Bayesian semiparametric regression analysis of multicategorical time-space data", *Annals of the institute of Statistical Mathematics*, 53, 11-30, 2001.

[6] Fahrmeir L., and Tutz G., *Multivariate statistical modelling based on generalized linear models.,* Springer Verlag, New York, 2001.

[7] EMEA. "Note for guidance on clinical investigation of medicinal products for the treatments of multiple sclerosis", 2001, URL: www.emea.eu.int/pdfs/human/ewp/056198en.pdf.

[8] Gilks et al. Markov Chain Monte Carlo in Practice, Chapman & Hall, 1996.
[9] Hedeker D., and Gibbons R., "A random effects ordinal regression model for multilevel analysis", *Biometrics,* 50, 933-944, 1994.
[10] Lamina C., "Modelling the Disease Course of Multiple Sclerosis with Mixed Effect, Diplom Arbeit, Department of Statistics, University of Munich, 2002.
[11] Marx, B., and Eilers, P., "Direct Generalized Additive Modeling with Penalized Likelihood", *Computational Statistics and Data Analysis*, 28, 193-209, 1998.
[12] Noseworthy J., *et al.* , "Interrater variability with the expanded disability status scale (EDSS), and functional systems (FS) in multiple sclerosis clinical trial", *Neurology ,* 40, 971-974, 1990.
[13] Searle, S.R., Casella, G. and McCulloch, C. E. (1992). *Variance Components.* Wiley, New York.
[14] Tutz G, *Die Analyse kategorialer Daten,* Oldenbourg Wissenschaftesverlag, Munic, 2000.
[15] VerbekeG, and Molenberghs, *Linear Mixed Models for Longitudinal Data,* Series in Statistics, Springer, New York, 2000.
[16] Vonesh, E. F. and Chinchilli, V. M. (1997). Linear and Nonlinear Models for the Analysis of Repeated Measures, Marcel Dekker, NewYork.

9. APPENDIX I

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	$\text{course}^{(1)} = \begin{cases} 1 \text{ if course = pp or pr} \\ 0 \text{ otherwise} \end{cases}$ $\text{course}^{(2)} = \begin{cases} 1 \text{ if course = sp} \\ 0 \text{ otherwise} \end{cases}$ Reference category is relapsing-remitting

Table A: Description of the covariates included in the analysis of SLCMSR data set.

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

Table B: Categorization of EDSS change.

E-mail address: University Vita-Salute S. Raffaele, Milan.
URL: diserio.clelia@hsr.it