Gehrmann, Hellriegel, Neiss, Fahrmeir:

# Analysis of the time to sustained progression in Multiple Sclerosis using generalised linear and additive models

Sonderforschungsbereich 386, Paper 354 (2003)

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

*Discussion paper 354 - SFB 386 - LMU Munich*

# Analysis of the time to sustained progression in Multiple Sclerosis using generalised linear and additive models.

Gehrmann, U. [1], Hellriegel, B. [1,2], Neiss, A. [1,2] and Fahrmeir, L. [3]

[1] Institute for Medical Statistics and Epidemiology, TU Munich

[2] Sylvia Lawry Centre for MS Research

[3] Institute for Statistics, LMU Munich

September 2003

## Abstract

The course of multiple sclerosis (MS) is generally difficult to predict. This is due to the great inter-individual variability with respect to symptoms and disability status. An important prognostic endpoint for MS is the expected time to sustained disease progression. Using the Expanded Disability Status Scale (EDSS) this endpoint is here defined as a rise of 1.0 or 0.5 compared to baseline EDSS (≤5.5 or >5.5) which is confirmed for at least six months. The goal of this paper was threefold. It aimed at identifying covariates which significantly influence sustained progression, determining size and form of the effect of these covariates and estimating the survival curves for given predictors. To this end a piecewise exponential model utilizing piecewise constant hazard rates and a Poisson model were devised. In order to improve and simplify these models a method for piecewise linear parameterization of non-parametric generalized additive models (GAMs) was applied. The models included fixed and random effects, the posterior distribution was estimated using Markov Chain Monte Carlo methods (MCMC) as well as a penalized likelihood approach and variables were selected using Akaikes information criterium (AIC). The models were applied to data of placebo patients from worldwide clinical trials that are pooled in the database of the Sylvia Lawry Centre for Multiple Sclerosis Research (SLCMSR). Only with a pure exponential model and fixed effects, baseline EDSS and the number of relapses in the last 12 month before study entry had an effect on the hazard rate. For the piecewise exponential model with random study effects there was no effect of covariates on the hazard rate other than a slightly decreasing effect of time. This reflects the fact that unstable patients reach the event early and are therefore eliminated from the analysis (selection effect).

*Key words: Piecewise exponential model, piecewise constant hazard rates, Markov Chain Monte Carlo methods (MCMC),spline, regression, duration models, Expanded Disability Status Scale (EDSS), relapse*

## 1 Introduction

Multiple Sclerosis (MS) is one of the most common neurodegenerative diseases (prevalence in Germany ~ 0.1% ). Disease onset lies between 20 and 40 years of age and women are affected twice as often as men. This complex autoimmune disease destroys the myelin layer of nerve fibres which leads to errors in signal transduction. MS is known to be influenced by genetic and environmental factors but its cause is still unknown. Moreover, the great heterogeneity in disease course results in insecurity of diagnosis and prognosis. A common measure for the degree of disability in MS is the Expanded Disability Status Scale (EDSS, ordinal scale ranging from 0 (=no symptoms)

to10 (death due to MS) with steps of 0.5). The disease course is thought to consist of two distinct phases which are not necessarily both passed through. The first phase is characterized by a stable level of disability with or without superimposed relapses, i.e. sudden increases in EDSS, which are (almost) completely resolved. In the second phase the disease progresses continuously. The time point when the disease course „switches“ from its more or less stable to its progressive phase is not easily determined. The corresponding event is termed "sustained/confirmed progression" in disability and the time taken to reach this change point is used as endpoint in clinical trials (e.g. Kappos et al.. 1998; Liu and Blumhardt 2000; Leary et al. 2003).

The goals of this study were to identify important predictors, to determine size and form of their effect and to prognose the risk with respect to sustained progression for covariates such as age at disease onset, gender, number of relapses in the last one or two years before and EDSS at entry into the study. The main model used to reach these goals was a piecewise exponential model using piecewise constant hazard rates (Fahrmeir and Tutz 2001). It combined parametric (exponential, GLM) and non-parametric (piecewise, GAM) modeling in a flexible way and also accounted for fixed and random effects. Variables were selected using Akaikes information criterium (AIC). In order to improve the model, non-parametric GAMs were transformed into simpler GLMs by parameterizing them piecewise linearly. The data employed in the analyses came from a large pooled database of placebo patients from clinical trials which have been donated to the Sylvia Lawry Centre for Multiple Sclerosis Research (SLCMSR) by institutions and companies (Noseworthy et al. 2003).

## 2 Database and patient population

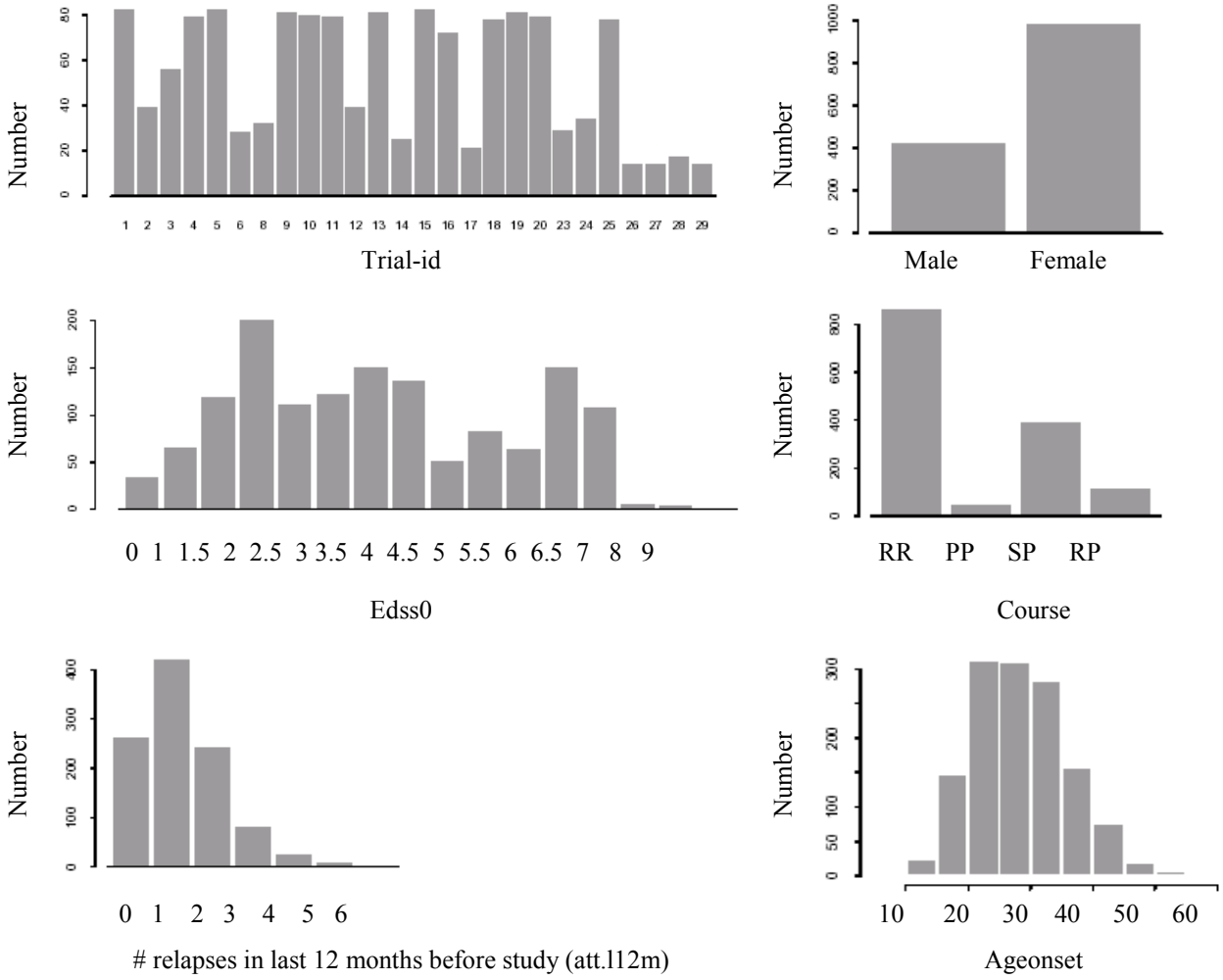
The SLCMSR database combines data from placebo arms of clinical trials and from patient registries from all over the world (see acknowledgements). For confidentiality it is not possible to identify the individual sources of the pooled data. Moreover, to warrant a high quality of statistical analyses performed at the SLCMSR the database is randomly split into two subsets, an open part which is used for model building and hypotheses generation and a closed part which is administered by trustees. It is the declared policy of the SLCMSR to validate analyses performed on the open part of the database using its closed counterpart. The results presented here have not yet undergone the validation procedure and have, therefore, to be considered as preliminary.

Sufficient information was available on the following covariates (<5% missing values, N=1396): gender; age at onset of disease (ageonset, [years]); age at entry into the study (ageentry, [years]); duration of MS [months]; EDSS at entry (edss0); and disease course (RR: relapsing remitting, SP: secondary progressive, RP: relapsing progressive, PP: primary progressive). In addition, the following covariates with more missing values (NA) were included because of their potential importance: number of relapses during the last 12 (att.l12m, 25.9% NA) and/or 24 months (att.l24m, 14.1% NA) before entering the study. There was also a label indicating patients that belonged to the same study (trial-id). The distributions of a selection of these covariates are shown in Fig. 1, especially the ones proving to be important.

The analyses utilized the data on placebo patients from 26 clinical trials gathered in the SLCMSR database which in release 005 contains 1754 patients. Only 1396 of them had a definite diagnosis of MS, an EDSS value

also for the first visit and a censoring time point greater than zero.

**Figure 1**. Histogramms for important covariates.




## 3 Endpoint and time scale

The endpoint of interest was the time point when the disease course switches to its progressive phase. This change point was not explicitly given in the database but had to be determined from the time series of EDSS values. As disease progression has to be confirmed for some time, the event of interest is referred to as "sustained/confirmed progression".

**Definition 1. Baseline EDSS**

The EDSS value at entry into the study, i.e. the first observation (edss0), is termed baseline EDSS.

**Definition 2. Sustained progression**

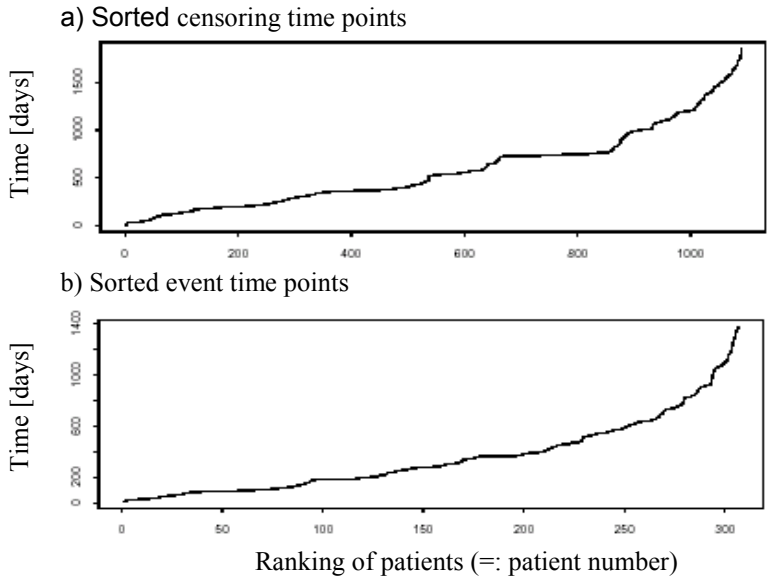
If baseline EDSS $\leq 5.5$ then increase level = baseline EDSS + 1.0

If baseline EDSS $\geq 6.0$ then increase level = baseline EDSS + 0.5

The event occurs if EDSS ≥ increase level for at least six months. The time point at which this event occurs is given by the start of this six months period when the increase level is reached first.

Usually, in clinical studies observations are collected at regular time intervals. This seems to suggest the use of a fixed grid of discrete time steps and discrete survival models for the analyses. However, when the assumption of discrete steps was checked for the pooled placebo data it did not hold. Plots of event and censoring time points sorted by magnitude showed rather continuous curves (see Fig. 2), probably due to the fact that studies with different visit schedules were pooled. Therefore, a continuous time scale was assumed and adequate models were chosen.

**Figure 2**. Sorted event and censoring time points.




## 4 Selection of a Poisson model

Assuming (piecewise) exponentially distributed survival times leads to Poisson models. These are special generalized linear models (GLM) where the response variable $y_i$ follows a Poisson distribution with Parameter $\lambda_i$ (for i=1,2, …., n observations) and *log* is the natural link function. The EDSS data used here followed a Poisson process, i.e. the number of events occurring in an interval of length $\Delta_i$ (time under risk), were Poisson distributed with the parameter $\Delta_i \lambda_i$. That is, there was an additional dependence on the length of the interval in which the i-th observation was made, leading to a Poisson model with offset (e.g. Tutz 2000)

$$P(y_i \mid x_i) = \begin{cases} \dfrac{(\Delta_i \lambda_i)^{y_i}}{y_i!} e^{-\Delta_i \lambda_i} & \text{for } y_i \in N_0 \\ 0 & \text{otherwise} \end{cases}$$

This resulted in an additive constant in addition to the linear predictor $x_i'\beta$ of the Poisson model

$$\log(\mu_i) = \log(\Delta_i \lambda_i) = \log(\Delta_i) + x_i'\beta$$

and statistical inference was possible using maximum likelihood (ML) methods. For generalized additive models (GAM) the predictor is augmented by a non-parametric term for continuous covariates (see Hasit & Tibshirani 1990)

In order to select the best model Akaike's information criterium (AIC) was used. For GLMs and more specifically for Poisson models a modified AIC* can be estimated that derives from AIC through multiplication with the dispersion parameter $\phi$ (e.g. Tutz 2000):

$$AIC^* \propto \Phi \cdot AIC = -2\sum_{i=1}^{n} y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) + 2p\Phi \propto -2\sum_{i=1}^{n} y_i \log(\hat{\mu}_i) + 2p\Phi$$

where $\hat{\mu}_i = \exp(x_i'\hat{\beta})$ and $\hat{\beta} \overset{as}{\sim} N\left(\beta, F^{-1}(\hat{\beta})\right)$ with $F$ being the Fisher matrix.
While for GLMs $p$ is the number of parameters fitted, for GAMs p additionally accounts for the effective number of parameters or "degrees of freedom" (df).

The moment estimator for the dispersion parameter $\phi$ in the Poisson model can be determined using the Pearson statistics for goodness of fit ($\chi_p^2$) . For model comparison $\phi$ is estimated from

$$\hat{\Phi} = \frac{1}{n-p}\sum_{i01}^{n}\frac{(y_i-\hat{\mu}_i)^2}{v(\hat{\mu}_i)} = \frac{1}{n-p}\chi_p^2,$$

based on the greatest possible GLM.

To be able to include GAMs and to account for varying numbers of missing values for the different covariates the following approach was taken:

(i) In a first step, only covariates with no or few missings were considered. For those a full GAM with splines with df=4 was fitted to detect characteristic forms of influence of covariates. Then variables were selected using AIC and splines with df between 1 and 4 in steps of 1/10. Finally the selected GAM was parameterized using the functional form detected if this lead to a decrease in AIC.

(ii) In a second step, covariates with greater numbers of missing values were included into the model if this decreased the AIC. This was done for each covariate separately so that the number of missings did not sum up. The rest of the procedure was as described above.

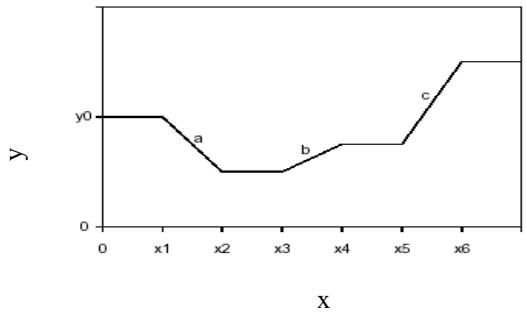
*Parameterizing a GAM to a GLM*

Once a specific GAM is selected as optimal one in principal could stop. However, there are two arguments for transforming this GAM into a suitable GLM. First, the penalty term of the AIC is higher for GAMs because it also accounts for the degrees of freedom of the fitted splines (see above). Second, GLMs are usually considered to be more comprehensible and tractable, because they can be described by simple formulae.

In order to find an adequate parameterization the GAM has to be inspected. In the case of the placebo data only piecewise linear functions with a maximum of three different slopes proved to be suitable (Fig. 3), although in

principle any kind of polynomial function of the covariates would be possible.

**Figure 3**. Piecewise linear influence with three slopes. The parameters $y_0$, a, b, and c have to be estimated.




## 5 Models with mixed effects

Mixed effect models are obtained if random effects are additionally included into the predictor of a GLM or GAM. Including random effects is especially advised in cases of overdispersion, i.e. $\hat{\Phi} >> 1$. Their inclusion structures the data into clusters. With q clusters and $n_i$ observations a single observation is described by $(y_{ij},x_{ij})$ $(i \in \{1,...,q\}, j \in \{1,...,n_i\})$ and the i-th cluster has the following form

$$(y_i, x_i) = (y_{i1},...,y_{in_i}, x_{i1},...,x_{in_i})$$

For estimation of a GLM with mixed effects the ML method is still valid like in the case of fixed effects only.

For the specific case of a Poisson model with offset different approaches are possible. Following Draper (1996) in this study a model with normally distributed random effects was chosen which has the following form

$$y_{ij} \mid \lambda_i \overset{indep}{\sim} Poisson(\lambda_i)$$ with *log* as link function,

$$\eta_{ij} = \log(\mu_{ij}) = \log(\Delta_{ij}\lambda_{ij}) = \log(\Delta_{ij}) + x_{ij}'\beta + z_{ij}'b_i$$ and the random effects

$$b_i \mid \sigma^2 \overset{indep}{\sim} Normal(0,\sigma^2)$$

This model can be estimated using the procedure NLMIXED in SAS which allows for the inclusion of normally distributed random effects and uses adaptive Gauß quadrature (after Pinheiro and Bates 1995) for integration and Quasi Newton optimization for maximizing the log-likelihood (Dennis and Mei 1979; Gay 1983).

GAMs with mixed effects can best be estimated using Bayesian models. Here the covariates are sorted into three types: $x_{ij}$ for $p_s$ non-linear effects with spline modeling, $w_{ij}$ for $p_f$ linear effects and $z_{ij}$ for $p_z$ random effects.
The predictor $\eta_{ij}$ then has the more general form

$$\eta_{ij} = \alpha + f_1(x_{ij1}) + ... + f_{p_s}(x_{ijp_s}) + w_{ij}\gamma + z_{ij}b_i$$

where $\gamma$ is the $p_f$-dimensional vector of the linear fixed effects and $b_i$ the $p_z$-dimensional vector of the random effects of the cluster i. The smooth functions $f_1$, …, $f_{ps}$ are linear combinations of basis functions such as

polynomial splines, B-splines or wavelets (Fahrmeir et al. 1996). The basis functions used for the EDSS data were cubic B-splines $f_{k_S}(x_{ijk_S}) = \xi_{ijk_S}\beta_{k_S} \quad with\ k_S = 1,...,p$ and

$\xi_{ijk_S} := (B_1(x_{ijk_S}),...,B_r(x_{ijk_S}))$ being the ij-th row of the B-spline matrix for effect $k_s$ (e.g. Lang and Brezger 2003).

The prior distribution for the fixed effects with spline modeling then is

$$\beta_{k_S} \propto \exp\left(-\frac{1}{2\tau^2_{k_S}}\beta'_{k_S}K_{k_S}\beta_{k_S}\right)$$

where $K_{ks}$ is the penalty matrix. The parameter $\tau^2_{ks}$ which regulates the global smoothness of the splines is commonly assumed to be inversely gamma distributed $\tau^2_{k_S} \sim IG(a,b)$ , with a=1 and b=0.005 being reasonable choices.

The priors for linear and random effects are, respectively

$\gamma_{k_f} \propto const.$

$b_{ik_z} \sim Normal(0,\sigma^2_{k_z})$ .

With stochastic independence of $\beta_{ks}$, $\gamma_{kf}$ and $b_{ikz}$ the common posterior distribution for the mixed effects is then given as

$$p(\delta) = \prod_{k_S=1}^{p_S} p(\beta_{k_S} \mid \tau^2_{k_S})p(\tau^2_{k_S}) \cdot \prod_{k_f=1}^{p_f} p(\gamma_{k_f}) \cdot \prod_{i=1}^{q}\prod_{k_z=1}^{p_z} p(b_{ik_z} \mid \sigma^2_{k_z})p(\sigma^2_{k_z})$$

where $\delta$ is the combined parameter vector for all mixed effects.

The joint posterior distribution for all effects, i.e. $p(\delta|y)$, has then to be calculated using simulation techniques. This was done here using MCMC as implemented in BayesX (Brezger et al. 2002).

## 6 Piecewise exponential survival model

The exponential model is the most simple parametric approach in survival analysis. The hazard rate $\lambda(t|x_i)=\lambda_i$ is time independent and for estimating $\lambda_i$ the relation between Poisson process and exponential distribution can be used. With $\Delta_i := t_i$, $y_i := 1$ in case of an event and $y_i := 0$ in case of censoring this leads to the Poisson model with offset discussed above, for which GLM- and GAM-based estimation methods can be used.

The additional advantage of the piecewise exponential model is that the hazard rate is constant only on certain time intervals but that it can jump arbitrarily between them:

$$\lambda(t \mid x_i) = \begin{cases} \lambda_{is} & for\ t \in (a_{s-1}, a_s], \quad s = 1,...,q \\ \lambda_{ik} & for\ t > a_q \end{cases}$$

The hazard rate can then vary at least piecewise and survival function S and density f have the following forms

$$S(t \mid T \geq a_{s-1}, x_i) = \exp(-\lambda_{is}(t - a_{s-1})) \quad for\ t \in (a_{s-1}, a_s]$$

$$f(t \mid T \geq a_{s-1}, x_i) = \lambda_{is} \exp(-\lambda_{is}(t - a_{s-1})) \quad for\ t \in (a_{s-1}, a_s]$$

where T is the time taken until the event occurs or until censoring.

In order to estimate the model, the data set has to be extended by generating a row for each individual i and each interval s where the individual is under risk (e.g. Fahrmeir and Tutz 2001):

$$y_{is} := \begin{cases} 1 & if\ individual\ i\ has\ an\ event\ in\ \text{int}erval\ s \\ 0 & otherwise \end{cases}$$

$$\Delta_{is} := \begin{cases} Event\ time - a_{s-1} & in\ case\ of\ an\ event \\ Censoring\ time - a_{s-1} & in\ case\ of\ censoring \\ a_s - a_{s-1} & in\ case\ of\ surviving\ the\ interval \end{cases}$$

$$x_{is} := x_i \quad \forall s$$

$\lambda_{is}$ is parameterized by $\exp(\eta_{is})$ with predictor $\eta_{is}$ in complete analogy to mixed log-linear or log-additive Poisson models.

## 7 Data analyses

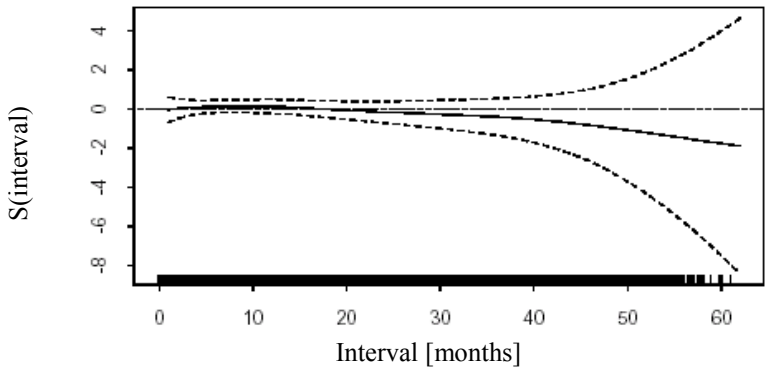
*Piecewise exponential model*

The expansion of the placebo data set was performed using fixed time intervals of 30 days and the time index was termed *interval*. In accordance with the approach described above (see Selection of a Poisson model) cubic splines with df=4 were calculated and plotted for this new time variable. Thereby only the intercept but no covariates were modeled (Fig. 4). The spline value for month *interval* gives the logarithm of the factor by which the hazard rate of this month deviates from the mean hazard rate:

$$\lambda_i \mid interval_i, x_{i2,....}, x_{ip} = (\lambda_i \mid x_{i2,....}, x_{ip}) \cdot \exp(s(interval_i))$$

That the confidence band widens for higher time indices results from the low frequency of observations (see bars in Fig. 4). The fact that the zero line (S(*interval*)=0) never leaves this band implies that over the whole range there was no time influence. This result was confirmed by estimating GAMs with combinations of the covariates without missing values (interval, baseline EDSS, disease course and gender) using cubic splines with df 1 to 4. On basis of the AIC the covariate *interval* was never selected as having a significant influence, implying that a time independent hazard rate can be assumed. This simplifies the models and avoids the expansion of the data set.

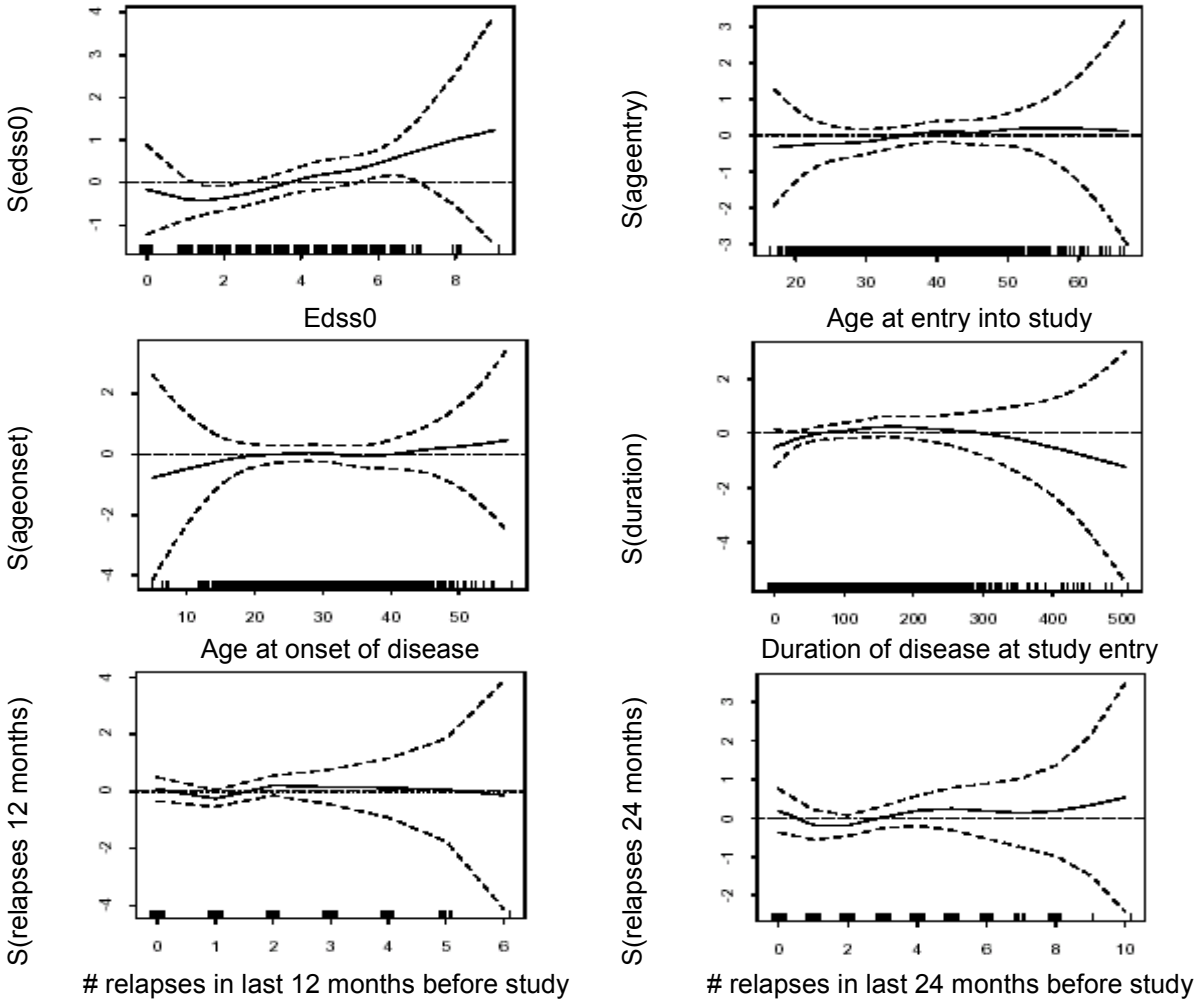
**Figure 4**. Spline of *interval* in piecewise exponential model. The two dashed lines indicate the 95% confidence band and the bars above the x axes give the frequency of observations.




*Pure exponential model*

Possible factors of influence that were considered for the pure exponential model were: baseline EDSS (edss0), disease course, gender, age at disease onset and at entry into the study, disease duration, number of relapses in the last 12 and/or 24 months before entry into the study. Except for gender and disease course the covariates are continuous so that modelling by splines makes sense (Fig. 5).

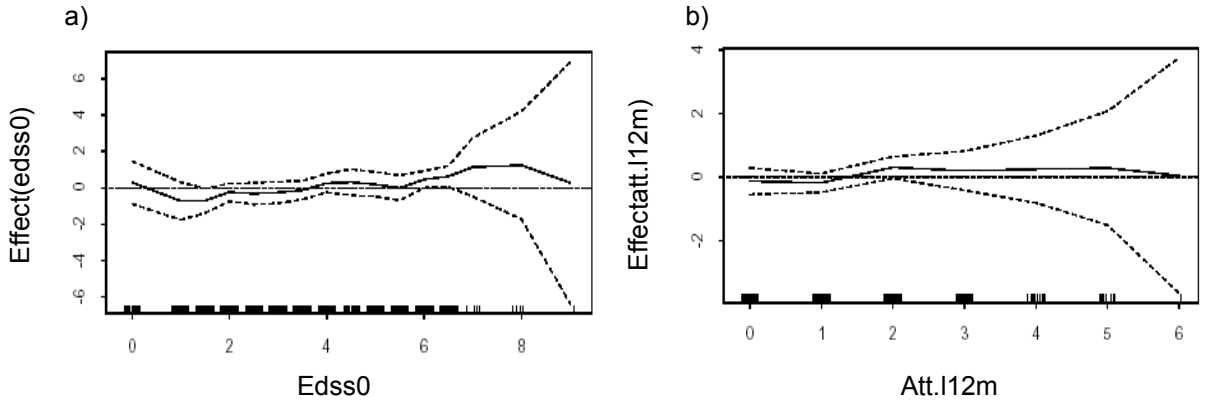
**Figure 5**. Splines in the pure exponential model.



As can be seen from Fig. 5 edss0 has a significant influence which seems to be either purely or piecewise linear. None of the other factors had any detectable influence. Only for the number of relapses in the last 12 months before study entry the confidence band touched the zero line and the form of the influence was suitable for parameterization.

Selection of a GAM with df 1 to 4 for covariates without missing values (edss0, gender and course) revealed edss0 as the only factor suitable for spline modelling. The cubic spline fitted had df=1, leading to a pure GLM. It could , however, still be sensible to parameterize edss0 in a piecewise manner, because of a kink between 1.0 and1.5. Parameterization indeed lowered the AIC if the average effect was 0 for edss0=0, the hazard rate was low at edss0=1 and followed by a linearly increasing effect until 7 and a constant influence from 7 onwards (cf. Figs 6a and 7a).

**Figure 6**. a) Spline of edss0 with degree 10 illustrating the effect of choosing splines of higher degree.
b) Spline of the number of relapses in the last 12 months before study entry in the piecewise exponential model for a given influence of edss0.




Of the variables with missing values only the number of relapses in the last 12 months before study entry (att.l12m) had a noticeable influence (Fig. 5). When edss0 and att.l12m were parameterized simultaneously the spline for att.l12m had to be newly determined. Although its effect was still non-significant the form of the influence suggested a parameterization which was confirmed by AIC: average effect 0 for at most one relapse and a higher hazard for two and more relapses (see Fig. 7b).

**Figure 7**. Parameterized influences of a) edss0 and b) att.l12m.

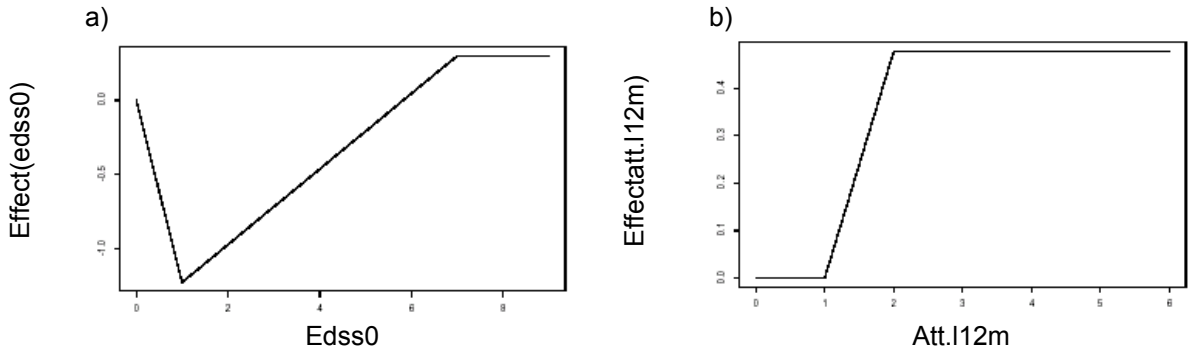


**Table 1**. AIC values for the models with fixed effects only. "Basis with edss0 for att.l12m" is the model with parameterized edss0 but applied to a reduced database excluding the missing values of att.l12m .

| **Model** | **AIC** |
|---|---|
| Null model | 1413.806 |
| Edss0 linear | 1386.216 |
| Edss0 parameterized | 1385.304 |
| Basis with edss0 for att.l12m | 1065.082 |
| Edss0 and att.l12m parameterized | 1059.085 |

The investigations described above and summarized in Table 1 lead to the following model

**Model 1. without random effects**

$$E(\lambda \mid edss0, att.l12m) = \exp(-7.516) \cdot \exp(-1.227 \cdot (I(edss0 \le 1) edss0 + I(edss > 1)) \cdot \exp(0.254 \cdot (I(edss0 \in (1,8]) \cdot (edss0 - 1) + I(edss > 8) \cdot 7)) \cdot \exp(0.477 \cdot (Iatt.l12m = 2) \cdot (att.l12m - 1) + I(att.l12m > 2)))$$

In summary, this means that only the covariates edss0 and att.l12m had a noteworthy influence on the hazard rate and therefore on the survival function.

*Estimation with random patient effects*

Estimation of the dispersion parameter for the piecewise exponential model resulted in the high value of 5.36. A possible reason for this could be cluster formation in the data. In the piecewise exponential model the expansion of the data had produced the cluster "patient". It is therefore obvious to include the covariate patient (patnr) as random effect. As doing so could also reveal a time effect a new piecewise exponential model with normally distributed random patient effects and a time effect *(interval*) with a cubic spline was put up (Model 2).

**Model 2. Bayesian additive mixed model with random patient effects**

$$E(\lambda \mid \text{int}erval) = \exp(\text{int}ercept) \cdot \exp(random(patnr)) \cdot \exp(s(\text{int}erval, \deg ree = 3, RW\ of\ 2nd \deg ree))$$

with patnr ~ $N(0,\sigma^2)$.

Estimation was carried out using MCMC procedures in BayesX (Brezger et al. 2002). Both, using the default settings (52000 iterations in total and 2000 for the burn-in, thinning step of width 50) and drastically increasing the parameter values (550000 iterations in total and 200000 for the burn-in, thinning step width 70) could not resolve the important problem of an unsatisfying auto-correlation structure. For the variance of patnr the auto-correlations of two elements of the sampling path with a distance of up to 45 and with distances between 100 and 145 were larger than 0.1 and therefore too large (see Fig. 8 bottom). The autocorrelations for individual patients, however, were well behaved (Fig. 8). The described structural problem implied that random patient effects

could not be modelled satisfyingly in BayesX. In addition, looking at the estimated time effect did not indicate a possibility to achieve significance (Fig. 9). That is, model 2 anyways would not have brought new insights.

**Figure 8**. Auto-correlations for random patient effects (maximum lag 250).

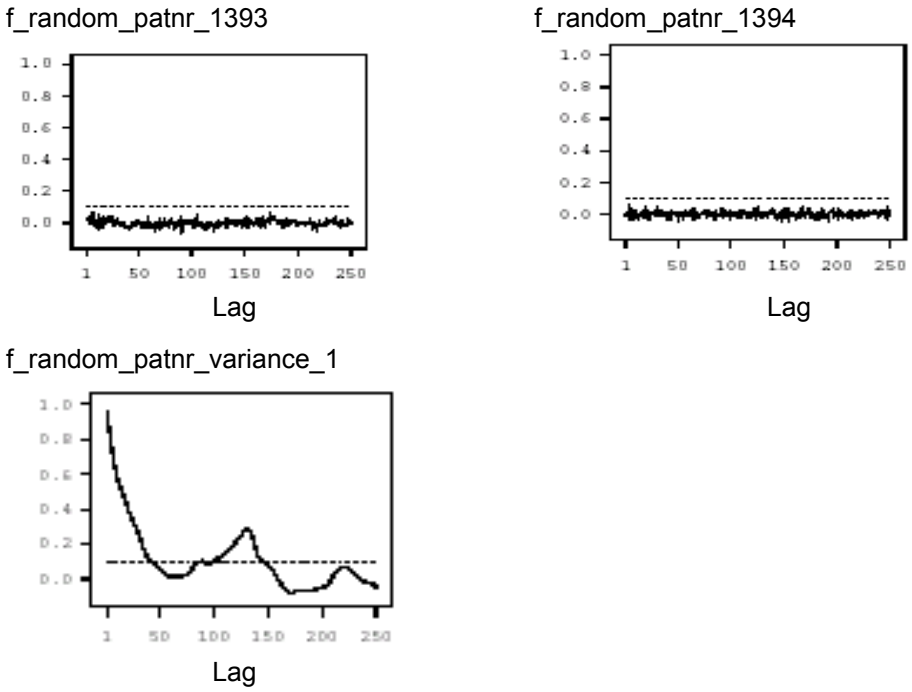
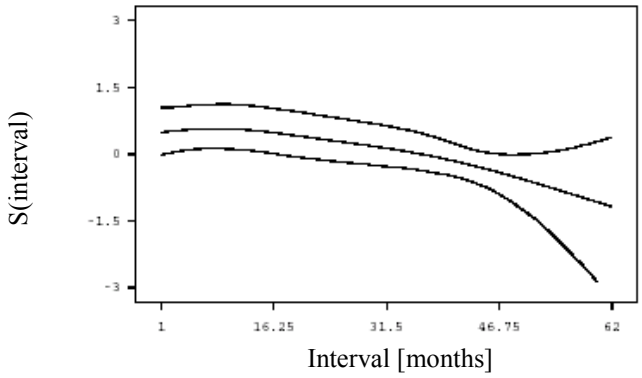



**Figure 9**. Time effect with random patient effects.




*Estimation with random study effects*

Including the covariate trial.id as random effect resulted in a more coarse way of clustering which on the one hand came from summarizing patient effects, e.g. through inclusion/exclusion criteria, and on the other hand from specific study effects such as through methodological specialties of (teams of) clinicians. In this case a piecewise exponential model with normally distributed random study effects and a time effect (*interval*) with cubic splines was used.

Estimation was again carried out using MCMC procedures in BayesX (Brezger et al. 2002). In order to produce proper sample paths the following settings were used: 110000 iterations in total and 40000 for the burn-in, and thinning steps of width 70. Under these conditions, both auto-correlations and sampling paths behaved well (Figs 10 and 11), and the same was true for *intercept* and *interval*.

**Figure 10**. Auto-correlations for random study effects (maximum lag 250).

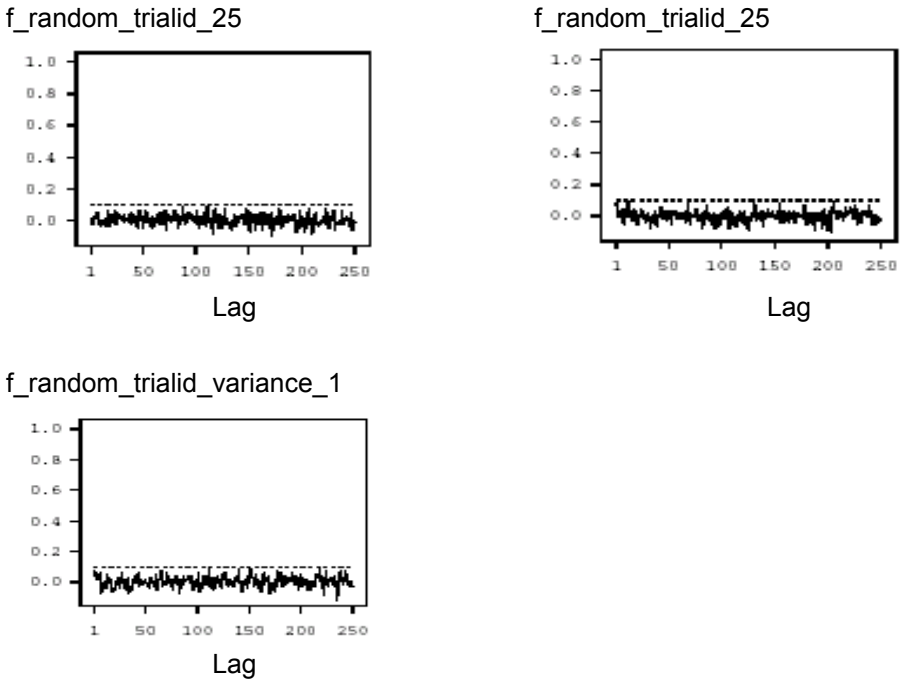
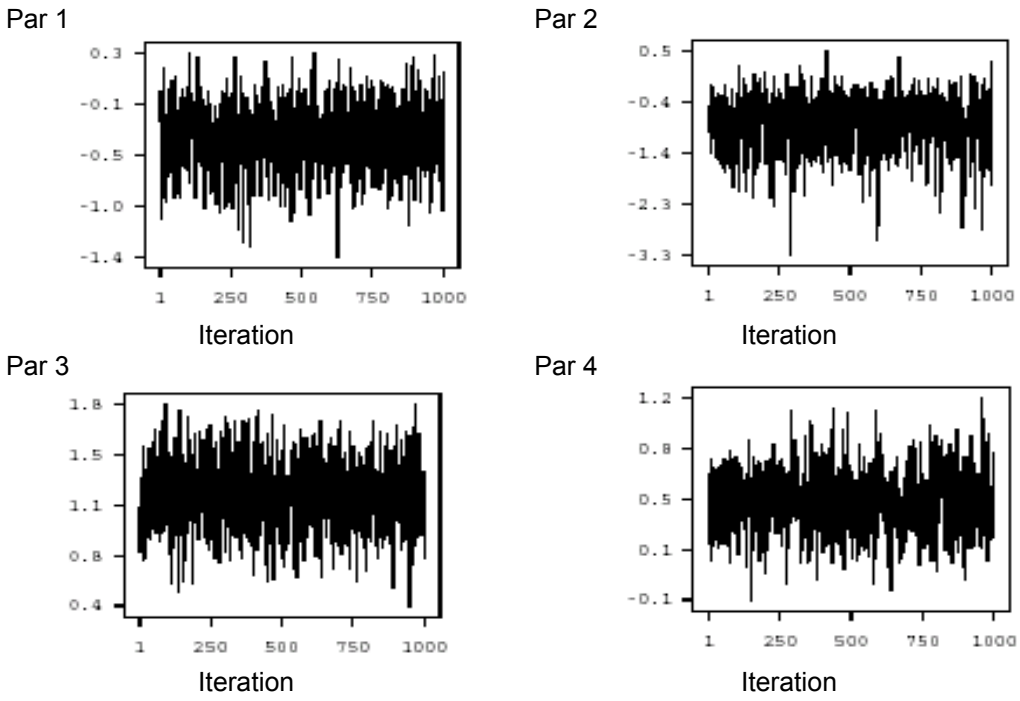



**Figure 11**. Sampling paths with random study effects (maximum 1000).



With random study effects the estimated influence of *interval* had about the same shape as before but was moved up, thereby generating significance (Fig. 12a). At least for about 16 months the confidence band ran above the zero line. This suggested to deal with further covariates and their parameterization. The next step was, therefore, to simultaneously model *interval* and edss0. For *interval* the form of the spline did not change considerably and for edss0 the form resembled that without study effects (Fig. 12b). The dispersion parameter, however, was higher when study effects were considered than without them (4.82 vs. 3.44), implying that according to the AIC significance was harder to achieve. The parameterization of edss0 influence fell shortly below the significance level. Other covariates also did not show an influence.

**Figure 12**. a) Time effect with random study effects and b) edss0 effect with random study effects.

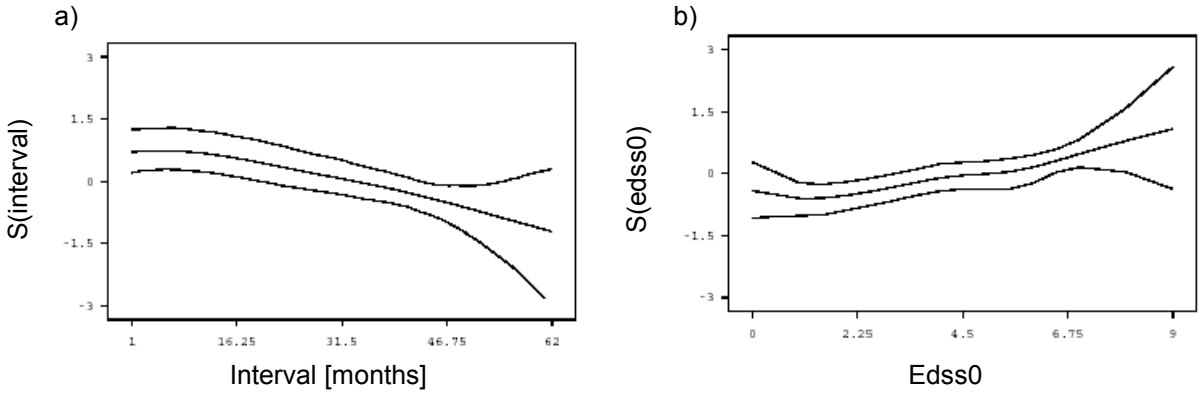



**Figure 13**. Parameterized influences of *interval*.

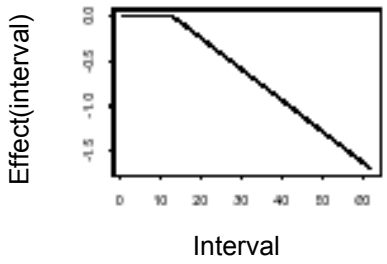



**Table 2**. AIC values for the models with study effects.

| **Model** | **AIC** |
|---|---|
| Null model | 3366.293 |
| *Interval* linear | 3354.091 |
| *Interval* parameterized | 3351.283 |
| + edss0 linear | too large |
| + edss0 parameterized | too large |

The investigations described above and summarized in Table 2 lead to the following model

**Model 3. Parameterized model with random study effects**

$$E(\lambda \mid \text{interval}) = \exp(-8.017) \cdot \exp(random(trial.id)) \cdot \exp(-0.038 \cdot (I(\text{interval} > 13) \cdot (\text{interval} - 13)))$$

with trial.id ~ N(0,0.513).

The main result here was that a time effect was detected when study effects were also included which had not been seen without them. This effect occurred only after month 13 (Fig. 13) and then was linearly decreasing. This decreasing trend of the hazard rate seemed surprising, because one would expect an increasing risk for the event to occur as the disease proceeds. This decrease, however, can be interpreted as a selection effect: unstable patients progress early. In the event modeling approach they, therefore, drop out of the observation. The remaining patients are more stable, leading to the observed decrease in the hazard rate.

## 8 Summary and conclusions

The goals of this study were to identify covariates with a significant impact on the time to sustained progression, to determine size and form of their effect and to estimate the survival function for given covariates. The main model was a piecewise exponential model using piecewise constant hazard rates (Fahrmeir and Tutz 2001). The length of the constant segments was set to 30 days. After an appropriate expansion of the data set the Poisson model with offset could be used for parameter estimation (Tutz 2000). As this is a special case of generalized linear and additive models standard statistics software could be utilized. Variables were selected using AIC. In order to improve the models the non-linear GAMs were transformed into simpler GLMs by parameterizing them piecewise linearly (on at most three segments). Besides fixed effects of covariates such as age at disease onset, gender and EDSS at entry into the trial, study effects were considered to be random. The models with fixed effects only were estimated using S-Plus (maximum likelihood method, Venables and Ripley 1997). For mixed effects models other software packages had to be used: for GLM the SAS procedure NLMIXED with adaptive Gauß quadrature for integration and Quasi Newton optimization for maximizing the log-likelihood (SAS 2000) and for GAM Bayesian models where fixed and random effects are modeled through prior distributions. The posterior distribution was estimated using Markov Chain Monte Carlo methods (MCMC) and the software BayesX (Brezger et al. 2002).

With respect to the data on placebo patients from 26 studies gathered in the SLCMSR database this approach proved to be suitable for explorative analyses and it allowed the direct estimation of the survival function for given covariates. Only with a pure exponential model and fixed effects, baseline EDSS and the number of relapses in the last 12 month before study entry had an effect on the hazard rate. For the piecewise exponential model with random study effects there was no effect of covariates on the hazard rate other than a weak time effect which implied in a slight decrease. This decrease, however, can be interpreted as a selection effect: unstable patients progress early. In the event modelling approach they, therefore, drop out of the observation and thus the hazard rate of the remaining collective decreases. The expected long-term increase could not be detected in the placebo data set because of the relatively short duration of clinical trials (longest time series about five years). This suggests that an analogous analysis should be conducted with data from patient registries (natural history data) which span much longer time intervals. The influence of the number of relapses before entry into the study should also be considered in further analyses because here the necessary information was missing for about a quarter of the patients.

A great difficulty in analyzing the MS placebo data was related to the scale used for measuring disability. Although EDSS is the gold standard since about 40 years it has a number of disadvantages such as a lack of

reliability, an ordinal scale, a bimodal distribution and an overemphasis on walking ability (Kappos et al. 1998). For future analyses it would, therefore, be better to use a more comprehensive scale, e.g. the MS Functional Composite (MSFC). In this context it would also make sense to include time dependent covariates, e.g. the time course of MSFC values or of scores for particular functional systems.

## Acknowledgements

We thank Martin Daumer, Andrea Hennerfeind, Alexander Jerak, Stefanie Kessner, Claudia Lamina, Stefan Lang, Michael Scholz, and Philip Young for discussions and comments and Ludwig Kappos for medical advice. The SLCMSR thanks all its data donors which are listed at http://www.slcmsr.org/en/partner.htm. BH is funded by the SFB 386 of the German Science Foundation and the SLCMSR by the Multiple Sclerosis International Federation (MSIF).

## References

Brezger, A, 2000. Bayesianische P-Splines. Diplomarbeit, Ludwig Maximilian University, Munich.

Brezger, A., Kneib, T., and Lang, S., 2002. BayesX, Version 0.9, Documentation, Munich.

Dennis, J.E., and Mei, H.H.W., 1979. Two New Unconstrained Optimization Algorithms which Use Function and Gradient Values. *Journal of Optimization Theory and Applications* 28, 453-482.

Draper, D., 1996 Discussion of the Paper by Lee and Nelder. *Journal of the Royal Statistical Society*, Series B 58, 662-663.

Fahrmeir, L., and Tutz, G., 2001. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer, New York.

Fahrmeir, L., Hamerle, A., and Tutz, G., 1996. *Multivariate Statistische Verfahren*. De Gruyter, Berlin.

Gay, D.M., 1983. Subroutines for Unconstrained Minimization, *ACM Transactions on Mathematical Software* 9, 503-524, Association for Computing Machinery, New York.

Hastie, T., andTibshirani, R., 1990. *Generalized additive models*. Chapman & Hall, London.

Kappos, L., Lechner-Scott, J., Lienert, C., Baumhackl, U., Hartung, H.-P., Mamoli, B., and Rieckmann, P., 1998. Expanded Disability Status Scale (EDSS), BIOGEN (ed.).

Kappos, L., Polman,C., Pozzilli, C., et al. , 1998. Placebo-controlled multicentre randomised trial of interferon beta-1b in treatment of secondary progressive multiple sclerosis. Lancet 352, 1491-1497.

Lang, S. and Brezger, A., 2003. Bayesian P-splines. J.C.G.F. (to appear).

Leary, S.M., Miller, D.H., Stevenson, V.L., Brex, P.A., Chard, D.T., and Thompson, A.J., 2003. Interferon beta-1a in primary progressive MS - An exploratory, randomized, controlled trial. Neurology 60, 44-51.

Liu C., and Blumhardt L.D., 2000. Disability outcome measures in therapeutic trials of relapsing-remitting multiple sclerosis: effects of heterogeneity of disease course in placebo cohorts. Journal of Neurology Neurosurgery and Psychiatry 68, 450-457.

Noseworthy, J., Kappos, L. and Daumer, M. (2003). Competing interests in multiple sclerosis research. Lancet 361, 350-351.

Pinheiro, J.C., and Bates, D.M., 1995. Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics*. 4 (1), 12-35.

SAS Institute Inc., 2000. *SAS OnlineDoc*. Version 8.

Tutz, G., 2000. *Die Analyse kategorialer Daten*. Oldenbourg, Munich.

Venables, W.N., and Ripley, B.D., 1997. *Modern Applied Statistics with S-PLUS*. 2nd ed., Springer, New York.