Statistica Sinica **17**(2007), 549-569

BAYESIAN ACCELERATED FAILURE TIME MODEL FOR CORRELATED INTERVAL-CENSORED DATA WITH A NORMAL MIXTURE AS ERROR DISTRIBUTION

Arnošt Komárek and Emmanuel Lesaffre

Katholieke Universiteit Leuven

Abstract: A Bayesian approach is proposed for an accelerated failure time model with interval-censored data. The model allows for structured correlated data by inclusion of a random effect part that might depend on covariates, as in a linear mixed model. The error distribution is modelled as a normal mixture with an unknown number of components. Also, the means and variances of the components are not prespecified so as to accommodate most continuous distributions. This results, among other things, in a nearly correct estimation of the shape of the hazard and survivor curves. A Markov chain Monte Carlo algorithm is described that samples from the posterior distribution. The approach is evaluated using a simulation study, and is illustrated by modeling the emergence times of eight permanent teeth using data from the Signal Tandmobiel® study.

Key words and phrases: Clustered Data, multicenter study, regression, reversible jump Markov chain Monte Carlo, survival data.

1. Introduction

Correlated survival data are encountered in many medical problems, for instance when the observations are clustered. Moreover, in some problems, the occurrence of the event can only be recorded at regular intervals leading to *interval-censored* data.

Correlated interval-censored survival times were encountered in the Signal Tandmobiel® study. This was a longitudinal, prospective (1996–2001) oral health screening project performed in Flanders, Belgium. The 4,468 school-children, born in 1989, were examined on a yearly basis by one of 16 trained dental-examiners. The details of the study design can be found in Vanobbergen et al. (2000). For oral health researchers, two questions are of interest. First, what is the effect of decayed primary predecessors (described by a binarised $dmft$ score) on the emergence times of the permanent premolars (teeth 14, 15, 24, 25, 34, 35, 44, 45 in European dental notation)? The emergences were observed only at yearly intervals. Second, what is the correlation between the emergence times of different teeth? Leroy et al. (2003) have shown that there is *horizontal*

symmetry, i.e., information about the correlation structure is available as the same emergence distribution can be assumed at horizontally symmetric positions (e.g., for teeth 14 and 24). Further, it is known that factors, like gender, have an impact on the emergence time and thus should be controlled for.

In multicenter clinical trials it is often necessary to control for center, and to check the center by treatment interaction, when evaluating the treatment effect. An important center effect or center by treatment interaction can be due to center differences in socio-economic characteristics of the patients, different training of the medical staff, differences in the administration of treatment, etc. To obtain a valid statistical conclusion on treatment efficacy, one should control for such effects. Furthermore, in situations where disease progression can only be revealed by a laboratory assessment (AIDS, some types of cancer), the observed event times are interval-censored as well.

The paper is organized as follows. In Section 2, we review models for correlated censored data. In Section 3 we propose a new approach – the Bayesian mixture MEAFT model. In Section 4, a Markov chain Monte Carlo algorithm that samples from the posterior distribution is discussed. The approach is evaluated using a simulation study in Section 5, and the Signal Tandmobiel® data in Section 6.

2. Models for Correlated Censored Data

Several approaches to analyze correlated right-censored survival times have been proposed. One approach is to extend the Cox's proportional hazards (PH) model (Cox (1972)) by including a cluster-specific random effect, called *frailty* in the expression of the hazard function (see, e.g., Hougaard (2000) and Therneau and Grambsch (2000)). The frailty component is most often assumed to have a parametric distribution such as gamma or log-normal. However, the frailty PH model has some important drawbacks. First, the implied correlation structure is too simple, e.g., in the analysis of the multicenter clinical trials only the center effect and not the center by treatment interaction can be controlled for. Second, the choice of the frailty distribution can have a crucial impact on the results for the regression parameters of interest (Hougaard (2000, Chap.7)). Third, the PH model is generally not robust toward neglected covariates (Hougaard (1999)).

A possible alternative to the PH model is the accelerated failure time (AFT) model which assumes that the covariates speed up or slow down the expected event time. We refer to Chapter 7 of Kalbfleisch and Prentice (2002) for an extensive review of classical approaches to the AFT model. In contrast to the PH model, neglected covariates in the AFT model do not cause bias in estimating

the regression parameters for the included covariates (Hougaard (1999)). An extension of the AFT model – the mixed effects accelerated failure time (MEAFT) model – takes into account the within-cluster correlations explicitly by including random effects in the regression expression, as in a classical linear mixed model of Laird and Ware (1982). Thus

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

where $T_{i,l}$ is the event time of the lth observation of the ith cluster, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is the unknown regression coefficient vector, $\boldsymbol{x}_{i,l}$ is the covariate vector for fixed effects, $\boldsymbol{b}_i = (b_{i,1}, \ldots, b_{i,q})^T$, $i = 1, \ldots, N$ are i.i.d. random effects vectors with a density $g(\boldsymbol{b})$, $\boldsymbol{z}_{i,l}$ is the covariate vector for random effects, and $\varepsilon_{i,l}$, $i = 1, \ldots, N$, $l = 1, \ldots, n_i$ are i.i.d. error random variables with a density $f(\varepsilon)$. As the roles of the regression parameters and dispersion parameters are clearly separated in the MEAFT, the regression parameters are robust against misspecification of the random effects distribution $g(\boldsymbol{b})$ (Keiding, Andersen and Klein (1997) and Lambert, Collett, Kimber and Johnson (2004)). Model (1), restricted to the case of $z_{i,l} \equiv 1$, has been studied by Pan and Louis (2000) and Pan and Connett (2001) for right-censored data. They make a working assumption concerning the normality of both $\varepsilon_{i,l}$ and b_i, and use frequentist techniques for uncensored data, combined with a Monte Carlo EM algorithm and multiple imputation, respectively, to overcome the problem of censoring.

Assume now that the (i,l)th true log-event time $y_{i,l}$ is only known to lie in the interval $(y_{i,l}^L, y_{i,l}^U]$, $-\infty \le y_{i,l}^L \le y_{i,l}^U \le \infty$. 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 $y_{i,l}^L = -\infty$. The likelihood contribution of the ith cluster is given by

$$L_i = \int_{\mathbb{R}^q} \Big\{ \prod_{l=1}^{n_i} \int_{y_{i,l}^L}^{y_{i,l}^U} f(y - \boldsymbol{\beta}^T \boldsymbol{x}_{i,l} - \boldsymbol{b}^T \boldsymbol{z}_{i,l})\, dy \Big\} g(\boldsymbol{b})\, d\boldsymbol{b}, \tag{2}$$

where the convention $\int_a^a f(s)\, ds \equiv f(a)$ applies to also accommodate uncensored observations. Due to multiple integration in the likelihood (2), it is rather cumbersome to use maximum-likelihood based methods for the MEAFT model with interval-censored observations, even with $f(\varepsilon)$ and $g(\boldsymbol{b})$ being parametrically specified. While stochastic versions of standard estimation techniques can be used, as was done by Pan and Louis (2000) or Pan and Connett (2001), we believe that a Bayesian approach is more natural, and easier to use here.

Furthermore for small samples or in situations when prediction, and not only regression parameters themselves, are of interest, it is desirable to avoid full parametric assumptions (like normality) concerning the error density $f(\varepsilon)$ that

determines the shape and character of resulting survival and hazard curves that are to be estimated from the data. For that reason, we sought a method with enough flexibility in specifying the error density, while still being computationally tractable for both interval-censored data and a general covariate vector $\boldsymbol{z}_{i,l}$.

3. A Bayesian Mixture MEAFT Model

To our best knowledge, the MEAFT model with a general q-variate random effects covariate vector $\boldsymbol{z}_{i,l}$ is required to solve the problems outlined in the introduction, and a flexible error density has not yet been considered in the literature. To model unknown distributional shapes, *finite mixture* distributions have been advocated by, e.g., Titterington, Smith and Makov (1985), Section 2.2 as appealing *semi-parametric* structures. However, until the last decade the statistical analysis of mixtures has not been straightforward. The use of a reversible jump MCMC algorithm (Green (1995)) to estimate unknown mixture parameters, suggested by Richardson and Green (1997), is a breakthrough in this area. We adopt their approach to model the error density in the MEAFT model, and argue that this offers a rich family of distributions of various shapes suitable for modeling practically any survival data. See Section 5 for some examples of densities and corresponding hazard or survivor functions as approximated by normal mixtures.

At the same time, the MCMC methodology easily overcomes a problem of the difficult likelihood as (2). Indeed, there is no need to maximize this likelihood since the sample from the posterior distribution obtained using the MCMC method is used to draw inferences. Furthermore, MCMC replaces both integrals in (2) with the sampling of exact event times and values of latent random effects from appropriate, ease-to-sample distributions, as will be shown in Sections 4.1 and 4.4.

We assume a Bayesian mixture MEAFT model (1) with a hierarchical structure graphically represented by the directed acyclic graph (DAG) given in Figure 1, where the usual convention of graphical models is used: square boxes represent fixed or observed quantities and circles the unknown parameters; solid lines represent stochastic and dashed lines express deterministic dependencies, respectively. The joint prior distribution is given by the product of the conditional distributions of each node given its parents, as discussed in this section. As the DAG indicates, the unknown parameters can be split into two parts connected only through the node of true log-event times.

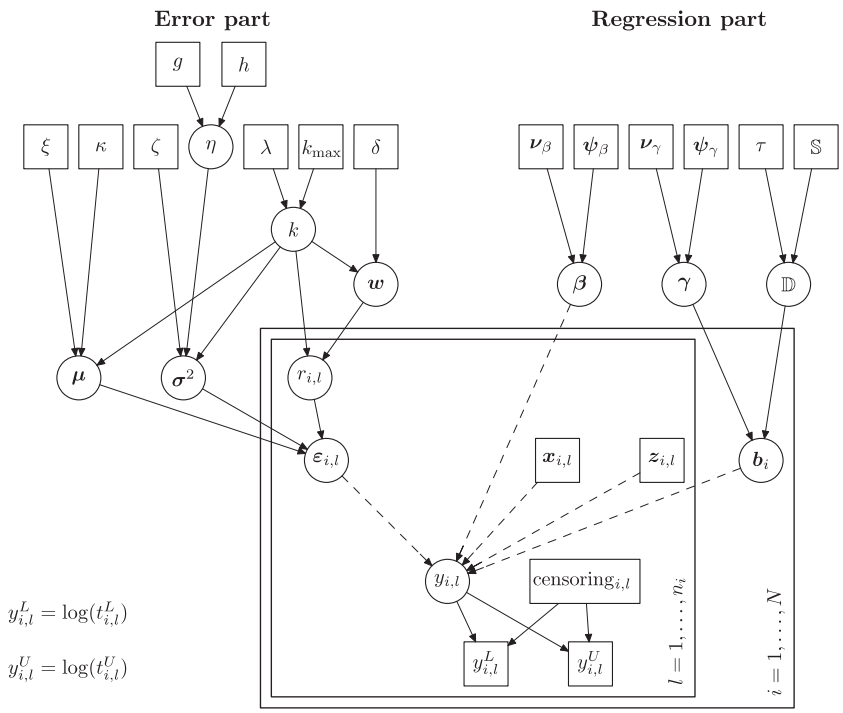

Figure 1. DAG for the Bayesian AFT model.

3.1. Prior specification of the error part

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

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

with $\varphi(\cdot \mid \mu_j, \ \sigma_j^2) \equiv$ density of $\mathcal{N}(\mu_j, \sigma_j^2)$. Note that the number of mixture components, k, is unknown, as well as the mixture weights $\boldsymbol{w} = (w_1, \ldots, w_k)^T$, means $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^T$, and variances $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_k^2)^T$. It is well-known (McLachlan and Basford (1988, Chapt. 2)) that a heteroscedastic mixture (3) leads to an unbounded likelihood if the parameter space for variances is unconstrained. In a Bayesian analysis, this difficulty is solved by using an appropriate prior distribution for the variances.

To improve the computation of the posterior distribution, it is useful to assume that $\varepsilon_{i,l}$, $i = 1, \ldots, N$, $l = 1, \ldots, n_i$ come from a heterogeneous population consisting of groups $j = 1, \ldots, k$, of sizes proportional to w_j, and to introduce latent allocation variables $r_{i,l}$ denoting the label of the group from which each

random error variable $\varepsilon_{i,l}$ is drawn. The corresponding DAG conditional distributions are then given by

$$\varepsilon_{i,l} \,|\, \boldsymbol{\mu},\, \boldsymbol{\sigma}^2,\, r_{i,l} \sim \mathcal{N}(\mu_{r_{i,l}}, \sigma^2_{r_{i,l}}), \qquad i = 1, \ldots, N,\ l = 1, \ldots, n_i, \tag{4}$$

$$\Pr(r_{i,l} = j \,|\, k,\, \boldsymbol{w}) = w_j, \qquad j = 1, \ldots, k. \tag{5}$$

DAG conditional distributions of the remaining parameters of the error part of the model 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 (1) a Poisson distribution with mean equal to a hyper-parameter λ truncated at some prespecified (relatively large) value $k_{\max}$, and (2) a uniform distribution on $\{1, \ldots, k_{\max}\}$ (the node λ in the DAG in Figure 1 then becomes redundant). The prior for the mixture weights $\boldsymbol{w}$ is taken to be a symmetric k-dimensional Dirichlet with prior 'sample size' equal to $k\,\delta$, i.e., $\boldsymbol{w}\,|k, \delta \ \sim\ \mathrm{D}(\delta, \delta, \ldots, \delta)$, where δ is a fixed hyper-parameter. Further, the mixture means μ_j and variances σ_j^2 are taken as independent with normal and inverse-gamma priors $\mu_j | k, \xi, \kappa \sim \mathcal{N}(\xi, \kappa)$ and $\sigma_j^2 | k, \zeta, \eta \sim \mathrm{IG}(\zeta, \eta)$, respectively. As in Richardson and Green (1997) we let η have a gamma distribution $\mathrm{G}(g, h)$ with fixed hyper-parameters g and h, see Section 3.4 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 establishing 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.

3.2. Prior specification of the regression part

The regression part of the model has the structure of a classical Bayesian linear mixed model (see, e.g., Gelman, Carlin, Stern and Rubin (2004, Chap.5)).

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 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 in general results in better behavior of the MCMC algorithm (Gelfand, Sahu and Carlin (1995)). Finally, since f does not have mean zero, we do not allow a column of ones in the matrix $\mathbb{X}$ and thus avoid identifiability problems.

The prior distribution for each regression coefficient β_j is assumed to be $\mathcal{N}(\nu_{\beta,j}, \psi_{\beta,j})$, $j = 1, \ldots, p$, 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 hyper-parameters.

The prior distribution for the random effect vector $\boldsymbol{b}_i$ is

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

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^T$. The prior distribution for each γ_j, is $\mathcal{N}(\nu_{\gamma,j}, \psi_{\gamma,j})$, 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 a 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 (equivalently, $\nu_{\gamma,1} = 0,\ \psi_{\gamma,1} = 0$).

The prior distribution for the covariance matrix $\mathbb{D}$ of random effects is assumed to be an inverse-Wishart IW(τ, $\mathbb{S}$) (parametrized such that the mean is $(\tau - q - 1)^{-1}\mathbb{S}$, where τ denotes 'degrees of freedom' ($\tau > q - 1$), and $\mathbb{S}$ is a scale matrix).

Finally, the DAG conditional distributions of the (unknown) log-event times, i.e., nodes that connect the regression and error parts of the DAG, are all deterministic and given by the MEAFT model (1).

3.3. Censoring

To complete the specification of the DAG we need to specify $p(y_{i,l}^L, y_{i,l}^U \,|\, y_{i,l},$ censoring). First, the censoring mechanism in this paper is assumed to be non-informative 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 the data, and the data consist of the realized censoring variables as well.

After omitting subscripts i, l for clarity, the form of $p(y^L, y^U \,|\, y,$ censoring) is 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 discrete 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, the density $p(y^L, y^U \,|\, y, c_1, \ldots, c_m)$ is again a discrete 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 $c_0 = -\infty$, $c_{m+1} = \infty$.

3.4. 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 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 hyper-parameters for the covariance matrix $\mathbb{D}$ that give 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 $\mathcal{N}(\xi, \kappa)$ for μ_j is achieved by letting ξ equal $\bar{e} = n^{-1} \sum_{j=1}^{n} e_j$, 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 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, which is necessary to avoid a problem of unbounded likelihood, without being informative about their absolute size. Finally they suggest 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 an explicitly included intercept and scale parameters. This can be done using such 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.

4. Markov Chain Monte Carlo Algorithm

Details of the implementation of the MCMC algorithm for the parameters of the error part of the model are given in Richardson and Green (1997). Their guidelines, now based on residuals $\varepsilon_{i,l} = y_{i,l} - \boldsymbol{\beta}^T \boldsymbol{x}_{i,l} - \boldsymbol{b}_i^T \boldsymbol{z}_{i,l}$, 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, Sec. 9) for the dimension changing steps (split-combine and birth-death moves).

For the regression part of the model, each iteration of the MCMC is conducted using the Gibbs sampler (Geman and Geman (1984)). The full conditional distributions needed to implement the Gibbs sampler are given below. The notation $|\ldots$ indicates that conditioning is done on all remaining parameters.

4.1. True log-event times $y_{i,l}$

The full conditional distribution of each $y_{i,l}$ is

$$y_{i,l} \mid \cdots \sim \mathcal{N}(\mu_{r_{i,l}} + \boldsymbol{\beta}^T \boldsymbol{x}_{i,l} + \boldsymbol{b}_i^T \boldsymbol{z}_{i,l},\ \sigma^2_{r_{i,l}}) \text{ truncated on } (y^L_{i,l}, y^U_{i,l}]. \tag{7}$$

4.2. Fixed effects $\boldsymbol{\beta}$

Let $\boldsymbol{\beta}_{(S)}$ be an arbitrary sub-vector of vector $\boldsymbol{\beta}$, and $\boldsymbol{x}_{i,l(S)}$ the corresponding sub-vectors of covariate vectors $\boldsymbol{x}_{i,l}$, and further let $\boldsymbol{x}_{i,l(-S)}$ be their complementary sub-vectors. Similarly, let $\boldsymbol{\nu}_{\beta(S)}$ and $\boldsymbol{\psi}_{\beta(S)}$ be appropriate sub-vectors of hyper-parameters $\boldsymbol{\nu}_\beta$ and $\boldsymbol{\psi}_\beta$, respectively. Finally, let $\Psi_{\beta(S)} = \operatorname{diag}(\boldsymbol{\psi}_{\beta(S)})$. Then

$$\boldsymbol{\beta}_{(S)} \mid \cdots \sim \mathcal{N}\Big(E[\boldsymbol{\beta}_{(S)} \mid \cdots], \operatorname{Var}[\boldsymbol{\beta}_{(S)} \mid \cdots]\Big), \tag{8}$$

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

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

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

4.3. Means of random effects $\boldsymbol{\gamma}$

There is no loss of generality in assuming that $\boldsymbol{\gamma} = (\boldsymbol{\gamma}^T_{(S)}, \boldsymbol{\gamma}^T_{(-S)})^T$, with $\boldsymbol{b}_{i(S)}$, $\boldsymbol{b}_{i(-S)}$, $\boldsymbol{\nu}_{\gamma(S)}$, $\boldsymbol{\psi}_{\gamma(S)}$ the corresponding sub-vectors or complementary sub-vectors of indicated quantities and $\Psi_{\gamma(S)} = \operatorname{diag}(\boldsymbol{\psi}_{\gamma(S)})$. Furthermore, let

$$\mathbb{D}^{-1} = \begin{pmatrix} \mathbb{V}_{(S)} & \mathbb{V}_{(S,-S)} \\ \mathbb{V}^T_{(S,-S)} & \mathbb{V}_{(-S)} \end{pmatrix}. \tag{9}$$

Then

$$\boldsymbol{\gamma}_{(S)} \mid \cdots \sim \mathcal{N}\Big(E[\boldsymbol{\gamma}_{(S)} \mid \cdots], \operatorname{Var}[\boldsymbol{\gamma}_{(S)} \mid \cdots]\Big), \tag{10}$$

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

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

4.4. Random effects $\boldsymbol{b_i}$

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

$$\boldsymbol{b}_i \mid \cdots \sim \mathcal{N}\Big(E[\boldsymbol{b}_i \mid \cdots], \mathrm{Var}\,[\boldsymbol{b}_i \mid \cdots]\Big), \qquad i = 1, \ldots, N, \tag{11}$$

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

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

4.5. 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 of $\mathbb{S} + \sum_{i=1}^{N} (\boldsymbol{b}_i - \boldsymbol{\gamma})(\boldsymbol{b}_i - \boldsymbol{\gamma})^T$.

4.6. Software

Programs in C++ have been written with an interface to the R language (R Development Core Team (2006)) as a contributed package `bayesSurv` and can be downloaded, together with a comprehensive description of how to perform the analyzes presented in this paper, from the Comprehensive R Archive Network (CRAN) on `http://www.R-project.org`.

5. Simulation Study

A simulation study was carried out to explore the performance of the proposed method. The setting mimics a typical multicenter study with a possible center by treatment interaction. 'True' uncensored data were generated according to the MEAFT model

$$\log(T_{i,l}) = 1.5 + \beta\, x_{i,l} + b_{i,1} + b_{i,2}\, z_{i,l} + \varepsilon_{i,l}, \quad i = 1, \ldots, N,\ l = 1, \ldots, n_i, \tag{12}$$

where $\beta = 0.4$, $(b_{i,1},\, b_{i,2})' \sim \mathcal{N}_2\big((0, \gamma)', \mathbb{D}\big)$, $\gamma = -0.8$, $\mathrm{var}(b_{i,1}) = 0.5^2$, $\mathrm{var}(b_{i,2}) = 0.1^2$, and $\mathrm{corr}(b_{i,1}, b_{i,2}) = 0.4$. The covariate $x_{i,l}$ was generated according to the extreme-value distribution of a minimum, with location equal to 8.5 and scale equal to 1, inspired more or less by the $\log_2(1+\text{CD4 count})$ covariate in the AIDS dataset analyzed by Komárek, Lesaffre and Hilton (2005). The covariate $z_{i,l}$ (*treatment vs. placebo*) was binary, taking a value of 1 with probability 0.4. The error term $\varepsilon_{i,l}$ was generated from a standard normal distribution, from a Cauchy distribution, from a Student t_2 distribution, from a standardized extreme value distribution, and from a normal mixture $0.4\mathcal{N}(-2.000,\, 0.25) + 0.6\mathcal{N}(1.333,\, 0.36)$. Two sample sizes were considered: (1) $N = 50$, $n_i = 5$ for

all i (small sample size), and (2) $N = 100$, $n_i = 10$ for all i (large sample size). Each simulation involved 100 replications.

All event times were interval-censored by simulating 120 consecutive 'assessment times' for each 'patient' in the dataset (the first assessment time was drawn from $\mathcal{N}(7, 1)$, times between consecutive assessments from $\mathcal{N}(6, 0.25)$). At each assessment, between 0.2% and 0.6% of randomly selected patients were withdrawn from the study, resulting in approximately 15% right-censored observations. For each dataset, the estimates were computed using the Bayesian mixture MEAFT model, using the Bayesian MEAFT model with a normal error, and using the maximum-likelihood AFT model with a normal error and ignoring the random effects structure.

Table 1. Simulation study. Results for the regression parameters: average estimate, mean squared error ($\times 10^{-4}$) in brackets.

Setting	Bayesian mixture	Bayesian normal	ML, no random eff.
	$\beta = 0.4$		
Normal, small	0.3966 (26.3)	0.3973 (25.6)	0.3992 (34.7)
large	0.4016 (7.3)	0.4018 (7.3)	0.4022 (9.0)
Cauchy t_1, small	0.4122 (52.5)	0.3832 (68.2)	0.3783 (124.6)
large	0.3921 (13.2)	0.3608 (41.8)	0.3571 (50.7)
Student t_2, small	0.3933 (58.1)	0.3859 (49.6)	0.3823 (72.9)
large	0.3944 (11.5)	0.3794 (19.0)	0.3780 (21.6)
Extr. value, small	0.3928 (17.9)	0.3954 (20.9)	0.3952 (25.9)
large	0.4036 (4.3)	0.4035 (5.4)	0.4022 (6.8)
Mixture, small	0.3942 (17.6)	0.4324 (68.1)	0.4436 (127.3)
large	0.3997 (3.5)	0.4480 (45.6)	0.4505 (52.9)
	$\gamma = -0.8$		
Normal, small	−0.8128 (240.4)	−0.8105 (222.8)	−0.8121 (235.7)
large	−0.7981 (47.0)	−0.7983 (48.2)	−0.7982 (60.4)
Cauchy t_1, small	−0.7656 (512.7)	−0.7192 (716.4)	−0.7210 (704.0)
large	−0.8097 (107.2)	−0.7360 (234.5)	−0.7383 (238.9)
Student t_2, small	−0.7777 (479.0)	−0.7593 (401.3)	−0.7614 (415.5)
large	−0.7933 (99.9)	−0.7610 (123.7)	−0.7601 (132.1)
Extr. value, small	−0.8150 (191.1)	−0.8106 (192.4)	−0.8094 (202.3)
large	−0.7969 (47.4)	−0.7999 (56.6)	−0.8022 (66.9)
Mixture, small	−0.7868 (95.7)	−0.8693 (895.0)	−0.8635 (840.5)
large	−0.8040 (26.6)	−0.9264 (366.8)	−0.9227 (369.4)

Table 1 shows the average estimates of the regression parameters and their mean squared errors. It is seen that, in most cases, the Bayesian mixture approach performs better than the incorrectly specified models. A large difference

in favour of the Bayesian mixture model is seen in the case of a normal mixture or a Cauchy as the error distribution. Additionally, when the Bayesian mixture approach is used the error distribution, and consequently also the hazard or survivor functions, are reproduced closely. This is not always the case when the Bayesian normal model is used. Figure 2 shows the behaviour of our estimated hazard functions for two heavy-tailed distributions, small sample size, using either the Bayesian mixture or the Bayesian normal model. A similar comparison for the survivor functions of the extreme value or mixture error is shown in Figure 3. Figure 4 shows the behaviour of the Bayesian mixture method when the sample size increases, for the extreme value distribution and mixture distribution.

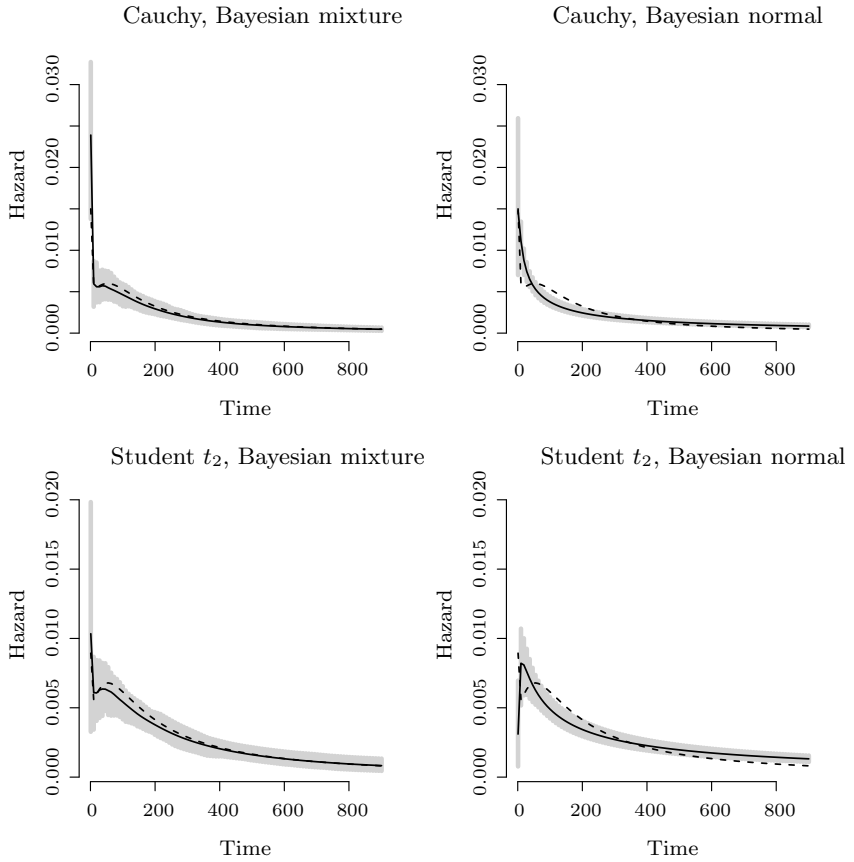

Figure 2. Simulation study – hazard functions. Results for a small sample size ($N = 50$, $n_i = 5$) and Cauchy and Student t_2 error distributions. Left column: estimates based on the Bayesian mixture model; right column: estimates based on the Bayesian normal model. Solid line: estimate of the hazard function for $x = 8.13$ – median value and $z = 0$; gray region: simulation based 95% point-wise confidence interval; dashed line: true hazard function.

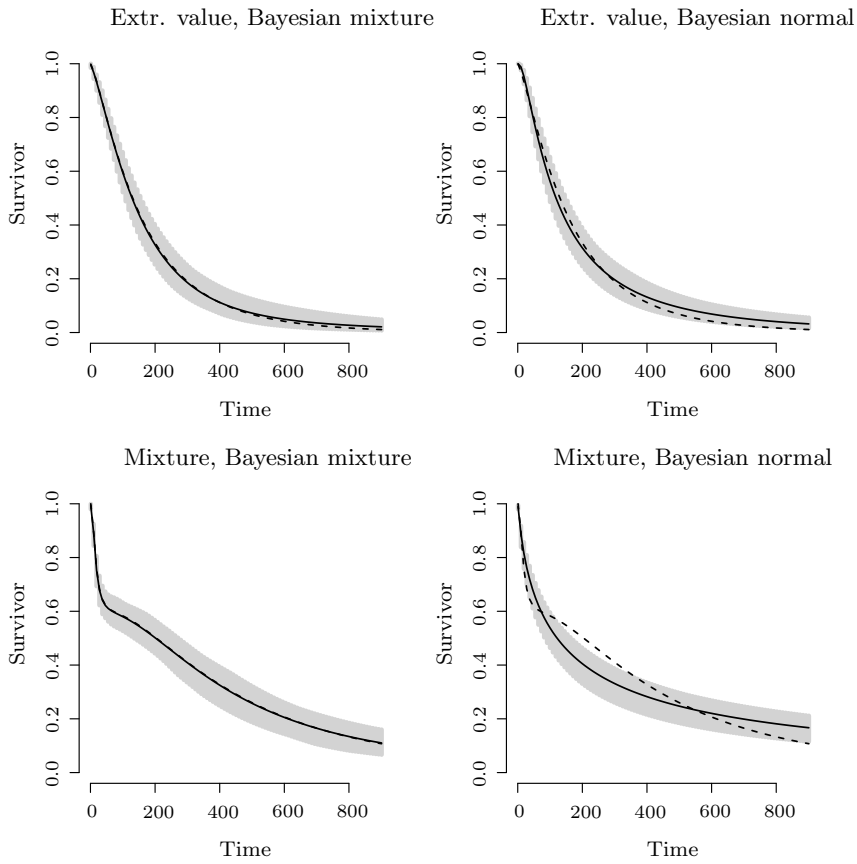

Figure 3. Simulation study – survivor functions. Results for a small sample size ($N = 50$, $n_i = 5$) and extreme value and normal mixture error distributions. Left column: estimates based on the Bayesian mixture model; right column: estimates based on the Bayesian normal model. Solid line: estimate of the survivor function for $x = 8.13$ – median value and $z = 0$; gray region: simulation based 95% point-wise confidence interval; dashed line: true survivor function.

6. Analysis of Signal Tandmobiel® Data

The first research question outlined in the introduction was considered by Lesaffre, Komárek and Declerck (2005) who analyzed each tooth separately using the penalized AFT model of Komárek, Lesaffre and Hilton (2005). With the Bayesian MEAFT model of this paper, we analyze all teeth jointly and answer also the second research question. A random sample of 500 boys and 500 girls is used for the inference.

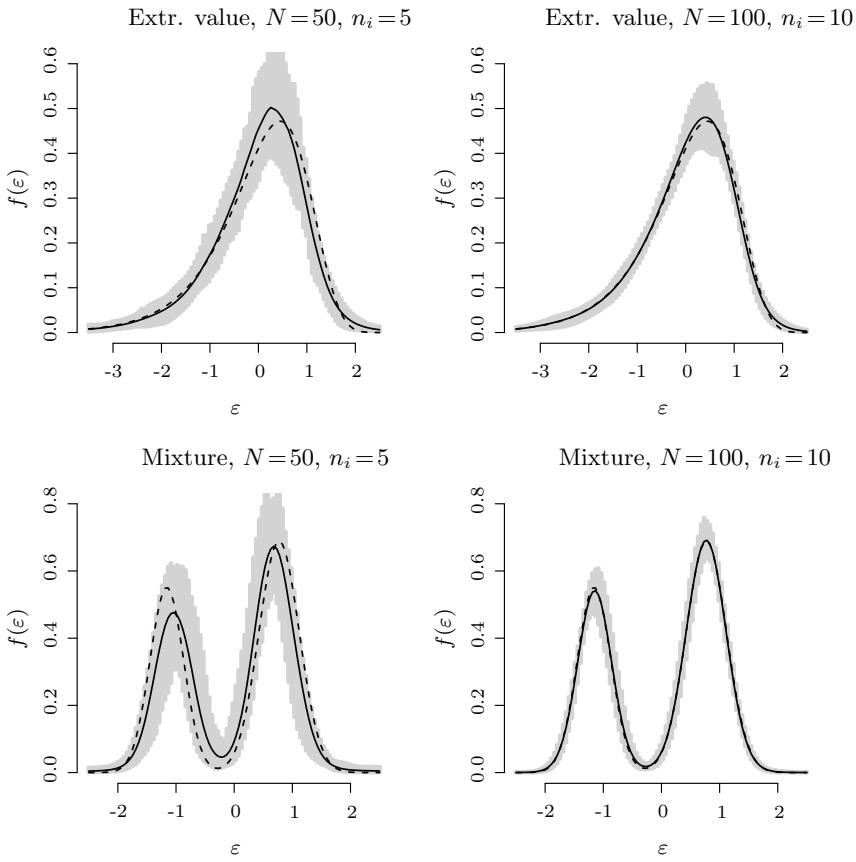

Figure 4. Simulation study – error densities. Results for both sample sizes and extreme value and normal mixture error distributions, estimates based on the Bayesian mixture model. Solid line: estimate of the standardized error density; gray region: simulation based 95% point-wise confidence interval; dashed line: true standardized error density.

For a better fit, we shifted the time origin of the MEAFT model to 5 years of age, replacing $T_{i,l}$ by $T_{i,l}-5$ at (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 $man4_{i,l}$, $max5_{i,l}$, $man5_{i,l}$ 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 at (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. As fixed effects we used $gender \equiv girl$, $dmft$, and all two-way

interaction terms between *girl*, $dmft$ and dummies for the pairs of horizontally symmetric teeth.

The initial maximum-likelihood AFT model, for each tooth separately, with a normal error distribution and without random effects, estimated the intercept as 1.8 and the scale as 0.25. According to the suggestions of Section 3.4 we used the following values of hyper-parameters: $\xi = 1.8$, $\kappa = (3 \cdot 0.25)^2$, $\zeta = 2$, $g = 0.2$, $h = 0.1$, $\delta = 1$. For the number of mixture components, k, a truncated Poisson prior with $\lambda = 5$ reflected our prior belief that the error distribution is skewed, and $k_{\max} = 30$ was used. All β and γ parameters were assigned a $\mathcal{N}(0, 100)$ prior. 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. The prior scale matrix $\mathbb{S}$ was diag(0.002) (corresponding to inverse-gamma$(\tau, 0.001)$ in the univariate case).

We sampled two chains, each of length 20,000 with 1:3 thinning. This took about 27 hours on a Pentium IV 2 GHz PC. The first 1,500 iterations of each chain were discarded. The convergence was evaluated by a critical examination of the trace and autocorrelation plots, and by using the method of Gelman and Rubin (1992).

6.1. Regression parameters

In this analysis, the main interest lies in the effect of $dmft$ on emergence. This can be evaluated from Table 2 where posterior summary statistics for the effect of $dmft > 0$ (appropriate linear combinations of β parameters) for the two genders and the four pairs of horizontally symmetric teeth are given. As a point estimate we report the posterior median, which can easily be obtained from an MCMC sample and still corresponds roughly to the maximum-likelihood estimate. Indeed, if the (log-)posterior distribution is unimodal and symmetric (which happened for practically all regression parameters) the posterior median is the same as the posterior mode. For skewed posterior distributions (variance components), the log-posterior median is practically the same as the log-posterior mode.

It is seen that bad status of the primary predecessor (*dmft*$= 1$) accelerates the emergence of the permanent successor in the case of maxillary teeth, and significantly. For the mandibular teeth, a slight effect is observed only for the first premolar on boys. Additionally, besides the effect of $dmft$, the emergence process for girls precedes that of boys.

Table 2. Signal Tandmobiel® data. Posterior medians, 95% equal-tail credibility intervals, and Bayesian two-sided p-values for the effect of $dmft > 0$ for the two genders and different teeth.

maxilla 4		maxilla 5	
girl	boy	girl	boy
-0.0352	-0.0457	-0.0212	-0.0317
$(-0.0522,\ -0.0185)$	$(-0.0631,\ -0.0284)$	$(-0.0390,\ -0.0035)$	$(-0.0500,\ -0.0135)$
$p < 0.001$	$p < 0.001$	$p = 0.019$	$p = 0.001$
mandible 4		mandible 5	
girl	boy	girl	boy
-0.0098	-0.0201	0.0015	-0.0090
$(-0.0267,\ 0.0070)$	$(-0.0378,\ -0.0032)$	$(-0.0162,\ 0.0193)$	$(-0.0283,\ 0.0098)$
$p = 0.255$	$p = 0.021$	$p = 0.870$	$p = 0.353$

6.2. Predictive emergence curves

In dentistry, predictive cumulative distribution functions (cdf) are preferred over the survivor functions in the case of emergence, and are known as *emergence curves*. Let $\boldsymbol{\theta}$ denote all unknown quantities of the model. For a specific value of covariates, say $\boldsymbol{x}_{new}$ and $\boldsymbol{z}_{new}$, the predictive cdf is given by

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

for any $t > 0$. Further

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

where $\Phi(\cdot \mid \mu_j, \sigma_j^2)$ is a cumulative distribution function of $\mathcal{N}(\mu_j, \sigma_j^2)$. The MCMC estimate of the predictive cdf is then given by $\hat{F}(t \mid \text{data}) = M^{-1} \sum_{m=1}^{M} F(t \mid \boldsymbol{\theta}^{(m)})$, where $\boldsymbol{\theta}^{(m)}, m = 1, \ldots, M$, is the MCMC sample from the posterior (predictive) distribution. All components of $\boldsymbol{\theta}^{(m)}$ are directly available except $\boldsymbol{b}^{(m)}$. These last must be additionally sampled from $\mathcal{N}_q(\boldsymbol{\gamma}^{(m)}, \mathbb{D}^{(m)})$. Predictive survivor or hazard curves can be obtained in an analogous manner.

Predictive emergence curves for the maxillary first premolar are shown in Figure 5. As a model check, Figure 5 also shows non-parametric estimates of the emergence curves computed separately in each group using the classical method of Turnbull (1976).

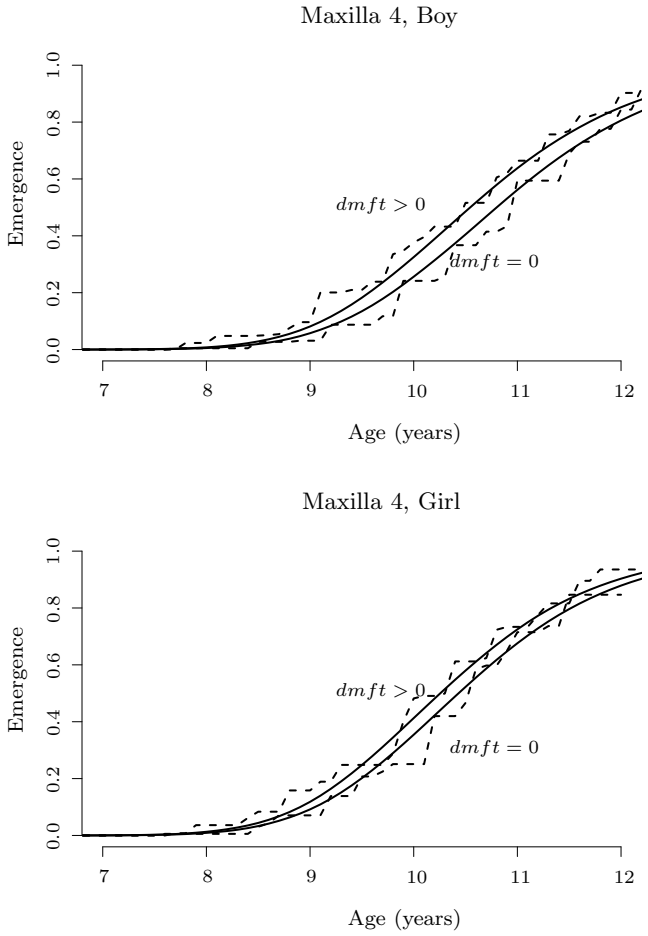

Figure 5. Signal Tandmobiel® data. Predictive emergence curves (solid lines) compared to the non-parametric estimate of Turnbull (dashed lines) for maxillary first premolars.

6.3. Inter-teeth relationship

Finally, Table 3 shows posterior summary statistics for variances and correlations of the 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 that the child effect is important, and that the different teeth in one mouth are strongly correlated. The posterior medians of variance parameters in Table 3 are all about 0.04, which is approximately four times higher than the posterior median of the variance of the error distribution, 0.01. Posterior medians of all correlation parameters lie between 0.79 and 0.91.

Table 3. Signal Tandmobiel® data. Posterior medians, 95% equal-tail credibility intervals for variances, and correlations between tooth-specific linear combinations of random effects.

$\mathrm{Var}\,(d_{\mathrm{max4}})$	$\mathrm{Var}\,(d_{\mathrm{man4}})$	$\mathrm{Var}\,(d_{\mathrm{max5}})$	$\mathrm{Var}\,(d_{\mathrm{man5}})$
0.042	0.039	0.042	0.041
(0.037, 0.047)	(0.035, 0.045)	(0.036, 0.049)	(0.035, 0.048)
$\mathrm{Corr}(d_{\mathrm{max4}}, d_{\mathrm{man4}})$	$\mathrm{Corr}(d_{\mathrm{max4}}, d_{\mathrm{max5}})$	$\mathrm{Corr}(d_{\mathrm{max4}}, d_{\mathrm{man5}})$	$\mathrm{Corr}(d_{\mathrm{man4}}, d_{\mathrm{max5}})$
0.887	0.914	0.842	0.793
(0.856, 0.914)	(0.887, 0.938)	(0.804, 0.874)	(0.749, 0.832)
$\mathrm{Corr}(d_{\mathrm{man4}}, d_{\mathrm{man5}})$	$\mathrm{Corr}(d_{\mathrm{max5}}, d_{\mathrm{man5}})$		
0.895	0.847		
(0.864, 0.923)	(0.810, 0.880)		

7. 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. Subsequently, the effect of covariates can be evaluated jointly with the association among clustered responses. Further, interval-, right-, or left-censored data are easy to handle and, finally, the MCMC sampling-based implementation of the model offers a straightforward way to obtain credibility intervals of model parameters as well as predictive survivor or hazard curves.

Observe that the Bayesian approach is used here mainly for technical convenience. Indeed, in practice the likelihood at (2) is hardly tractable using the maximum-likelihood method. On the other hand, Bayesian estimation using the MCMC does not pose any real difficulties. Further, since all our prior distributions are non-informative (at least approximately) and we use (on a proper scale) more or less posterior modes as point estimates, the classical maximum-likelihood estimation would lead to almost the same results.

The proposed methodology contributes to the area of *semi-parametric* modeling of *correlated*, and at the same time, *interval-censored* data. Furthermore, our approach allows one to bring in a structure into the dependencies between observations in one cluster. For example, in multicenter studies, the vector $\boldsymbol{z}_{i,l} = (1, \mathit{treatment}_{i,l})'$ at (1) allows consideration of not only the random center effect, but also a random center-by-treatment interaction that can sometimes be substantial.

According to our best knowledge, no approach is available which tackles this complex data structure. With varying amounts of effort, some of the existing

semi-parametric approaches mentioned in Section 2 could be, of course, extended to handle regression with correlated interval-censored data.

Unfortunately, our approach cannot handle time-dependent covariates. However, the same is true for any model in which the distribution of the response is specified by the density and not by the hazard function. To include time-dependent covariates, Cox's proportional hazards model is commonly used. For example, Kooperberg and Clarkson (1997), Betensky et al. (1999), and Goetghebeur and Ryan (2000) consider independent interval-censored data. Vaida and Xu (2000) offer an approach based on the proportional hazards linear mixed model with right-censored data.

Finally, our approach can be quite easily extended, along the lines presented in Komárek and Lesaffre (2007), to also handle data where the response is given as the difference of two interval-censored observations.

Acknowledgements

The work of the first author was financially supported by Research Grants BDB-B/05/10 and PDM/06/242, Katholieke Universiteit Leuven. The authors further acknowledge support from the Interuniversity Attraction Poles Programs P5/24 and P6/03 – Belgian State – Federal Office for Scientific, Technical and Cultural Affairs.

Data collection for the example in Section 6 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

Betensky, R. A., Lindsey, J. C., Ryan, L. M. and Wand, M. P. (1999). Local EM estimation of the hazard function for interval-censored data. *Biometrics* **55**, 238-245.

Brooks, S. P., Giudici, P. and Roberts, G. O. (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions (with discussion). *J. Roy. Statist. Soc. Ser. B* **65**, 3-55.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

Diebolt, J. and Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *J. Roy. Statist. Soc. Ser. B* **56**, 363-375.

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

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian Data Analysis*. Second edition. Chapman & Hall/CRC, Boca Raton, ISBN 1-58488-388-X.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulations using multiple sequences (with discussion). *Statist. Sci.* **7**, 457-511.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayes restoration of image. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721-741.

Goetghebeur, E. and Ryan, L. (2000). Semiparametric regression analysis of interval-censored data. *Biometrics* **56**, 1139-1144.

Green, P. J. (1995). Reversible jump Markov chain computation and Bayesian model determination. *Biometrika* **82**, 711-732.

Hougaard, P. (1999). Fundamentals of survival data. *Biometrics* **55**, 13-22.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer-Verlag, New York. ISBN 0-387-98873-4.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Second edition. Wiley, Chichester, ISBN 0-471-36357-X.

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. *Statist. Medicine* **16**, 215-225.

Komárek, A. and Lesaffre, E. (2007). Bayesian accelerated failure time model with multivariate doubly-interval-censored data and flexible distributional assumptions. To appear in *J. Amer. Statist. Assoc.*.

Komárek, A., Lesaffre, E. and Hilton, J. F. (2005). Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *J. Comput. Graph. Statist.* **14**, 726-745.

Kooperberg, C. and Clarkson, D. B. (1997). Hazard regression with interval-censored data. *Biometrics* **53**, 1485-1494.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.

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. *Statist. Medicine* **23**, 3177-3192.

Leroy, R., Bogaerts, K., Lesaffre, E. and Declerck, D. (2003). The emergence of permanent teeth in Flemish children (Belgium). *Community Dentistry and Oral Epidemiology* **31**, 30-39.

Lesaffre, E., Komárek, A. and Declerck, D. (2005). An overview of methods for interval-censored data with an emphasis on applications in dentistry. *Statist. Meth. Medical Res.* **14**, 539-552.

McLachlan, G. J. and Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, New York. ISBN 0-8247-7691-7.

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

R Development Core Team (2006). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org. ISBN 3-900051-07-0.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with unknown number of components (with discussion). *J. Roy. Statist. Soc. Ser. B* **59**, 731-792.

Roeder, K. and Wasserman (1997). Practical Bayesian density estimation using mixtures of normals. *J. Amer. Statist. Assoc.* **92**, 894-902.

Stephens, M. (2000). Dealing with label switching in mixture models. *J. Roy. Statist. Soc. Ser. B* **62**, 795-809.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York. ISBN 0-387-98784-3.

Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons, Chichester. ISBN 0-471-90763-4.

Turnbull, B. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser. B* **37**, 290-295.

Vaida, F. and Xu, R. (2000). Proportional hazards model with random effects. *Statist. 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 J. Paediatric Dentistry* **2**, 87-96.

Biostatistical Centre, Katholieke Universiteit Leuven, Kapucijnenvoer 35, B–3000, Leuven, Belgium.

E-mail: arnost.komarek@med.kuleuven.be

Biostatistical Centre, Katholieke Universiteit Leuven, Kapucijnenvoer 35, B–3000, Leuven, Belgium.

E-mail: emmanuel.lesaffre@med.kuleuven.be

(Received October 2004; accepted October 2005)