Bayesian Accelerated Failure Time Model for Correlated Censored Data with a Normal Mixture as an Error Distribution

Arnošt Komárek*

Biostatistical Centre, Katholieke Universiteit Leuven,

Kapucijnenvoer 35, B–3000, Leuven, Belgium

and

Emmanuel Lesaffre

September 9, 2004

Summary. A fully Bayesian approach for an accelerated failure time model is proposed. The model allows for structured correlated data by inclusion of a random effect part that might depend on a general vector of covariates as in linear mixed models. The error distribution is modelled as a normal mixture with an unknown number of components. Also the means and variances are not specified to accomodate most continuous distributions. A Markov chain Monte Carlo algorithm is described and the approach is illustrated on two survival applications: (1) data giving times between recurrent events for patients with chronic granulotomous disease; and (2) times to emergence of permanent teeth. In the practical examples, we illustrate how a predictive error distribution, and predictive survival or hazard curves for future observations can be obtained.

Key words: Clustered Data; Recurrent Events; Regression; Reversible Jump Markov

* *email:* arnost.komarek@med.kuleuven.ac.be

Chain Monte Carlo; Survival Data.

1. Introduction

Correlated survival times are encountered in many medical problems, e.g. when there are recurrent events on an individual or when the observations are clustered (multicenter studies, multivariate survival times). When the effect of covariates on the survival time is of interest, several approaches for right-censored correlated survival times have been proposed (see, e.g., Hougaard, 2000) which are extensions of the Cox's proportional hazard model (Cox, 1972). Here we present a Bayesian accelerated failure time (AFT) regression model allowing for possible correlation between the event times while making only moderate distributional assumptions on the error term. Moreover, the approach allows not only for right- or left-censored data but also for interval-censored data.

In an AFT model the covariates are assumed to speed up or slow down the expected time to failure, see e.g. Kalbfleisch and Prentice (2002, Section 2.3.3). An extension of the AFT model to incorporate correlated survival data could consist in including random effects in the regression expression as in a classical linear mixed model (Laird and Ware, 1982), i.e.

$$\log(T_{i,l}) = Y_{i,l} = \boldsymbol{\beta}^T \boldsymbol{x}_{i,l} + \boldsymbol{b}_i^T \boldsymbol{z}_{i,l} + \varepsilon_{i,l}, \qquad i = 1, \ldots, N, \quad l = 1, \ldots, n_i, \tag{1}$$

where $T_{i,l}$ is the event time of the lth observation of the ith cluster or the time of the lth recurrent event on the ith patient, $Y_{i,l}$ its logarithmic (or any other monotone) transformation, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is the unknown regression coefficient vector, $\boldsymbol{x}_{i,l}$ the covariate vector for fixed effects, $\boldsymbol{b}_i = (b_{i,1}, \ldots, b_{i,q})^T$ is the random effect vector causing the possible correlation for the components of $\boldsymbol{Y}_i = (Y_{i,1}, \ldots, Y_{i,n_i})^T$, $\boldsymbol{z}_{i,l}$ is the covariate vector for random effects and $\varepsilon_{i,l}$ are independent and identically distributed random variables. Along the lines of Gelman et al. (2004, Chapter 15) we use the terms 'fixed'

and 'random' effects throughout the paper even in a Bayesian context where all unknown parameters are treated as random quantities.

For recurrent events, usually $\boldsymbol{z}_{i,l} = 1$ for all i and l and $\boldsymbol{b}_i = b_{i,1}$ expresses an individual-specific deviation from an overall mean log-event time which is not explained by fixed effects covariates. For clustered data, the vector $\boldsymbol{z}_{i,l}$ may define further sub-clusters (as in the example of Section 4.2) allowing for closer dependence of observations within sub-clusters given by common values of appropriate components of the vector $\boldsymbol{b}_i$ while keeping the dependence also across the sub-clusters through the correlation between the components of $\boldsymbol{b}_i$.

Unlike for a classical linear mixed model where a normal distribution for the error term $\varepsilon_{i,l}$ is often a reasonable assumption there is no gold standard for the error distribution when dealing with survival data so that semi-parametric procedures are usually preferred. For uncorrelated survival data, a classical semi-parametric approach for fitting AFT model is given by the Buckley and James (1979) method. Another classical approach makes use of linear rank statistics (Louis, 1981; Tsiatis, 1990; Jin et al., 2003). Recently, Komárek, Lesaffre and Hilton (2004) suggested an AFT model with a flexible error distribution estimated by a penalized maximum likelihood method.

Early approaches to the semi-parametric AFT model used a GEE approach (Liang and Zeger, 1986) for analyzing

$$\log(T_{i,l}) = Y_{i,j} = \boldsymbol{\beta}^T \boldsymbol{x}_{i,l} + \varepsilon_{i,l}, \qquad i = 1, \ldots, N, \quad l = 1, \ldots, n_i, \tag{2}$$

with correlated $T_{i,1}, \ldots, T_{i,n_i}$, see e.g. Lin and Wei (1992) and Lee, Wei and Ying (1993). In the first step they estimate the regression coefficient $\boldsymbol{\beta}$ using a linear rank statistics approach and ignoring the correlation. In the second step, they correct the standard errors using a GEE approach. However, ignoring the dependence generally does not take

full advantage of the information in the data and is likely not to be efficient.

Methods that deal with the correlation among survival times in the estimation part appeared more recently in the literature (Hornsteiner and Hamerle, 1996; Pan and Kooperberg, 1999; Pan and Connett, 2001). They specify the AFT model as either (1) or (2) and the estimation procedure generally iterates between (a) estimating the marginal distribution of $Y_{i,l} - \boldsymbol{\beta}^T \boldsymbol{x}_{i,l}$ using either the Kaplan-Meier estimate or the bivariate log-spline density estimate of Kooperberg (1998); (b) single imputation (Buckley and James method) or multiple imputation (Rubin, 1987) of censored event times; (c) estimation of regression coefficients $\boldsymbol{\beta}$ using methods for uncensored data (GEE of Liang and Zeger, 1986 or maximum-likelihood). Since the same distribution of all $Y_{i,l} - \boldsymbol{\beta}^T \boldsymbol{x}_{i,l} = (\boldsymbol{b}_i^T \boldsymbol{z}_{i,l}) + \varepsilon_{i,l}$ is required, only a univariate random effect b_i with $z_{i,l} \equiv 1$ in the model specification (1) is allowed. A related approach was taken by Pan and Louis (2000) who use a Monte Carlo EM algorithm (Tanner, 1996) in step (c). Again only a univariate random effect with $z_{i,l} \equiv 1$ is considered however the authors note that their approach can be extended to accomodate more general $\boldsymbol{z}_{i,l}$.

With regard to Bayesian approaches to AFT modelling, semi-parametric approaches were suggested by Christensen and Johnson (1988), Kuo and Mallick (1997) and Walker and Mallick (1999), they used a Dirichlet process, a mixture of Dirichlet processes and a Pólya tree, respectively for the error random variables $\varepsilon_{i,l}$. Only Walker and Mallick (1999) relax the assumption of the independence of $\varepsilon_{i,l}, i = 1, \ldots, N, l = 1, \ldots, n_i$ and allow for clustered data. Recently, Kottas and Gelfand (2001), Hanson and Johnson (2002; 2004) contributed further to semi-parametric Bayesian AFT modelling. The former paper uses a mixture of unimodal parametric densities or step-functions with a Dirichlet process prior placed on the mixing distribution for the error distribution, the later two papers further improves the proposal of Walker and Mallick (1999) by assuming a mixture

of fully specified Pólya tree priors.

Clearly, there is a need for a method that allows the inclusion of a general covariate vector $\boldsymbol{z}_{i,l}$ to allow for more structured modelling of the correlation among the survival times. To this end we have developed a full Bayesian model. We believe that the Bayesian approach is particularly appealing here since it accomodates not only right- or left- but also interval censoring without significant problems.

In this paper we suggest to model the distribution of the error term $\varepsilon_{i,l}$ using a mixture of normals to express in a parsimonious and flexible way most continuous distributions. To this end we will use the reversible jump MCMC approach (Green, 1995; Richardson and Green, 1997) to estimate the number of normal components as well as the normal distributions themselves. However we will take a suitable parametric distribution for the random effect term $\boldsymbol{b}_i$. The fact that we put more emphasis on a correct specification of the distribution of the error term $\varepsilon_{i,l}$ than on a specification of the distribution of random effects $\boldsymbol{b}_i$ is driven by the following reasoning.

For an AFT model, the regression parameters $\boldsymbol{\beta}$ express the effect of covariates ($\boldsymbol{x}_{i,l}$) both conditionally (given $\boldsymbol{b}_i$) and marginally (after integrating $\boldsymbol{b}_i$ out). Both interpretations do not change when different distributional assumptions are made on $\boldsymbol{b}_i$. Further, with a correctly specified distribution of $\varepsilon_{i,l}$ the conditional model is always correctly specified. However, when the distribution of $\varepsilon_{i,l}$ is incorrect neither conditional nor marginal models are specified correctly. Further, Keiding et al. (1997) show that for univariate (single-spell) Weibull AFT model the regression parameters are robust against the misspecification of the frailty distribution. This finding, also for non-Weibull models is further supported by the empirical results of Lambert et al. (2004).

Programs in C++ have been written with an interface to the R language (R Development Core Team, 2004) as a contributed package `bayesSurv` and can be down-

loaded together with the scripts used to perform analyses presented in Section 4 from the Comprehensive R Archive Network (CRAN) on `http://www.R-project.org`. In the program, sampled values of model parameters are stored in files and can be subsequently checked for the convergence using R packages `coda` or `boa`.

The paper is further organized in the following way. Section 2 describes the Bayesian model in detail, the third section continues with a description of the inference based on the model outlined in Section 2. Section 4 is devoted to two real data illustrations. The paper is finalized by a discussion.

2. A Bayesian AFT model

Assume that the true log-event time of the (i, l)th observation $y_{i,l}$ satisfies $y_{i,l}^L \leq y_{i,l} \leq y_{i,l}^U$ whereby the observed log-event time is given by the pair $(y_{i,l}^L, y_{i,l}^U)$, $-\infty \leq y_{i,l}^L \leq y_{i,l}^U \leq \infty$. For an uncensored observation: $y_{i,l}^L = y_{i,l}^U$, for a right censored observation: $y_{i,l}^U = \infty$ and for a left censored observation: $y_{i,l}^L = -\infty$.

2.1 *Error Structure*

The density $f(e)$ of the error term $\varepsilon_{i,l}$ in model (1) is specified as

$$f(e) = \sum_{j=1}^{k} w_j \, \varphi(e \mid \mu_j, \, \sigma_j^2), \tag{3}$$

with $\varphi(\cdot \mid \mu_j, \, \sigma_j^2) \equiv$ density of $\mathrm{N}(\mu_j, \sigma_j^2)$. Note that the number of mixture components, k, is unknown as well as mixture weights $\boldsymbol{w} = (w_1, \ldots, w_k)^T$, means $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^T$, and variances $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_k^2)^T$. Further, we may assume that $\varepsilon_{i,l}$ $i = 1, \ldots, N$, $l = 1, \ldots, n_i$ come from a heterogeneous population consisting of groups $j = 1, 2, \ldots, k$ of sizes proportional to w_j. Let $r_{i,l}$ denote the (unknown) label of the group from which each random error variable $\varepsilon_{i,l}$ is drawn, i.e.

$$p(r_{i,l} = j \mid k, \boldsymbol{w}) = w_j, \qquad j = 1, \ldots, k. \tag{4}$$

Given the value of $r_{i,l}$, the random error variable $\varepsilon_{i,l}$ is drawn from $\mathrm{N}(\mu_{r_{i,l}}, \sigma^2_{r_{i,l}})$.

2.2 *The Bayesian Hierarchical Model*

AFT model (1) has an hierarchical structure. We assume a directed acyclic graph (DAG) structure for our model which is graphically represented in Figure 1 with the usual convention of graphical models that square boxes represent fixed or observed quantities and circles the unknowns. As is indicated in the DAG, the unknown parameters can be split into two parts connected only through the node of true log-event times. The regression part of the model has a hierarchical model structure (see e.g., Gelman et al., 2004, Chapter 5). Unlike the classical hierarchical model, the error part is somewhat more complicated and given by the structure outlined in Paragraph 2.1. In this section we will specify in detail the distributional aspects of the components in Figure 1.

`[Figure 1 about here.]`

First of all, the conditional distribution of each (unknown) log-event time, nodes that connect the regression and error parts of the DAG is

$$y_{i,l} \mid r_{i,l}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{x}_{i,l}, \boldsymbol{z}_{i,l} \;\sim\; \mathrm{N}(\mu_{r_{i,l}} + \boldsymbol{\beta}^T \boldsymbol{x}_{i,l} + \boldsymbol{b}_i^T \boldsymbol{z}_{i,l},\; \sigma^2_{r_{i,l}}) \tag{5}$$

independently for $i = 1, \ldots, N$, $l = 1, \ldots, n_i$.

2.2.1 Regression part Let $\mathbb{X}$ be an $\sum_{i=1}^N n_i \times p$ matrix with vectors $\boldsymbol{x}_{1,1}^T, \ldots, \boldsymbol{x}_{N,n_N}^T$ as rows. Similarly, let $\mathbb{Z}$ be an $\sum_{i=1}^N n_i \times q$ matrix with vectors $\boldsymbol{z}_{1,1}^T, \ldots, \boldsymbol{z}_{N,n_N}^T$ as rows. Further, we will assume that the matrix $(\mathbb{X}, \mathbb{Z})$ is of full column rank $(p+q)$. In other words, covariates included in $\boldsymbol{x}_{i,l}$ are not included in $\boldsymbol{z}_{i,l}$ and vice versa. This gives rise to hierarchical centering which generally results in a better behavior of the MCMC algorithm (see Gelfand et al., 1995). Finally, since a general density (3) does not have zero mean we do not allow a column of ones in the matrix $\mathbb{X}$.

The prior distribution for each regression coefficient β_j, is assumed to be $\mathrm{N}(\nu_{\beta,j}, \psi_{\beta,j})$, $j = 1, \ldots, p$ and the β_j are assumed to be a priori independent. The vectors $\boldsymbol{\nu}_\beta = (\nu_{\beta,1}, \ldots, \nu_{\beta,p})^T$ and $\boldsymbol{\psi}_\beta = (\psi_{\beta,1}, \ldots, \psi_{\beta,p})^T$ are fixed hyperparameters.

The (prior) distribution for the random effect vector $\boldsymbol{b}_i$ is assumed to be (multivariate) normal, i.e.

$$\boldsymbol{b}_i \,|\, \boldsymbol{\gamma}, \mathbb{D} \;\sim\; \mathrm{N}_q(\boldsymbol{\gamma}, \mathbb{D}), \quad \text{independently for } i = 1, \ldots, N, \tag{6}$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^T$. The prior distribution for each γ_j, is $\mathrm{N}(\nu_{\gamma,j}, \psi_{\gamma,j})$, independently for $j = 1, \ldots, q$. The vectors $\boldsymbol{\nu}_\gamma = (\nu_{\gamma,1}, \ldots, \nu_{\gamma,q})^T$ and $\boldsymbol{\psi}_\gamma = (\psi_{\gamma,1}, \ldots, \psi_{\gamma,q})^T$ are fixed. Special care is needed when the random intercept is included in the model (i.e. when $\mathbb{Z}$ contains a column of ones, let say its first column). Hierarchical centering cannot be applied in this case since the overall intercept is given by the mean of the mixture (3). For that reason, γ_1 is fixed to zero (or equivalently, $\nu_{\gamma,1} = 0,\ \psi_{\gamma,1} = 0$).

Prior distribution for the covariance matrix $\mathbb{D}$ of random effects is assumed to be an inverse-Wishart with τ 'degrees of freedom' ($\tau > q - 1$) and a scale matrix $\mathbb{S}$ (parameterized such that the mean is $(\tau - q - 1)^{-1}\mathbb{S}$). In a special case of a univariate random effect ($q = 1$), we use d instead of $\mathbb{D}$ and s instead of $\mathbb{S}$ in the notation. Further, in the situation of $q = 1$, we considered alternatively (see Section 4.1) also the use of a spread uniform prior for standard deviation of the random effect, i.e. a priori $\sqrt{d} \sim \mathrm{Uniform}(0, \sqrt{s})$ which is often considered to be a better choice (Gelman et al., 2004; Gelman, 2004). The node τ becomes redundant then.

2.2.2 Error part Prior distributions of the parameters for the error distribution (3) of the AFT model (1) are inspired by the work of Richardson and Green (1997) (with some change in notation). We give a brief summary.

For the number of mixture components, k, we experimented with a Poisson distribution with mean equal to a hyperparameter λ truncated at some prespecified (relatively large) value k_{max} and a uniform distribution on $\{1, \ldots, k_{max}\}$ (the node λ in the DAG in Figure 1 becomes redundant then).

The prior for mixture weights $\boldsymbol{w}$ is taken to be a symmetric k-dimensional Dirichlet with prior 'sample size' equal to a hyperparameter δ, i.e.

$$\boldsymbol{w} \,|k, \delta \;\sim\; \mathrm{D}(\delta, \delta, \ldots, \delta). \tag{7}$$

Further, the mixture means μ_j and variances σ_j^2 are a priori all drawn independently with normal and inverse-gamma priors

$$\mu_j \,|\, k, \xi, \kappa \;\sim\; \mathrm{N}(\xi, \kappa) \quad \text{and} \quad \sigma_j^2 \,|\, k, \zeta, \eta \;\sim\; \mathrm{IG}(\zeta, \eta). \tag{8}$$

As in Richardson and Green (1997) we let η have a gamma distribution $\mathrm{G}(g, h)$ with fixed hyperparameters g and h, see the following section for more details.

Since the error model is invariant to permutations of labels $j = 1, \ldots, k$, the joint prior distribution of a vector $\boldsymbol{\mu}$ is restricted to the set $\{\boldsymbol{\mu} : \mu_1 < \cdots < \mu_k\}$ for identifiability reasons (see Stephens, 2000 for other approaches to reach identifiability). The joint prior distribution of the mixture means and variances is thus $k!$ times the product of the individual normal and inverse-gamma densities, restricted to above mentioned set of increasing means.

2.2.3 Censoring To finalize the list of conditional distributions in the DAG, Figure 1 we have to specify $p(y_{i,l}^L, y_{i,l}^U \,|\, y_{i,l}, \text{censoring})$. Firstly, the censoring mechanism in this paper is assumed to be noninformative about the failure distribution. A box called 'censoring' in the DAG represents a realization of the random variable(s) causing the

censoring. Note that there is no need to specify a measurement model for the censoring mechanism since the inference relies on the posterior distribution of parameters given data and the data consist of the realized censoring variables as well.

After omitting subscripts i, l for clarity, the form of $p(y^L, y^U \,|\, y, \text{censoring})$ is then rather obvious for most censoring mechanisms. In the case of right censoring driven by a censoring random variable C, $p(y^L, y^U \,|\, y,\, c)$ is a Dirac density with $P\big[(y^L, y^U) = (y, y) \,\big|\, y, c\big] = I[y \leq c]$, $P\big[(y^L, y^U) = (c, \infty) \,\big|\, y, c\big] = I[y > c]$. For interval censoring resulting from a realization of random variables $C_1, \ldots, C_m$ representing the times when a failure status was checked (e.g. visits to a clinic or laboratory examination), $p(y^L, y^U \,|\, y,\, c_1, \ldots, c_m)$ is again a Dirac density with $P\big[(y^L, y^U) = (c_j, c_{j+1}) \,\big|\, y, c_1, \ldots, c_m\big] = I[c_j < y \leq c_{j+1}]$, $j = 0, \ldots, m$ with $c_0 = -\infty,\ c_{m+1} = \infty$.

2.3 *Weak Prior Information*

In this paper, we have opted for specifying weak prior information on the parameters of interest. When a priori information is available, our prior assumptions could be appropriately modified.

For the regression part of the model, we use non-informative, however proper distributions, that is, the prior variances of regression parameters $\boldsymbol{\beta}$ ($\boldsymbol{\psi}_\beta$) and $\boldsymbol{\gamma}$ ($\boldsymbol{\psi}_\gamma$) are chosen such that the posterior variance of the regression parameters is at least 100 times lower (which must be checked from the results). Prior hyperparameters for the covariance matrix $\mathbb{D}$ giving a weak prior information correspond to choices of $\tau = q - 1 + d$ and $\mathbb{S} = \text{diag}(d, \ldots, d)$ with d being a small positive number.

In the error part of the model, it is not possible to be fully non-informative, i.e. to use priors $p(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 \,|\, k) \propto 1 \times \prod_{j=1}^{k} \sigma_j^{-2}$ and to obtain proper posterior distributions (Diebolt and Robert, 1994; Roeder and Wasserman, 1997). Richardson and Green (1997) offer, in

the context of i.i.d. observations, for say $e_1, \ldots, e_n$, the following alternative: A rather flat prior $\mathrm{N}(\xi, \kappa)$ for μ_j is achieved by letting ξ equal to $\bar{e} = n^{-1} \sum_{j=1}^{n} e_j$ and setting κ equal to a multiple of R^2, where $R = \max(e_i) - \min(e_i)$. They further point out that it might be restrictive to suppose that knowledge of the range or variability of the data implies much about the size of each single σ_j^2 and therefore introduced an additional hierarchical level by allowing η to follow a gamma distribution with parameters g and h. They further recommend taking $\zeta > 1 > g$ to express the belief that the σ_j^2 are similar, without being informative about their absolute size and setting the parameter h to a small multiple of $1/R^2$. Here, the residuals $y_{i,l} - \boldsymbol{\beta}^T \boldsymbol{x}_{i,l} - \boldsymbol{b}_i^T \boldsymbol{z}_{i,l}$ play the role of the observations e_i. A rough estimate of their location and scale can be obtained through a maximum-likelihood fit of the AFT model, even without random effects. (the scale of residuals can only increase), with explicitly included intercept and scale parameters in the model. This can be done using standard software packages as R, Splus, SAS. The estimated intercept from this model can then be used instead of $\bar{e}$ and a multiple of the estimated scale parameter instead of R.

3. Bayesian inference

3.1 *Posterior distribution*

For our Bayesian AFT model, the joint posterior distribution, with omitted hyper-parameters for a clarity, is given by

$$p\Big(\{y_{i,l}\}_{i,l},\, \boldsymbol{w},\, \boldsymbol{\mu},\, \boldsymbol{\sigma}^2,\, \{r_{i,l}\}_{i,l},\, k,\, \eta,\, \boldsymbol{\beta},\, \boldsymbol{\gamma},\, \{\boldsymbol{b}_i\}_i,\, \mathbb{D} \,\Big|\, \{(y_{i,l}^L,\, y_{i,l}^U)\}_{i,l},\, \text{censoring},\, \{\boldsymbol{x}_{i,l}\}_{i,l},\, \{\boldsymbol{z}_{i,l}\}_{i,l}\Big) \propto$$

$$\propto \prod_{i=1}^{N}\prod_{l=1}^{n_i}\Big\{p(y_{i,l}^L,\, y_{i,l}^U \mid y_{i,l},\, \text{censoring}) \;\times\; p(y_{i,l} \mid \boldsymbol{\mu},\, \boldsymbol{\sigma}^2,\, r_{i,l},\, \boldsymbol{\beta},\, \boldsymbol{b}_i,\, \boldsymbol{x}_{i,l},\, \boldsymbol{z}_{i,l})\Big\} \times$$
$$\times \prod_{i=1}^{N}\prod_{l=1}^{n_i} p(r_{i,l} \mid k,\, \boldsymbol{w}) \;\times\; p(\boldsymbol{\mu} \mid k) \;\times\; p(\boldsymbol{\sigma}^2 \mid k,\, \eta) \;\times\; p(\eta) \;\times\; p(\boldsymbol{w} \mid k) \;\times\; p(k) \times$$
$$\times\; p(\boldsymbol{\beta}) \;\times \prod_{i=1}^{N} p(\boldsymbol{b}_i \mid \boldsymbol{\gamma},\, \mathbb{D}) \;\times\; p(\boldsymbol{\gamma}) \;\times\; p(\mathbb{D}), \tag{9}$$

where all conditional distributions follow directly from Section 2.

3.2 *Markov chain Monte Carlo*

A convenient way to get posterior quantities of interest is by means of MCMC methods, see, e.g. Besag et al. (1995). In our problem, the regression part of the model is updated using the Gibbs sampler (Geman and Geman, 1984). The same approach is used to update the true log-event times $y_{i,l}$ and the parameters from the error part of the model whose dimension does not depend on k $(\eta, r_{i,l}, i = 1, \ldots, N, l = 1, \ldots, n_i)$. The parameters $\boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2$ and k itself are updated by a reversible jump MCMC algorithm of Green (1995).

Details of the implementation of both the reversible jump MCMC algorithm for parameters of the varying dimension and the Gibbs steps for the remaining parameters of the error part of the model are given in Richardson and Green (1997). Their guidelines, now based on residuals $e_{i,l} = y_{i,l} - \boldsymbol{\beta}^T \boldsymbol{x}_{i,l} - \boldsymbol{b}_i^T \boldsymbol{z}_{i,l}$ can be immediately applied with some obvious changes in notation. For the actual implementation of the reversible jump MCMC algorithm we additionally employed the auxiliary variable (AV) method of Brooks et al. (2003, Section 9) for the dimension changing steps (split-combine and birth-death moves).

The full conditional distributions of the parameters from the regression part of the model and of true log-event times needed to implement the Gibbs sampler follow. The notation $|\ldots$ indicates that conditioning is done on all remaining parameters.

3.2.1 True log-event times $y_{i,l}$ The full conditional distribution of each $y_{i,l}$ is a truncated normal, i.e.

$$y_{i,l} \mid \cdots \sim \mathrm{N}(\mu_{r_{i,l}} + \boldsymbol{\beta}^T \boldsymbol{x}_{i,l} + \boldsymbol{b}_i^T \boldsymbol{z}_{i,l},\ \sigma^2_{r_{i,l}}) \text{ truncated on } (y^L_{i,l}, y^U_{i,l}]. \tag{10}$$

3.2.2 Fixed effects $\boldsymbol{\beta}$ Let $\boldsymbol{\beta}_{(S)}$ be an arbitrary subvector of vector $\boldsymbol{\beta}$, and $\boldsymbol{x}_{i,l(S)}$ the corresponding subvectors of covariate vectors $\boldsymbol{x}_{i,l}$, and further $\boldsymbol{x}_{i,l(-S)}$ their complementary subvectors. Similary, let further $\boldsymbol{\nu}_{\beta(S)}$ and $\boldsymbol{\psi}_{\beta(S)}$ be appropriate subvectors of hyperparameters $\boldsymbol{\nu}_\beta$ and $\boldsymbol{\psi}_\beta$, respectively. Finally, let $\Psi_{\beta(S)} = \mathrm{diag}(\boldsymbol{\psi}_{\beta(S)})$. Then

$$\boldsymbol{\beta}_{(S)} \mid \cdots \sim \mathrm{N}\Big(E[\boldsymbol{\beta}_{(S)} \mid \cdots], \mathrm{var}[\boldsymbol{\beta}_{(S)} \mid \cdots]\Big), \tag{11}$$

$$\text{with} \quad \mathrm{var}[\boldsymbol{\beta}_{(S)} \mid \cdots] = \Big(\Psi^{-1}_{\beta(S)} + \sum_{i=1}^{N}\sum_{l=1}^{n_i} \sigma^{-2}_{r_{i,l}} \boldsymbol{x}_{i,l(S)} \boldsymbol{x}^T_{i,l(S)}\Big)^{-1},$$

$$E[\boldsymbol{\beta}_{(S)} \mid \cdots] = \mathrm{var}[\boldsymbol{\beta}_{(S)} \mid \cdots] \times \Big\{\Psi^{-1}_{\beta(S)} \boldsymbol{\nu}_{\beta(S)} + \sum_{i=1}^{N}\sum_{l=1}^{n_i} \sigma^{-2}_{r_{i,l}} \boldsymbol{x}_{i,l(S)} e^{(F)}_{i,l(S)}\Big\},$$

where $e^{(F)}_{i,l(S)} = y_{i,l} - \mu_{r_{i,l}} - \boldsymbol{\beta}^T_{(-S)} \boldsymbol{x}_{i,l(-S)} - \boldsymbol{b}_i^T \boldsymbol{z}_{i,l}$.

3.2.3 Means of random effects $\boldsymbol{\gamma}$ Let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}^T_{(S)}, \boldsymbol{\gamma}^T_{(-S)})^T$. More general case with unsorted components of the $\boldsymbol{\gamma}$ vector is obtained obviously with some costs on ease of notation. Analogically as above, let $\boldsymbol{b}_{i(S)}$, $\boldsymbol{b}_{i(-S)}$, $\boldsymbol{\nu}_{\gamma(S)}$, $\boldsymbol{\psi}_{\gamma(S)}$ the corresponding subvectors or complementary subvectors of indicated quantities and $\Psi_{\gamma(S)} = \mathrm{diag}(\boldsymbol{\psi}_{\gamma(S)})$. Further, let the inversion of the matrix $\mathbb{D}$ be decomposed in the following way.

$$\mathbb{D}^{-1} = \begin{pmatrix} \mathbb{V}_{(S)} & \mathbb{V}_{(S,-S)} \\ \mathbb{V}^T_{(S,-S)} & \mathbb{V}_{(-S)} \end{pmatrix}. \tag{12}$$

Then

$$\boldsymbol{\gamma}_{(S)} \mid \cdots \sim \mathrm{N}\Big(E[\boldsymbol{\gamma}_{(S)} \mid \cdots], \mathrm{var}[\boldsymbol{\gamma}_{(S)} \mid \cdots]\Big), \tag{13}$$

with

$$\text{var}[\boldsymbol{\gamma}_{(S)} \mid \cdots] = \Big(\Psi^{-1}_{\gamma(S)} + N\, \mathbb{V}_{(S)}\Big)^{-1},$$

$$E[\boldsymbol{\gamma}_{(S)} \mid \cdots] = \text{var}[\boldsymbol{\gamma}_{(S)} \mid \cdots] \times \Big\{\Psi^{-1}_{\gamma(S)}\boldsymbol{\nu}_{\gamma(S)} + \mathbb{V}_{(S)} \sum_{i=1}^{N} \boldsymbol{b}_{i(S)} + \mathbb{V}_{(S,-S)} \sum_{i=1}^{N} \big(\boldsymbol{b}_{i(-S)} - \boldsymbol{\gamma}_{(-S)}\big)\Big\}.$$

3.2.4 Random effects $\boldsymbol{b}_i$ For the random effects vectors $\boldsymbol{b}_i$:

$$\boldsymbol{b}_i \mid \cdots \sim \text{N}\Big(E[\boldsymbol{b}_i \mid \cdots], \text{var}[\boldsymbol{b}_i \mid \cdots]\Big), \qquad i = 1, \ldots, N, \tag{14}$$

$$\text{with} \quad \text{var}[\boldsymbol{b}_i \mid \cdots] = \Big(\mathbb{D}^{-1} + \sum_{i=1}^{N}\sum_{l=1}^{n_i} \sigma^{-2}_{r_{i,l}} \boldsymbol{z}_{i,l}\boldsymbol{z}^T_{i,l}\Big)^{-1},$$

$$E[\boldsymbol{b}_i \mid \cdots] = \text{var}[\boldsymbol{b}_i \mid \cdots] \times \Big\{\mathbb{D}^{-1}\boldsymbol{\gamma} + \sum_{i=1}^{N}\sum_{l=1}^{n_i} \sigma^{-2}_{r_{i,l}} \boldsymbol{z}_{i,l}(y_{i,l} - \mu_{r_{i,l}} - \boldsymbol{\beta}^T\boldsymbol{x}_{i,l})\Big\}.$$

3.2.5 Covariance matrix of random effects $\mathbb{D}$ Finally, $\mathbb{D} \mid \cdots$ is an inverse-Wishart distribution with degrees of freedom equal to $\tau + N$ and a scale matrix equal to $\mathbb{S} + \sum_{i=1}^{N}(\boldsymbol{b}_i - \boldsymbol{\gamma})(\boldsymbol{b}_i - \boldsymbol{\gamma})^T$.

4. Real data examples

To demonstrate the performance of our approach on a real data set and show possible outputs from such an analysis we give two examples.

4.1 *CGD data: recurrent events analysis*

The first example uses the data set from a multicenter placebo-controlled randomized trial of gamma inferon in patients with chronic granulotomous disease (CGD). The data set can be found in Appendix D.2 of Fleming and Harrington (1991). There were 128 patients randomized to either gamma inferon ($n = 63$) or placebo ($n = 65$). For each patient the times from study entry to initial and any recurrent serious infections are available. There is a minimum of one and a maximum of eight (recurrent) infection

times per patient, with a total of 203 records.

The problem of recurrent events in this data set was discussed by several authors. Among others, Therneau and Hamilton (1997) used the CGD data to illustrate several approaches for recurrent event analysis based on the Cox's proportional hazards (PH) model (Cox, 1972). Vaida and Xu (2000) used this dataset to illustrate the PH model with random effects. They specify the hazard function for the $(i,l)th$ event as $\hbar_{i,l}(t) = \hbar_0(t)\ \exp(\boldsymbol{\beta}^T \boldsymbol{x}_{i,l} + \boldsymbol{b}_i \boldsymbol{z}_{i,l})$ and use a normal distribution for $\boldsymbol{b}_i$.

In this section, we present AFT model (1) with response the time from entry or previous infection to the next infection in days. Each patient represents a cluster, i.e. $i = 1, \ldots, 203$, $l = 1, \ldots, n_i,\ n_i \leq 8$. Dependencies between the times of recurrent events of one patient are introduced by a univariate random effect b_i with $z_{i,l} = 1$ for all i and l. As fixed effects covariates, we used the same covariates as Vaida and Xu (2000), see Table 1 for their list.

[Table 1 about here.]

The initial maximum-likelihood AFT model with a normal error distribution and without random effects gave an estimate of the intercept equal to 3.66 and a scale equal to 1.69. Along the suggestions made in Section 2.3 we used the following values of hyperparameters: $\xi = 3.66$, $\kappa = 25 \approx (3 \cdot 1.69)^2$, $\zeta = 2$, $g = 0.2$, $h = 0.1$, $\delta = 1$. For the number of mixture components, k, a truncated Poisson prior with $\lambda = 5$ reflecting our prior belief that the error distribution is skewed and $k_{max} = 30$ was used. Prior means of all regression parameters were equal to 0 and their prior variances to 1000.

For the variance d of the random effect we tried either an inverse-gamma$(0.001, 0.001)$ prior ($\tau = 0.002,\ s = 0.002$ in the terms of the inverse-Wishart distribution used in the DAG, Figure 1) or a uniform Unif$(0, \sqrt{s})$ prior on $\sqrt{d}$ (Spiegelhalter et al., 2004,

Section 5.7) with s equal to 100^2, 50^2 and 10^2. Different priors for this parameter had only negligible effect on the posterior distributions of all remaining parameters. However, the posterior distribution of d was strongly driven by the inverse-gamma prior (showing two modes, one of them located close to zero). This was not the case when the uniform prior was used. Additionally, all uniform priors led to essentially identical posterior distributions. All results presented below are then based on Unif(0, 100) prior on $\sqrt{d}$.

Posterior summary statistics of the model can be found in Table 1. It is seen that the treatment significantly increases the time to the infection. Further, the posterior mean of $\exp\{\beta(trtmt)\}$ is equal to 4.01 with 95% CI = (1.60, 9.18) which means that on average, the treatment increases the time to the next event 4.01 times.

Further, the first panel of Figure 2 shows posterior means and 95% posterior credibility intervals of random effects b_i for all patients, sorted according to number of infections they underwent. It is clearly seen that the random effects of patients with higher numbers of total infections on average decrease (consequently the same is true for the time to the next event).

[Figure 2 about here.]

4.1.1 Predictive error densities Averaging the error density (3) across the MCMC run, conditionally on fixed values of k, gives a Bayesian predictive error density estimate of the mixture with k components, i.e. an estimate of $E\{f(\cdot\ |k, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2\ )\big|\ k, \text{data}\}$. Averaging further across values of k gives an estimate of $E\{f(\cdot\ |k, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2\ )\big|\ \text{data}\}$, the overall Bayesian predictive density estimate of the error distribution. In our sample, the number of mixture components k ranged from 1 to 18 while mixtures with $k \in \{4, 5, 6, 7\}$ occupied each more than 10% of the sample, with the highest frequency for $k = 6$ (13.0%).

Mixtures with $k \geq 11$ took each less than 3% of the sample. Apparently, the model did not suffer from the technical restriction given by $k_{max} = 30$. Predictive error density estimates are shown in the second panel of Figure 2. Note that only $k \in \{1, 2\}$ (14.8% of the sample) gives an appreciably different estimate from the unconditional estimate and conditional estimates for $3 \leq k \leq 10$ (79.3% of the sample).

4.1.2 Predictive survivor curves Further, we present estimates of predictive survivor curves for a specific value of covariates, say $\boldsymbol{x}_{new}$ and $\boldsymbol{z}_{new}$. Denoting all unknown quantities in the model by $\boldsymbol{\theta}$ and omitting $\boldsymbol{x}_{new}$ and $\boldsymbol{z}_{new}$ in the notation, the predictive survivor function is given by

$$S(t \mid \text{data}) = \int S(t \mid \boldsymbol{\theta}, \text{data})\, p(\boldsymbol{\theta} \mid \text{data})\, d\boldsymbol{\theta}$$

for any $t > 0$. Further

$$S(t \mid \boldsymbol{\theta}, \text{data}) = S(t \mid \boldsymbol{\theta}) = \sum_{j=1}^{k} w_j \Big[1 - \Phi\big\{\log(t) - \boldsymbol{\beta}^T \boldsymbol{x}_{new} - \boldsymbol{b}^T \boldsymbol{z}_{new} \mid \mu_j, \sigma_j^2\big\}\Big],$$

where $\Phi(\cdot \mid \mu_j, \sigma_j^2)$ is a cumulative distribution function of $\mathrm{N}(\mu_j, \sigma_j^2)$. The MCMC estimate of the predictive survivor function is then given by

$$\hat{S}(t \mid \text{data}) = M^{-1} \sum_{m=1}^{M} \sum_{j=1}^{k^{(m)}} w_j^{(m)} \Big[1 - \Phi\big\{\log(t) - \boldsymbol{\beta}^{(m)T} \boldsymbol{x}_{new} - \boldsymbol{b}^{(m)T} \boldsymbol{z}_{new} \mid \mu_j^{(m)}, \sigma_j^{(m)2}\big\}\Big],$$

where M denotes number of MCMC iterations. All quantities are available, except $\boldsymbol{b}^{(m)}$. This must be additionally sampled from $\mathrm{N}_q(\boldsymbol{\gamma}^{(m)}, d^{(m)})$. Predictive survivor curves for males and females taking treatment or placebo while controlling for remaining covariates are shown in the left part of Figure 3.

[Figure 3 about here.]

4.1.3 Predictive hazard functions Also predictive hazard functions can be computed. For any $t > 0$

$$\hbar(t \mid \boldsymbol{\theta}, \text{data}) = \hbar(t \mid \boldsymbol{\theta}) = \frac{p(t \mid \boldsymbol{\theta})}{S(t \mid \boldsymbol{\theta})},$$

where $p(t \mid \boldsymbol{\theta}) = t^{-1} \sum_{j=1}^{k} w_j \varphi\big\{\log(t) - \boldsymbol{\beta}^T \boldsymbol{x}_{new} - \boldsymbol{b}^T \boldsymbol{z}_{new} \mid \mu_j, \sigma_j^2\big\}$. The MCMC estimate of the predictive hazard function is then given by

$$\hat{\hbar}(t \mid \text{data}) = M^{-1} \sum_{m=1}^{M} \frac{t^{-1} \sum_{j=1}^{k^{(m)}} w_j^{(m)} \varphi\big\{\log(t) - \boldsymbol{\beta}^{(m)T} \boldsymbol{x}_{new} - \boldsymbol{b}^{(m)T} \boldsymbol{z}_{new} \mid \mu_j^{(m)}, \sigma_j^{(m)2}\big\}}{\sum_{j=1}^{k^{(m)}} w_j^{(m)} \Big[1 - \Phi\big\{\log(t) - \boldsymbol{\beta}^{(m)T} \boldsymbol{x}_{new} - \boldsymbol{b}^{(m)T} \boldsymbol{z}_{new} \mid \mu_j^{(m)}, \sigma_j^{(m)2}\big\}\Big]}.$$

Predictive hazard curves for same combination of covariates as before are shown in the right part of Figure 3.

4.2 *Signal Tandmobiel®: interval-censored clustered data*

In the second example, we show an analysis of a subset of the Signal Tandmobiel® dataset (Vanobbergen et al., 2000) involving clustered interval-censored observations. This dental longitudinal study performed in Flanders in 1996–2001 collected oral health data at tooth and tooth-surface level from schoolchildren born in 1989. The children were examined annualy by one of 16 dentists. Annual examinations give then rise to interval censoring. In this paper, we analysed a random sample of 500 boys and 500 girls.

Lesaffre, Komárek and Declerck (2004) analysed the effect of gender and caries on the primary predecessor (described by a binarised dmf score) on the emergence time of the permanent premolars given in years (teeth 14, 15, 24, 25, 34, 35, 44, 45 in European dental notation). For each tooth separately they used the penalized AFT model of Komárek et al. (2004). With the current Bayesian approach, all eight teeth can be analysed jointly while accounting for possible correlation among teeth of a single child.

The cluster is constituted by a child now ($i = 1, \ldots, 1\,000$, $l = 1, \ldots, 8$). Based on a preliminary modelling we will assume so called *horizontal symmetry*, i.e. the same

emergence distribution is assumed for left and right tooth at the same position of the jaw (e.g. for teeth 14 and 24). For a better fit, we shifted the time origin of the AFT model to 5 years of age, i.e. by replacing $T_{i,l}$ by $T_{i,l} - 5$ in the model (1).

The random effect vector $\boldsymbol{b}_i = (b_{i,1}, \ldots, b_{i,4})'$ with $\boldsymbol{z}_{i,l} = (1, man4_{i,l}, max5_{i,l}, man5_{i,l})'$ where $man4_{i,l}, max5_{i,l}, man5_{i,l}$, respectively are dummies for the mandibular first premolars (teeth 34, 44), maxillary second premolars (teeth 15, 25) and mandibular second premolars (teeth 35, 45), respectively is assumed in the model (1). With such model specification, apart of the random variation given by the error term $\varepsilon_{i,l}$, the terms $d_{i,max4} = b_{i,1}$, $d_{i,man4} = b_{i,1} + b_{i,2}$, $d_{i,max5} = b_{i,1} + b_{i,3}$, $d_{i,man5} = b_{i,1} + b_{i,4}$ determine how the log-emergence time of a pair of horizontally symmetric teeth of a single child differ from the population average. Observe that our model allows an unstructured covariance matrix for the emergence times, apart of course from the assumed horizontal symmetry. As the fixed effects we used $gender \equiv girl$, dmf and all two-way interaction terms between $girl$, dmf and dummies for the pairs of horizontal symmetric teeth.

The same guidelines as in the case of CGD data were used to specify prior hyperparameters leading to $\xi = 1.8$, $\kappa = 0.75^2$, $\zeta = 2$, $g = 0.2$, $h = 0.1$, $\delta = 1$, $\lambda = 5$, $k_{max} = 30$. For the covariance matrix $\mathbb{D}$ of random effects we used an inverse Wishart prior with $\tau = 4$ which is a minimal possible value for prior degrees of freedom. Though, due to the fact that 1 000 clusters are involved in the data set even a higher value could be used with a negligible impact on results. Prior scale matrix $\mathbb{S}$ was equal to diag(0.002) (corresponding to inverse-gamma$(\tau, 0.001)$ in the univariate case). All β and γ parameters were assigned a spread $\mathrm{N}(0, 100)$ prior.

In this analysis, the main interest lies in the effect of dmf on emergence. This can be evaluated from Table 2 that shows posterior summary statistics for the effect of $dmf > 0$ (appropriate linear combinations of β parameters) for the two genders

and the four pairs of horizontally symmetric teeth. It is seen that caries on the primary predecessor accelerates significantly the emergence of the permanent successor in the case of maxillary teeth. For the mandibular teeth, a slight effect is observed only for the first premolar on boys. Further, Figure 4 shows predictive cumulative distribution functions which are preferred in dentistry to the survival functions in the case of emergence and are known as *emergence curves*. Clearly, besides the effect of dmf the emergence process for girls is ahead of boys.

[Table 2 about here.]

[Figure 4 about here.]

Finally, Table 3 shows posterior summary statistics for variances and correlations of above defined tooth-specific linear combinations $d_{i,max4}, d_{i,man4}, d_{i,max5}, d_{i,man5}$ of random effects $b_{i,1}, \ldots, b_{i,4}$. It shows how the child effect is important and how the different teeth in one mouth are strongly correlated. The posterior means of all variance parameters in Table 3 are all about 0.2 which is much higher than the posterior mean of the variance of the error distribution which was equal to 0.01. Posterior means of all correlation parameters lie between 0.79 and 0.91.

[Table 3 about here.]

5. Discussion

We have proposed a Bayesian accelerated failure time model whose error distribution is modelled in a flexible way as a finite normal mixture. An advantage of the full Bayesian approach is the fact that a general random effect vector can be easily included in the model. Further, interval-censored data do not convey any complexity to both algebra and computation and finally, the MCMC sampling-based implementation of the model

offers straightforward ways to obtain credibility intervals of model parameters as well as predictive survivor or hazard curves. The advantage of our approach, compared to semi-parametric (Bayesian) AFT models is the availability of the estimate of the error distribution which can serve as a basis for fully parametric models that, if appropriately used, can result in higher efficiency of the inference.

Acknowledgements

The first author was financially supported by Research Grants OE/03/29 and DB/04/031, Katholieke Universiteit Leuven. The authors further acknowledge support from the Interuniversity Attraction Poles Program P5/24 – Belgian State – Federal Office for Scientific, Technical and Cultural Affairs.

Data collection for the example in Section 4.2 was supported by Unilever, Belgium. The Signal Tandmobiel® project comprises the following partners: D. Declerck (Dental School, Catholic University Leuven), L. Martens (Dental School, University Ghent), J. Vanobbergen (Oral Health Promotion and Prevention, Flemish Dental Association), P. Bottenberg (Dental School, University Brussels), E. Lesaffre (Biostatistical Centre, Catholic University Leuven), K. Hoppenbrouwers (Youth Health Department, Catholic University Leuven; Flemish Association for Youth Health Care).

References

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems (with Discussion). *Statistical Science* **10**, 3–66.

Brooks, S. P., Giudici, P. and Roberts, G. O. (2003). Efficient construction of reversible

jump Markov chain Monte Carlo proposal distributions (with Discussion). *Journal of the Royal Statistical Society, Series B* **65**, 3–55.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429–436.

Christensen, R. and Johnson, W. (1988). Modelling accelerated failure time with a Dirichlet process. *Biometrika* **75**, 693–704.

Cox, D. R. (1972). Regression models and life-tables (with Discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–220.

Diebolt, J. and Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Series B* **56**, 363–375.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. John Wiley & Sons, New York.

Gelfand, A. E., Sahu, S. K. and Carlin, B. P. (1995). Efficient parametrisations for normal linear mixed models. *Biometrika* **82**, 479–499.

Gelman, A. (2004). Prior distributions for variance parameters in hierarchical models. Technical report, Columbia University.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian Data Analysis, Second Ed.* Chapman & Hall/CRC, Boca Raton.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayes restoration of image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Green, P. J. (1995). Reversible jump markov chain computation and bayesian model determination. *Biometrika* **82**, 711–732.

Hanson, T. and Johnson, W. O. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association* **97**, 1020–1033.

Hanson, T. and Johnson, W. O. (2004). A Bayesian semiparametric AFT model for interval-censored data. *Journal of Computational and Graphical Statistics* **13**, 341–361.

Hornsteiner, U. and Hamerle, A. (1996). A combined GEE/Buckley-James method for estimating an accelerated failure time model of multivariate failure times. Technical Report Discussion Paper 47, Ludwig-Maximilians-Universität München.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York.

Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–353.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data, Second Ed.*. John Wiley & Sons, Chichester.

Keiding, N., Andersen, P. K. and Klein, J. P. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine* **16**, 215–225.

Komárek, A., Lesaffre, E. and Hilton, J. F. (2004). Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *Being revised* .

Kooperberg, C. (1998). Bivariate density estimation with an application to survival analysis. *Journal of Computational and Graphical Statistics* **7**, 322–341.

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association* **96**, 1458–1468.

Kuo, L. and Mallick, B. (1997). Bayesian semiparametric inference for the accelerated failure time model. *The Canadian Journal of Statistics* **25**, 457–472.

Lambert, P., Collett, D., Kimber, A. and Johnson, R. (2004). Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in Medicine* **00**, in press.

Lee, E. W., Wei, L. J. and Ying, Z. (1993). Linear regression analysis for highly stratified failure time data. *Journal of the American Statistical Association* **88**, 557–565.

Lesaffre, E., Komárek, A. and Declerck, D. (2004). An overview of methods for interval-censored data with an emphasis on applications in dentistry. *Submitted* .

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Lin, J. S. and Wei, L. J. (1992). Linear regression analysis for multivariate failure time observations. *Journal of the American Statistical Association* **87**, 1091–1097.

Louis, T. A. (1981). Nonparametric analysis of an accelerated failure time model. *Biometrika* **68**, 381–390.

Pan, W. and Connett, J. E. (2001). A multiple imputation approach to linear regression with clustered censored data. *Lifetime Data Analysis* **7**, 111–123.

Pan, W. and Kooperberg, C. (1999). Linear regression for bivariate censored data via multiple imputation. *Statistics in Medicine* **18**, 3111–3121.

Pan, W. and Louis, T. A. (2000). A linear mixed-effects model for multivariate censored data. *Biometrics* **56**, 160–166.

R Development Core Team (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with unknown number of components (with Discussion). *Journal of the Royal Statistical Society, Series B* **59**, 731–792.

Roeder, K. and Wasserman (1997). Practical bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association* **92**, 894–902.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York.

Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons, Chichester.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society, Series B* **62**, 795–809.

Tanner, M. A. (1996). *Tools for Statistical Inference*. Springer, New York.

Therneau, T. M. and Hamilton, S. A. (1997). rhDNase as an example of recurrent event analysis. *Statistics in Medicine* **16**, 2029–2047.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics* **18**, 354–372.

Vaida, F. and Xu, R. (2000). Proportional hazards model with random effects. *Statistics in Medicine* **19**, 3309–3324.

Vanobbergen, J., Martens, L., Lesaffre, E. and Declerck, D. (2000). The Signal-Tandmobiel® project – a longitudinal intervention health promotion study in Flanders (Belgium): baseline and first year results. *European Journal of Paediatric Dentistry* **2**, 87–96.

Walker, S. and Mallick, B. K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics* **55**, 477–483.

Table 1

CGD Data. Posterior means, 95% equal-tail credibility intervals and Bayesian p-values for regression parameters $\boldsymbol{\beta}$: trtmt = treatment (yes), inher = pattern of inheritance (autosomal recessive), age = age in years, cortico = use of corticosteroids (yes), prophy = use of prophylactic antibiotics (yes), gender = female, hosp1 = hosp. category US – other, hosp2 = hosp. category Europe – Amsterdam, hosp3 = hosp. category Europe – other. Posterior summary statistics for intercept = mean of the error distribution, scale = standard deviation of the error distribution and standard deviation of the random effect.

trtmt	*inher*	*age*	*cortic*
1.303	−0.885	0.047	−2.533
(0.496, 2.214)	(−1.812, 0.035)	(0.005, 0.093)	(−5.311, −0.106)
$p = 0.001$	$p = 0.059$	$p = 0.027$	$p = 0.04$
prophy	*gender*	*hosp*1	*hosp*2
1.111	1.369	0.466	1.589
(0.069, 2.265)	(0.03, 2.821)	(−0.464, 1.473)	(0.143, 3.265)
$p = 0.036$	$p = 0.045$	$p = 0.333$	$p = 0.031$
*hosp*3	intercept	scale	std. dev. of $b_{i,1}$
1.213	3.852	1.871	0.826
(−0.071, 2.625)	(2.213, 5.465)	(1.259, 3.321)	(0.197, 1.473)
$p = 0.063$			

Table 2

Signal Tandmobiel® data. Posterior means, 95% equal-tail credibility intervals and Bayesian p-values for the effect of $dmf > 0$ for the two genders and different teeth.

maxilla 4		*maxilla* 5	
girl	*boy*	*girl*	*boy*
-0.0352	-0.0457	-0.0213	-0.0318
$(-0.0522,\ -0.0185)$	$(-0.0631,\ -0.0284)$	$(-0.0390,\ -0.0035)$	$(-0.0500,\ -0.0135)$
$p = 0$	$p = 0$	$p = 0.019$	$p = 0.001$
mandible 4		*mandible* 5	
girl	*boy*	*girl*	*boy*
-0.0098	-0.0203	0.0014	-0.0090
$(-0.0267,\ 0.0070)$	$(-0.0378,\ -0.0032)$	$(-0.0162,\ 0.0193)$	$(-0.0283,\ 0.0098)$
$p = 0.255$	$p = 0.021$	$p = 0.870$	$p = 0.353$

Table 3

Signal Tandmobiel® data. Posterior means, 95% equal-tail credibility intervals for variances and correlations between tooth-specific linear combinations of random effects.

$\mathrm{var}(d_{max4})$	$\mathrm{var}(d_{man4})$	$\mathrm{var}(d_{max5})$	$\mathrm{var}(d_{man5})$
0.204	0.198	0.205	0.202
(0.192, 0.218)	(0.186, 0.211)	(0.190, 0.221)	(0.187, 0.218)
$\mathrm{cor}(d_{max4}, d_{man4})$	$\mathrm{cor}(d_{max4}, d_{max5})$	$\mathrm{cor}(d_{max4}, d_{man5})$	$\mathrm{cor}(d_{man4}, d_{max5})$
0.886	0.914	0.841	0.792
(0.856, 0.914)	(0.887, 0.938)	(0.804, 0.874)	(0.749, 0.832)
$\mathrm{cor}(d_{man4}, d_{man5})$	$\mathrm{cor}(d_{max5}, d_{man5})$		
0.895	0.847		
(0.864, 0.923)	(0.810, 0.880)		

Figure 1. DAG for the Bayesian AFT model.

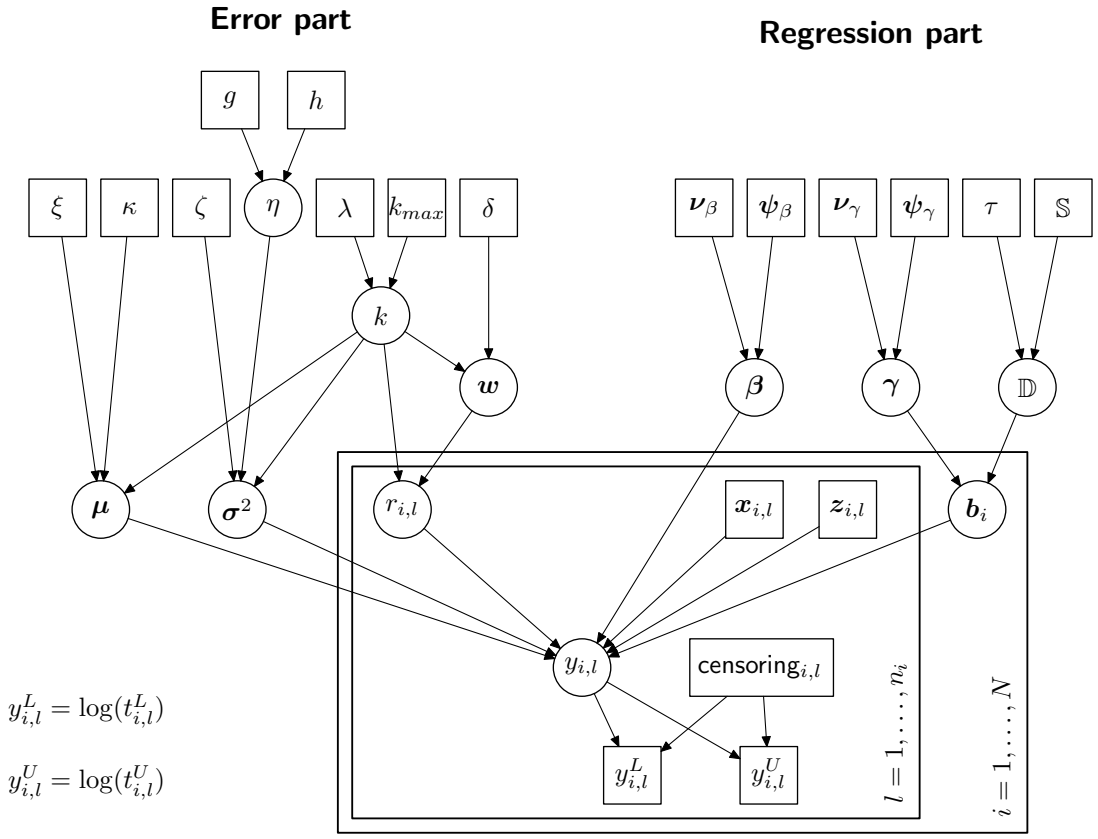

Figure 2. CGD Data – Recurrent Events Analysis. (a) posterior means and 95% PCI for random effects b_i; (b) predictive error densities; solid line: unconditionally, dotted line: $k = 1,\ 2$, dotted-dashed line: $k = 3,\ 4$, dashed line: $k = 5 - 10$.

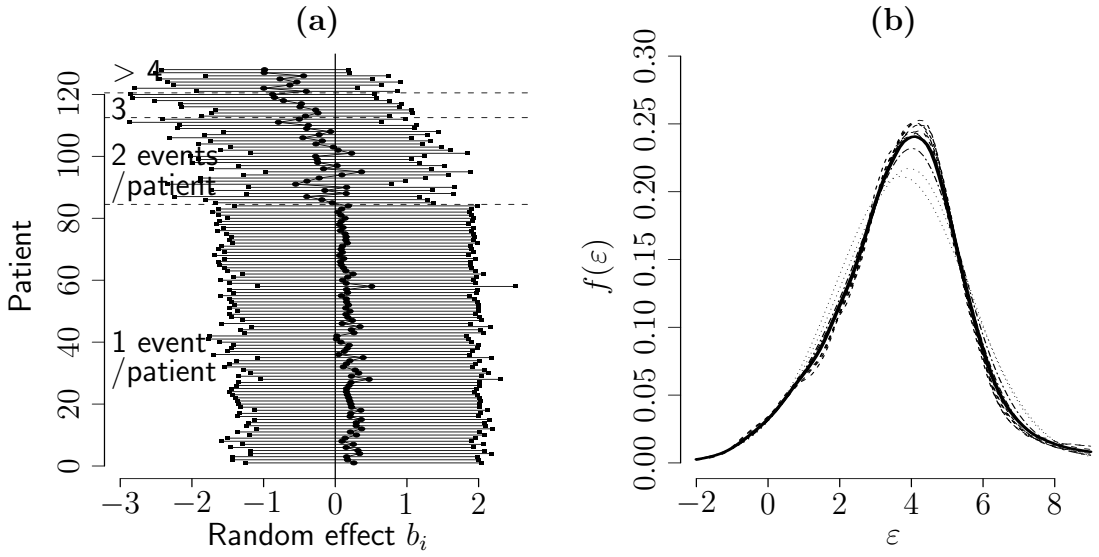

Figure 3. CGD Data – Recurrent Events Analysis. (a) predictive survivor and (b) hazard curves for males and females taking either treatment or placebo Remaining covariates were fixed to either mean value ($age = 14.6$) or to most common value (X-linked pattern of inheritance, no use of corticosteroids, use of prophylactic antibiotics and a hospital category US-other).

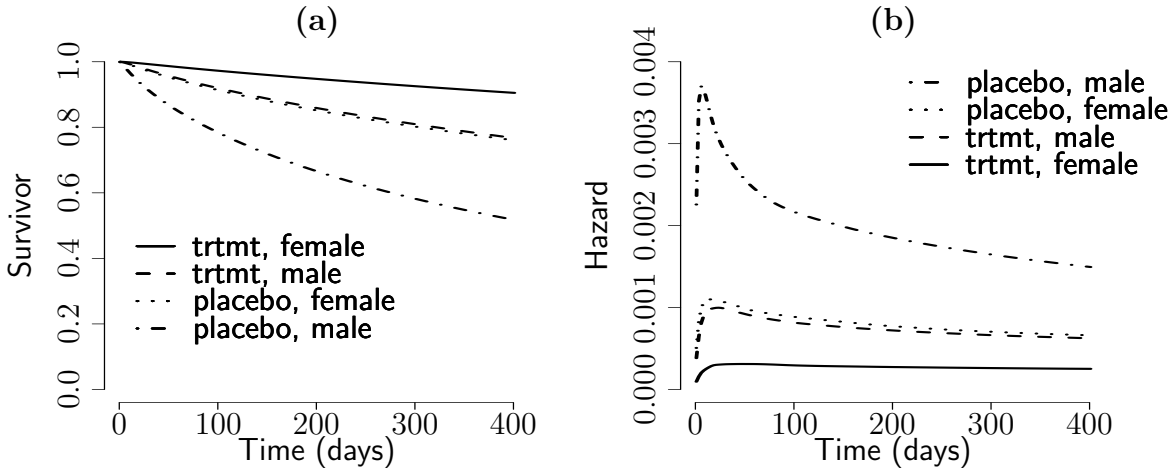

Figure 4. Signal Tandmobiel® data. Predictive emergence curves, dotted-dashed line for girls with $dmf > 0$, dotted line for girls with $dmf = 0$, dashed line for boys with $dmf > 0$ and solid line for boys with $dmf = 0$.