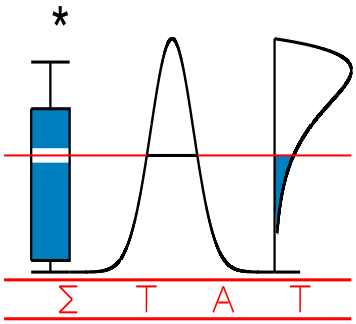

T E C H N I C A L

R E P O R T

0546

BAYESIAN ACCELERATED FAILURE TIME MODEL WITH MULTIVARIATE DOUBLY-INTERVAL-CENSORED DATA AND FLEXIBLE DISTRIBUTIONAL ASSUMPTIONS

A. KOMÁREK, E. LESAFFRE

I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

September 22, 2005

Bayesian Accelerated Failure Time Model with Multivariate Doubly-Interval-Censored Data and Flexible Distributional Assumptions

Arnošt Komárek and Emmanuel Lesaffre

Biostatistical Centre, Katholieke Universiteit Leuven,

Kapucijnenvoer 35, B–3000, Leuven, Belgium

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

emmanuel.lesaffre@med.kuleuven.be

Tel: +32-16-33 68 96;

Fax: +32-16-33 69 00

Summary: In this paper we consider the relationship of covariates to the time to caries of permanent first molars. This involves an analysis of multivariate doubly-interval-censored data. To describe this relationship we suggest an accelerated failure time model with random effects taking into account that the observations are clustered. Indeed, up to four permanent molars per child enter into the analysis implying up to four caries times for each child. Each distributional part of the model is specified in a flexible way as a penalized normal mixture with an overspecified number of mixture components. A Bayesian approach with the MCMC methodology is used to estimate the model parameters and a software package in the R language has been written.

Key words: Clustered Data; Density Smoothing; Gaussian Markov Random Field; Markov Chain Monte Carlo; Regression; Survival Data.

1 Introduction

In standard survival methods it is assumed that the time to the event is either exactly known or right-censored. However, in some areas of medical research (dentistry, HIV studies), the event can only be recorded at regular intervals (visits to the physician, dentist) which gives rise to *interval-censored data*. Moreover, often not only the failure time but also the onset time is recorded in an interval-censored manner resulting in *doubly-interval-censored data*. A typical example is the time to caries development on a tooth which is equal to the time from tooth emergence to onset of caries. When several teeth are examined jointly, one needs to take into account that there is clustering, i.e. that teeth from the same mouth are related. In this paper we aim to analyze the effect of brushing, plaque accumulation, presence of sealants and the status of the adjacent deciduous molars on the caries time of the four permanent first molars using the data from the Signal Tandmobiel® study.

De Gruttola and Lagakos (1989) suggested a non-parametric estimate of the survivor function for doubly-interval-censored data. Alternative methods were subsequently given by Bacchetti and Jewell (1991); Gómez and Lagakos (1994); Sun (1995); Gómez and Calle (1999). Further, Kim, De Gruttola, and Lagakos (1993) generalized the one-sample estimation procedure of De Gruttola and Lagakos (1989) to a Cox PH model. However, their method needs to discretize the data. Cox regression with the onset time interval-censored and the event time right-censored has been considered by Goggins, Finkelstein, and Zaslavsky (1999); Sun, Liao, and Pagano (1999); Pan (2001). To our best knowledge, regression with *multivariate* doubly-interval-censored survival data has not been discussed in the literature yet. In this paper, we propose a Bayesian method based on the accelerated failure time model. At the same time, our method aims to avoid strong parametric assumptions concerning the baseline survival time.

The following notation will be used throughout the paper. Let $\sum_{i=1}^{N} n_i$ observational units be divided into N clusters, the ith one of size n_i. Let $U_{i,l}$ and $V_{i,l}$, $i = 1, \ldots, N$, $l = 1, \ldots, n_i$ denote the true chronological onset and event times, respectively and $T_{i,l} = V_{i,l} - U_{i,l}$ the true time-to-event. Let $\boldsymbol{z}_{i,l}$ be a vector of covariates which can possibly influence the onset time $U_{i,l}$ of the(i,l)th unit and let $\boldsymbol{x}_{i,l}$ be a covariate vector having possibly an impact on the time-to-event $T_{i,l}$.

In our context, it is only known that $U_{i,l}$ occurred within an interval of time $[u_{i,l}^L,\ u_{i,l}^U]$, where $u_{i,l}^L \leq U_{i,l} \leq u_{i,l}^U$. Similarly, the event time $V_{i,l}$ is only known to lie in an interval $[v_{i,l}^L,\ v_{i,l}^U]$, with $v_{i,l}^L \leq V_{i,l} \leq v_{i,l}^U$, $i = 1, \ldots, N$, $l = 1, \ldots, n_i$, see also Figure 1. Note that exactly observed, right- and left-censored data are special cases of interval-censored data. It is assumed that the observed intervals result from an independent noninformative censoring process (e.g. pre-scheduled visits). Further, we assume that the true time-to-event $T_{i,l}$ is independent of the true onset time $U_{i,l}$ for all i and l, this issue will be discussed further in Section 3.1.

< Figure 1 about here >

In Section 2, we describe the data and the research question which motivated the development of our approach. Section 3 describes the assumed model. The prior distributions are described in Section 4. The resulting posterior distributions are given in Section 5. The suggested approach is validated using a simulation study in Section 6 and in Section 7 it is shown how it tackled our research question. The paper ends with a discussion.

2 Data and Research Question

The Signal Tandmobiel® study is a longitudinal prospective (1996–2001) oral health screening project performed in Flanders, Belgium. The children (2 315 boys and 2 153

girls) born in 1989 were examined on a yearly basis by one of 16 trained dentist-examiners. Additional data on oral hygiene and dietary habits were obtained through structured questionnaires, completed by the parents. The details of the study design and research methods can be found in Vanobbergen et al. (2000).

The primary interest of the present analysis is to address the influence of *sound* versus affected (*decayed/filled/missing due to caries*) deciduous second molars (teeth 55, 65, 75, 85, respectively in European dental notation) on the caries susceptibility of the adjacent permanent first molars (teeth 16, 26, 36, 46, respectively). Note that for about five years the deciduous second molars are in the mouth together with the permanent first molars.

It is possible that the caries processes on the primary and the permanent molar occur simultaneously. In this case it is difficult to know whether caries on the deciduous molar caused caries on the permanent molar or vice versa. For this reason, the permanent first molar was excluded from the analysis if caries was present when emergence was recorded. In total, 3 520 children were included in the analysis of which 187 contributed 1 tooth, 317 2 teeth, 400 3 teeth and 2 616 all 4 teeth.

Additionally, we considered the impact of gender (*boy/girl*), presence of sealants in pits and fissures of the permanent first molar (*none/present*), occlusal plaque accumulation on the permanent first molar (*none/in pits and fissures/on total surface*), and reported oral brushing habits (*not daily/daily*). Note that pits and fissures sealing is a preventive action which is expected to protect the tooth against caries development. The presence of plaque on the occlusal surfaces of the permanent first molars was assessed using a simplified version of the index described by Carvalho et al. (1989). All explanatory variables were obtained at the examination where the presence of the permanent first molar was first recorded.

The choice of explanatory variables is motivated by the results of Leroy et al. (2005) where a GEE multivariate log-logistic survival model was used for the caries times com-

bined with multiple imputation to correct for the interval-censored emergence times. Additionally, they included also the status of the deciduous first molars (teeth 54, 64, 74, 84, respectively) as a covariate in the model. We will not use this factor as an explanatory variable due to its high dependence with the status of the deciduous second molar (in all quadrants of the mouth, the χ^2 test statistics with 9 degrees of freedom exceeded 1 100).

The onset time $U_{i,l}$, $l = 1, \ldots, 4$ is the age (in years) of the ith child (ith cluster) at which the lth permanent first molar emerged. The failure time, $V_{i,l}$, indicates the onset of caries of the lth permanent first molar. The time from tooth emergence to the onset of caries, $T_{i,l}$, is doubly-interval-censored. Here, both the time of tooth emergence and the onset of caries experience are only known to lie in an interval of about 1 year.

Further, in our example about 85% of the permanent first molars had emerged at the first examination giving rise to a huge amount of left-censored onset times. However, at each examination the permanent teeth were scored according to their clinical eruption stage using a grading that starts at P0 (tooth not visible in the mouth) and ends with P4 (fully erupted tooth with full occlusion). Based on the clinical eruption stage at the moment of the first examination, all left-censored emergence times were transformed into interval-censored ones with the lower limit of the observed interval equal to the age at examination minus 0.25 year, 0.5 year and 1 year, respectively for the teeth with the eruption stage P1, P2 and P3, respectively and with the lower limit equal to 5 years for the teeth with the eruption stage P4. We refer to Leroy et al. (2005) for details and motivation.

3 Model

We allow both the true onset time $U_{i,l}$ and the true time-to-event $T_{i,l}$ to depend on covariates via the accelerated failure time (AFT). The AFT model is a valuable alternative to the Cox PH model (see, e.g., Reid, 1994; Keiding, Andersen, and Klein, 1997; Hougaard, 1999; Lambert et al., 2004) although far less used. To account for possible dependencies among teeth of the same child we use an AFT model with a random intercept included in the model and refer to it as the *cluster-specific* AFT model. Further, we have opted for a flexible expression for all distributional parts of the model implying a smoothly estimated survival and hazard curve with the shape driven by the data. To this end, we propose a penalized normal mixture with an overspecified number of mixture components.

3.1 Cluster-specific AFT model

Cluster-specific random effects d_i and b_i, $i = 1, \ldots, N$ are introduced to account for possible dependencies of different teeth within an individual. Namely, the following model is assumed

$$\log(U_{i,l}) = d_i + \boldsymbol{\delta}'\boldsymbol{z}_{i,l} + \zeta_{i,l}, \qquad i = 1, \ldots, N, \quad l = 1, \ldots, n_i, \quad (1)$$

$$\log(V_{i,l} - U_{i,l}) = \log(T_{i,l}) = b_i + \boldsymbol{\beta}'\boldsymbol{x}_{i,l} + \varepsilon_{i,l}, \qquad i = 1, \ldots, N, \quad l = 1, \ldots, n_i, \quad (2)$$

where $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ are unknown regression parameter vectors, $\zeta_{i,l}$, $i = 1, \ldots, N$, $l = 1, \ldots, n_i$ are i.i.d. random variables with some density $g_\zeta(\zeta)$. Analogously, the error terms $\varepsilon_{i,l}$, $i = 1, \ldots, N$, $l = 1, \ldots, n_i$ are i.i.d. random variables with density $g_\varepsilon(\varepsilon)$. The random effects d_i, $i = 1, \ldots, N$ and b_i, $i = 1, \ldots, N$, respectively are assumed to be i.i.d. with a density $g_d(d)$ and $g_b(b)$, respectively. Furthermore we assume that $\varepsilon_{i,l}$, $\zeta_{i,l}$, b_i and d_i are independent for all i and l.

Assumptions outlined above imply that $U_{i,l}$ and $T_{i,l}$ are independent for each i and

l. Specifically, here we assume that the caries process on a specific tooth only depends on the time when that tooth is at risk for caries and not on the chronological time when the tooth entered the risk group for caries. This assumption seems to be reasonable for the Signal Tandmobiel® data taking into account the results of Leroy et al. (2005) who evaluated also the effect of the emergence time on the time-to-caries and found it highly non-significant ($p = 0.78$). The above assumptions also imply that (a) whether a child is an early or late emerger is independent of whether a child is more or less sensitive against caries (independence of b_i and d_i) and (b) whether a specific tooth emerges early or late is independent of whether that tooth is more or less sensitive against caries (independence of $\varepsilon_{i,l}$ and $\zeta_{i,l}$).

3.2 Flexible distributional assumptions

To finalize the specification of our measurement model we have to specify the densities $g_\zeta,\ g_\varepsilon$ of the random errors and the densities $g_d,\ g_b$ of the random effects.

A generic density $g(y)$ of a random variable Y is modelled as a location-and-scale transformed weighted sum of normal densities over a *fixed fine* grid of knots $\boldsymbol{\mu} = (\mu_{-K}, \ldots, \mu_K)'$ centered around $\mu_0 = 0$. The means of the normal components are equal to the knots and their variances are all equal and fixed to σ^2, i.e.

$$Y \;=\; \alpha \,+\, \tau\, Y^*, \qquad Y^* \sim \sum_{j=-K}^{K} w_j\, \mathcal{N}\left(\mu_j,\ \sigma^2\right) \tag{3}$$

where the intercept term α and the scale parameter τ have to be estimated as well as the vector $\boldsymbol{w} = (w_{-K}, \ldots, w_K)'$ of weights that satisfy $w_j > 0$ for all j and $\sum_j w_j = 1$. Additionally, each element of $\boldsymbol{w}$ is expressed as a function of the elements of the vector $\boldsymbol{a} = (a_{-K}, \ldots, a_K)'$ as follows

$$w_j = \frac{\exp(a_j)}{\sum_{k=-K}^{K} \exp(a_k)}, \quad j = -K, \ldots, K \tag{4}$$

which allows for unconstrained estimation.

Our model can be considered as a limiting case of the B-spline smoothing (Eilers and Marx, 1996) of unknown functions. Instead of the B-splines as basis functions, the normal densities are used which are the limits of the B-splines as their degree tends to infinity (Unser, Aldroubi, and Eden, 1992). A similar approach was used by Ghidey, Lesaffre, and Eilers (2004) who used the expression similar to (3) to model the distribution of the random effects in a linear mixed model (with uncensored data). Komárek, Lesaffre, and Hilton (2005) employed this technique for the error distribution in an AFT model with univariate censored data.

The choice of the grid points μ_j and the basis standard deviation σ can be made independent of the location and the range of the true distribution of Y. In our analysis (Section 7), the same grid of equidistant knots of length 31 ($K = 15$) defined on $[-4.5,\ 4.5]$ is used with the basis standard deviation $\sigma = 2\,(\mu_j - \mu_{j-1})/3 = 0.2$; see Komárek et al. (2005) for a motivation.

3.3 Likelihood

Denoting p a generic density, the likelihood contribution of the ith cluster is given by

$$
\begin{aligned}
L_i &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\prod_{l=1}^{n_i}\int_{u_{i,l}^L}^{u_{i,l}^U}\int_{v_{i,l}^L-u_{i,l}}^{v_{i,l}^U-u_{i,l}} p(t_{i,l},\, b_i,\, u_{i,l},\, d_i)\,dt_{i,l}\,du_{i,l}\right\} db_i\,dd_i \qquad (5)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\prod_{l=1}^{n_i}\int_{u_{i,l}^L}^{u_{i,l}^U}\int_{v_{i,l}^L-u_{i,l}}^{v_{i,l}^U-u_{i,l}} p(t_{i,l}\,|\,b_i,\, u_{i,l},\, d_i)\,p(b_i\,|\,u_{i,l},\, d_i)\,p(u_{i,l}\,|\,d_i)\,p(d_i)\,dt_{i,l}\,du_{i,l}\right\} db_i\,dd_i\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\prod_{l=1}^{n_i}\int_{u_{i,l}^L}^{u_{i,l}^U}\left\{\int_{v_{i,l}^L-u_{i,l}}^{v_{i,l}^U-u_{i,l}} p(t_{i,l}\,|\,b_i)\,dt_{i,l}\right\}p(u_{i,l}\,|\,d_i)\,du_{i,l}\right] p(b_i)\,p(d_i)\,db_i\,dd_i,
\end{aligned}
$$

where $p(t_{i,l}\,|\,b_i) = t_{i,l}^{-1}\,g_\varepsilon\big\{\log(t_{i,l}) - b_i - \boldsymbol{\beta}'\boldsymbol{x}_{i,l}\big\}$ combines the AFT model (2) with the model of the type (3) for $g_\varepsilon(\varepsilon_{i,l})$ and similarly $p(u_{i,l}\,|\,d_i) = u_{i,l}^{-1}\,g_\zeta\big\{\log(u_{i,l}) - d_i - \boldsymbol{\delta}'\boldsymbol{z}_{i,l}\big\}$ combines the AFT model (1) with the model of the type (3) for $g_\zeta(\zeta_{i,l})$. Further, $p(b_i) = g_b(b_i)$ and $p(d_i) = g_d(d_i)$ are given by the model (3). Since it is not possible to identify between the intercept terms of the random effect and the error term, we fix the intercepts of the

random effects, i.e. the terms α^b and α^d to zero.

Ghidey et al. (2004) and Komárek et al. (2005) used penalized maximum-likelihood estimation of the unknown parameters. However, this method is rather cumbersome and computationally intractable for likelihood (5). Instead, a Bayesian approach together with the MCMC methodology (see, e.g., Besag et al., 1995) will be used here to avoid explicit integration and optimization.

4 Prior distributions

To specify the model from the Bayesian point of view, prior distributions for all unknown parameters, denoted by $\boldsymbol{\theta}$, have to be given. For our model we assume a hierarchical structure given by a directed acyclic graph (DAG) in Figure 2 with the usual convention that square boxes represent fixed or observed quantities, circles unknown parameters, solid arrows stochastic dependencies and dashed arrows deterministic dependencies. We use $p(\cdot \mid \cdot)$ to denote a generic conditional distribution and $p(\cdot)$ to denote a generic marginal distribution. The joint prior distribution of the whole model is given by the product of the conditional distributions of the nodes pertaining to unobserved quantities given their parents, namely

$$
\begin{aligned}
p(\boldsymbol{\theta}) \propto \prod_{i=1}^{N} \Bigg[\prod_{l=1}^{n_i} \Big\{ & p\big(v_{i,l} \mid u_{i,l},\, t_{i,l}\big) \times p\big(t_{i,l} \mid \boldsymbol{\beta},\, b_i,\, \varepsilon_{i,l}\big) \times p\big(u_{i,l} \mid \boldsymbol{\delta},\, d_i,\, \zeta_{i,l}\big) \times \\
& p\big(\varepsilon_{i,l} \mid \mathcal{G}_\varepsilon,\, r^{\varepsilon}_{i,l}\big) \times p\big(\zeta_{i,l} \mid \mathcal{G}_\zeta,\, r^{\zeta}_{i,l}\big) \Big\} \times p\big(b_i \mid \mathcal{G}_b,\, r^{b}_{i}\big) \times p\big(d_i \mid \mathcal{G}_d,\, r^{d}_{i}\big) \Bigg] \times \\
& p\big(\mathcal{G}_\varepsilon\big) \times p\big(\mathcal{G}_\zeta\big) \times p\big(\mathcal{G}_b\big) \times p\big(\mathcal{G}_d\big) \times p\big(\boldsymbol{\delta}\big) \times p\big(\boldsymbol{\beta}\big).
\end{aligned}
$$

The node $\mathcal{G}_\varepsilon$ refers to the set $\{\sigma^\varepsilon,\ \boldsymbol{\mu}^\varepsilon,\ \alpha^\varepsilon,\ \tau^\varepsilon,\ \boldsymbol{w}^\varepsilon,\ \boldsymbol{a}^\varepsilon,\ \lambda^\varepsilon\}$ which contains the parameters of formulas (3) and (4) and a smoothing parameter λ^ε which will be further discussed in Section 4.2. The sets $\mathcal{G}_\zeta, \mathcal{G}_b, \mathcal{G}_d$ are defined in an analogous manner. Further, let $\mathcal{G}$ be a generic symbol for its subscriped counterpart (i.e. for $\mathcal{G}_\varepsilon$, $\mathcal{G}_\zeta$, $\mathcal{G}_b$, $\mathcal{G}_d$) and let

y be a generic symbol for $\varepsilon_{i,l}$, $\zeta_{i,l}$, b_i, or d_i, $i = 1, \ldots, N$, $l = 1, \ldots, n_i$, respectively. The sub-DAG for the generic y random variable is shown in Figure 3 and the corresponding DAG conditional distributions are discussed in Sections 4.1 and 4.2.

< Figure 2 about here >

< Figure 3 about here >

4.1 DAG conditional distributions for the generic node y

To specify the generic DAG conditional distribution of y we introduce a latent *allocation variable* r taking values in $\{-K, \ldots, K\}$. Actually, data augmentation (Tanner and Wong, 1987) is introduced which simplifies the MCMC procedure. The DAG conditional distribution $p(y \,|\, \mathcal{G}, r)$ is simply a normal distribution:

$$p(y \,|\, \mathcal{G}, r) = p(y \,|\, \sigma, \boldsymbol{\mu}, \alpha, \tau, r) = \mathcal{N}\left(\alpha + \tau\, \mu_r,\ (\tau\, \sigma)^2\right).$$

Further, $p(r \,|\, \mathcal{G}) = p(r \,|\, \boldsymbol{w})$ is given by $\Pr\big(r = j \;\big|\; \boldsymbol{w}\big) = w_j$, $j \in \{-K, \ldots, K\}$ and the conditional distribution $p(y \,|\, \mathcal{G}) = p(y \,|\, \sigma, \boldsymbol{\mu}, \alpha, \tau, \boldsymbol{w})$ is a normal mixture given by the formulas (3) and (4).

4.2 DAG conditional distributions for $\mathcal{G}$

The prior distribution of a generic node $\mathcal{G}$ whose structure is given in Figure 3 equals $p(\mathcal{G}) \propto p(\boldsymbol{a} \,|\, \lambda)\, p(\lambda)\, p(\alpha)\, p(\tau)$. Although often the grid length $(2\,K + 1)$ is of moderate size it results in a rather large number of unknown $\boldsymbol{a}$ parameters. To avoid overfitting of the data and even identifiability problems, a restriction on the $\boldsymbol{a}$ parameters is needed. Komárek et al. (2005) used a penalty term for the transformed weights added to the log-likelihood for this purpose, which can be interpreted as an informative log-prior distribution (e.g., Silverman, 1985, Section 6). Therefore the prior distribution $p(\boldsymbol{a} \,|\, \lambda)$

is defined as the exponential of the penalty term of Komárek et al. (2005), i.e.

$$\begin{aligned} p(\boldsymbol{a} \mid \lambda) &\propto \exp\Big\{-\frac{\lambda}{2} \sum_{j=-K+m}^{K} \big(\Delta^m a_j\big)^2\Big\} \\ &= \exp\Big\{-\frac{\lambda}{2}\, \boldsymbol{a}'\, \mathbb{D}'\mathbb{D}\, \boldsymbol{a}\Big\}, \end{aligned} \tag{6}$$

where Δ^m denotes a difference operator of order m (e.g., $\Delta^3\, a_j = a_j - 3\, a_{j-1} + 3\, a_{j-2} - a_{j-3}$ is used in the analysis in Section 7) and $\mathbb{D}$ the corresponding difference operator matrix. The hyperparameter λ controls the smoothness of the resulting density $g(y)$.

Expression (6) is that of a multivariate normal density with zero mean and covariance matrix $\lambda^{-1}\big(\mathbb{D}'\mathbb{D}\big)^-$, where $\big(\mathbb{D}'\mathbb{D}\big)^-$ denotes a generalized inverse of the matrix $\mathbb{D}'\mathbb{D}$. This distribution is known as a Gaussian Markov random field (GMRF) extensively used in spatial statistics. Although the distribution (6) is improper (the matrix $\mathbb{D}'\mathbb{D}$ has a deficiency of m in its rank) the resulting posterior distribution is proper as soon as there is some informative data available, see Besag et al. (1995).

Prior distribution (6) is concentrated in areas where the coefficients of $\boldsymbol{a}$ that correspond to the mixture components with closer means do not substantially differ. Consequently, prior distribution (6) favours smooth estimates of the estimated densities ($g_\varepsilon, g_\zeta, g_b$ or g_d). Due to the correspondence of the prior (6) with the penalty term in the penalized maximum-likelihood approach we will call the mixture model (3) with this prior *a penalized Gaussian mixture*.

The smoothing hyperparameter λ can be interpreted as the component of prior precision of the transformed weights $\boldsymbol{a}$. In penalized maximum-likelihood estimation, the optimal value for λ is determined by cross-validation or Akaike's information criterion (see Ghidey et al., 2004; Komárek et al., 2005). Another way is to optimize the penalized likelihood also with respect to λ (see Kauermann, 2005). For our full Bayesian inference, the unknown smoothing parameter λ is considered stochastic and analogously to Kauermann (2005) is estimated simultaneously with all the remaining parameters of the model.

Therefore, here a hyperprior has been assigned to λ, i.e. a highly dispersed but proper Gamma prior, e.g., $p(\lambda) = \text{Gamma}(1,\, 0.005)$ is used in the analysis in Section 7.

Finally, in the case when the intercept term α is not fixed to zero (intercept of error distributions), a highly dispersed normal distribution has been taken for $p(\alpha)$. For the precision τ^{-2} we have taken a highly dispersed Gamma distribution. In the analysis in Section 7, $p(\alpha) = \mathcal{N}(0,\, 100)$, $p(\tau^{-2}) = \text{Gamma}(1,\, 0.005)$. Alternatively a uniform distribution on τ which is sometimes preferred for hierarchical models (Gelman et al., 2004, pp. 136, 390) could be taken.

4.3 DAG conditional distributions for the time variables and regression parameters

The DAG conditional distributions for the time variables are determined by the AFT expressions from Section 3.1. Namely, $p\big(u_{i,l}\big|\boldsymbol{\delta},\, d_i,\, \zeta_{i,l}\big)$, $p\big(t_{i,l}\big|\boldsymbol{\beta},\, b_i,\, \varepsilon_{i,l}\big)$, and $p\big(v_{i,l}\big|u_{i,l},\, t_{i,l}\big)$ are Dirac densities with $p\big(u_{i,l}\big|\boldsymbol{\delta},\, d_i,\, \zeta_{i,l}\big) = I\big[u_{i,l} = \exp(d_i + \boldsymbol{\delta}'\boldsymbol{z}_{i,l} + \zeta_{i,l})\big]$, $p\big(t_{i,l}\big|\boldsymbol{\beta},\, b_i,\, \varepsilon_{i,l}\big) = I\big[t_{i,l} = \exp(b_i + \boldsymbol{\beta}'\boldsymbol{x}_{i,l} + \varepsilon_{i,l})\big]$, and $p\big(v_{i,l} \,\big|\, u_{i,l},\, t_{i,l}\big) = I[v_{i,l} = u_{i,l} + t_{i,l}]$, respectively.

The conditional distributions $p\big(u^L_{i,l},\, u^U_{i,l}\big|u_{i,l},\, \text{censoring}_{i,l}\big)$ and $p\big(v^L_{i,l},\, v^U_{i,l}\big|v_{i,l},\, \text{censoring}_{i,l}\big)$ are, under the assumption of independent noninformative censoring, the Dirac densities as well. For example with interval censoring resulting from checking the survivor status at (random) times $\boldsymbol{C}_{i,l} = \{c_{i,l,0}, \ldots, c_{i,l,S+1}\}$, where $c_{i,l,0} = 0,\ c_{i,l,S+1} = \infty$ we have $p\big(u^L_{i,l} = c_{i,l,s},\, u^U_{i,l} = c_{i,l,s+1} \,\big|\, u_{i,l},\, \boldsymbol{C}_{i,l}\big) = I[c_{i,l,s} \le u_{i,l} \le c_{i,l,s+1}]$, $s = 0, \ldots, S$. With standard right-censoring driven by the (random) censoring time $C_{i,l}$ we have $p\big(u^L_{i,l} = u^U_{i,l} = u_{i,l}\big|u_{i,l},\, C_{i,l} = c_{i,l}\big) = I[u_{i,l} \le c_{i,l}]$ and $p\big(u^L_{i,l} = c_{i,l},\, u^U_{i,l} = \infty\big|u_{i,l},\, C_{i,l} = c_{i,l}\big) = I[u_{i,l} > c_{i,l}]$. Note however that, as soon as the censoring mechanism is independent and noninformative, we do not have to specify its model since it only acts as a multiplicative constant in the posterior.

Finally, the prior distributions of the regression parameters, $p(\boldsymbol{\beta})$ and $p(\boldsymbol{\delta})$, are taken to be products of independent highly dispersed normal distributions ($\mathcal{N}(0,\, 100)$ was used in the analysis in Section 7).

5 Posterior distribution

The joint posterior distribution, $p(\boldsymbol{\theta} \,|\, \text{data})$, is proportional to the product of all DAG conditional distributions, i.e.

$$p(\boldsymbol{\theta} \,|\, \text{data}) \propto p(\boldsymbol{\theta}) \times \prod_{i=1}^{N} \prod_{l=1}^{n_i} \Big\{ p\big(u_{i,l}^L,\, u_{i,l}^U \big|\, u_{i,l},\, \text{censoring}_{i,l}\big) \times p\big(v_{i,l}^L,\, v_{i,l}^U \big|\, v_{i,l},\, \text{censoring}_{i,l}\big) \Big\}.$$

Inference will be based on a sample from the posterior distribution obtained using the MCMC methodology. We derived full conditional distributions for all model parameters and use the Gibbs algorithm (Geman and Geman, 1984) to perform one iteration of the MCMC.

A software package, called `bayesSurv`, was written combining the R language (R Development Core Team, 2005) with programs in C++, and is available from the *Comprehensive R Archive Network* on `http://www.R-project.org`. Specifically, MCMC sampling is performed by the function `bayessurvreg3`, the estimates of the densities g_ζ, g_ε, g_d, $g(b)$ are obtained using the function `bayesGspline` and predictive survivor or hazard functions for specified combinations of covariates are computed using the function `predictive2`.

5.1 Full conditionals for parameters defining the penalized Gaussian mixture

The full conditionals for the mixture intercept α is a normal distribution and for the smoothing parameter λ a gamma distribution is obtained. However, the full condition-

als for the precision τ^{-2} and the elements of $\boldsymbol{a}$ do not belong to any of the standard distributions.

The full conditional of each element of $\boldsymbol{a}$ is given by

$$p(a_j \mid \cdots) \propto \frac{\exp(N_j\, a_j)}{\Big\{\sum_{k=-K}^{K} \exp(a_k)\Big\}^n} \times \exp\Bigg[-\frac{\Big\{a_j - \mathrm{E}\big(a_j \mid \boldsymbol{a}_{-(j)},\, \lambda\big)\Big\}^2}{2\ \mathrm{var}\big(a_j \mid \boldsymbol{a}_{-(j)},\, \lambda\big)}\Bigg], \tag{7}$$

where (a) in case the $\boldsymbol{a}$ coefficients determine the distribution of the error terms $\zeta_{i,l}$ or $\varepsilon_{i,l}$, respectively, n is equal to $\sum_{i=1}^{N} n_i$, the total sample size and N_j is the number of residuals $\zeta_{i,l}$ or $\varepsilon_{i,l}$, respectively for which the latent allocation variable $r_{i,l}^{\zeta}$ or $r_{i,l}^{\varepsilon}$, respectively is equal to j; (b) in case the $\boldsymbol{a}$ coefficients determine the distribution of the random effects d_i or b_i, respectively, n is equal to N, the number of clusters and N_j is the number of random effects d_i or b_i, respectively for which the latent allocation variable r_i^d or r_i^b, respectively is equal to j. Finally, $\mathrm{E}\big(a_j \mid \boldsymbol{a}_{-(j)},\, \lambda\big)$ and $\mathrm{var}\big(a_j \mid \boldsymbol{a}_{-(j)},\, \lambda\big)$ are the mean and the variance resulting from the GMRF prior (6). Distribution (7) is log-concave so we experimented both with the slice sampler of Neal (2003) as well as with the adaptive rejection sampling (ARS) method of Gilks and Wild (1992) to update the elements of $\boldsymbol{a}$. However, in our applications no method was found to be superior with respect to the performance of the MCMC. The results presented in Section 7 were obtained using slice sampling.

Suppose the prior of τ^{-2} is Gamma$(\xi_1^*,\, \xi_2^*)$, then the full conditional distribution of τ^{-2} has the form

$$p(\tau^{-2} \mid \cdots) \propto (\tau^{-2})^{\xi_1 - 1} \exp\Big(\xi_3 \sqrt{\tau^{-2}} - \xi_2 \tau^{-2}\Big), \tag{8}$$

with

$$\xi_1 = \xi_1^* + 0.5\, n, \qquad \xi_2 = \xi_2^* + 0.5\, \sigma^{-2} \sum_{i^*=1}^{n} (y_{i^*} - \alpha)^2, \qquad \xi_3 = \sigma^{-2} \sum_{i^*=1}^{n} \mu_{r_{i^*}} (y_{i^*} - \alpha),$$

where i^* goes either from 1 to N or from (1, 1) to (N, n_N). Distribution (8) is generally

not log-concave though it can easily be shown that it is always unimodal. For this reason we used slice sampling to update this parameter in an MCMC run.

5.2 Full conditionals for the remaining parameters

The full conditional distribution for the latent allocation variables $r^{\zeta}_{i,l}$, $r^{\varepsilon}_{i,l}$, r^{d}_{i} and r^{b}_{i}, respectively, $i = 1, \ldots, N$, $l = 1, \ldots, n_i$ is multinomial and thus is easily sampled from. Also the regression parameters $\boldsymbol{\beta}$, $\boldsymbol{\delta}$ and the random effects b_i and d_i can easily be updated since their full conditional is normal. The true onset times $u_{i,l}$, event times $v_{i,l}$ and times-to-event $t_{i,l}$ are all non-stochastically determined by the remaining variables when conditioned on them. Finally, the full conditional distributions for the residuals $\varepsilon_{i,l}$ and $\zeta_{i,l}$ are simply truncated normals. Namely, $p(\varepsilon_{i,l} \mid \cdots)$ is $\mathcal{N}\big(\alpha^{\varepsilon} + \tau^{\varepsilon}\,\mu^{\varepsilon}_{r^{\varepsilon}_{i,l}},\ (\tau^{\varepsilon}\,\sigma^{\varepsilon})^2\big)$ truncated on the interval $[\log(v^{U}_{i,l} - u_{i,l}) - b_i - \boldsymbol{\beta}'\boldsymbol{x}_{i,l},\ \log(v^{L}_{i,l} - u_{i,l}) - b_i - \boldsymbol{\beta}'\boldsymbol{x}_{i,l}]$, $i = 1, \ldots, N$, $l = 1, \ldots, n_i$ and analogously $p(\zeta_{i,l} \mid \cdots)$.

5.3 Posterior predictive distribution

Closely related to the posterior distribution is the posterior predictive distribution of the onset time or time-to-event for a new subject with covariate values $\boldsymbol{z}_{pred}$ and $\boldsymbol{x}_{pred}$. The posterior predictive survivor or hazard function can be computed easily from the MCMC output. For instance, the posterior predictive survivor function for the time-to-event equals

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

Further,

$$S(t \mid \boldsymbol{\theta},\, \text{data},\, \boldsymbol{x}_{pred}) = S(t \mid \boldsymbol{\theta},\, \boldsymbol{x}_{pred}) = \sum_{j=-K}^{K} w^{\varepsilon}_{j} \left[1 - \Phi\Big\{\frac{\log(t) - \alpha^{\varepsilon} - b - \boldsymbol{\beta}'\boldsymbol{x}_{pred} - \tau^{\varepsilon}\mu^{\varepsilon}_{j}}{\tau^{\varepsilon}\sigma^{\varepsilon}}\Big\}\right],$$

where Φ is a cumulative distribution function of $\mathcal{N}(0,\,1)$. Using the MCMC, this quantity is estimated by

$$\hat{S}(t \,|\, \text{data},\, \boldsymbol{x}_{pred}) = M^{-1} \sum_{m=1}^{M} S(t \,|\, \boldsymbol{\theta}^{(m)},\, \boldsymbol{x}_{pred}),$$

where $\boldsymbol{\theta}^{(m)}$ states for the values of unknown parameters sampled at the mth iteration of the MCMC consisting of a total of M iterations. All values of $\boldsymbol{\theta}^{(m)}$ are directly available, except $b^{(m)}$ which must be additionally sampled from the normal mixture given by $\mathcal{G}_b^{(m)}$. Analogously, the posterior predictive survivor function for the onset time or posterior predictive hazard functions are computed.

6 Simulation study

To validate our approach we conducted a simulation study which mimics to a certain extent the Signal Tandmobiel® data. From each of 150 clusters we simulated 4 observations. The onset time $U_{i,l}$ and the time-to-event $T_{i,l}$, $i = 1, \ldots, 150$, $l = 1, \ldots, 4$ were generated according to the AFT models (1) and (2) with $\boldsymbol{z}_{i,l} = (z_{i,l,1},\, z_{i,l,2})'$, $\boldsymbol{\delta} = (0.20,\, -0.10)'$ and $\boldsymbol{x}_{i,l} = (x_{i,l,1},\, x_{i,l,2})'$, $\boldsymbol{\beta} = (0.30,\, -0.15)'$. The covariates $z_{i,l,1}$ and $x_{i,l,1}$ are continuous and generated independently from a uniform distribution on (0, 1), the covariates $z_{i,l,2}$ and $x_{i,l,2}$ are binary with the equal probabilities for zeros and ones.

The error terms $\zeta_{i,l}$ and $\varepsilon_{i,l}$ are obtained from $\zeta_{i,l} = \alpha^{\zeta} + \tau^{\zeta}\, \zeta^*_{i,l}$ ($\alpha^{\zeta} = 1.75$, $\zeta^*_{i,l} \sim g^*_{\zeta}$) and $\varepsilon_{i,l} = \alpha^{\varepsilon} + \tau^{\varepsilon}\, \varepsilon^*_{i,l}$ ($\alpha^{\varepsilon} = 2.00$, $\varepsilon^*_{i,l} \sim g^*_{\varepsilon}$), respectively. Further, the random effects d_i and b_i are obtained from $d_i = \tau^d\, d^*_i$ ($d^*_i \sim g^*_d$) and $b_i = \tau^b\, b^*_i$ ($b^*_i \sim g^*_b$), respectively. The scale parameters were chosen such that $(\tau^d)^2 + (\tau^{\zeta})^2 = \tau^2_{onset} = 0.1$ and $(\tau^b)^2 + (\tau^{\varepsilon})^2 = \tau^2_{event} = 1.0$, see below for the individual values. The choice of τ^2_{onset} and τ^2_{event} was motivated by the results of the analysis in Section 7.

Two scenarios for the distributional parts of the model were considered. In scenario I, both densities g^*_{ζ} and g^*_{ε} (of the error terms) are a mixture of normals, i.e. equal to

$0.4\mathcal{N}(-2.000,\, 0.25) + 0.6\mathcal{N}(1.333,\, 0.36)$ standardized to have unit variance. For the densities g_{d*} and g_b^* (of the standardized random effects) the density of a standardized extreme value of minimum distribution was taken. In scenario II, we reversed the setting, i.e. we have taken an extreme value distribution for the error terms and a normal mixture for the random effects. Additionally, within each scenario, the variances τ^2_{onset} and τ^2_{event} were decomposed such that the ratios $\tau^d/\tau^\zeta = \tau^b/\tau^\varepsilon$ were equal to 5, 3, 2, 1, 1/2, 1/3, and 1/5, respectively.

The true onset and event times were interval-censored by simulating the 'visit' times for each subject in the data set. The first visit was drawn from $\mathcal{N}(1,\, 0.2^2)$. Each of the distances between the consecutive visits was drawn from $\mathcal{N}(0.5,\, 0.05^2)$.

Table 1 gives the results for the regression parameters and shows that they are estimated practically unbiasedly and with a reasonable precision. It is further seen that the precision of the estimation decreases when the within-cluster variability (variance of the error terms) increases compared to the between-cluster variability (variance of the random effects). In practice however, the between-cluster variability is often much higher than the within-cluster variability. Furthermore, the shape of the survivor curves is correctly estimated as is illustrated in Figure 4 which shows results for the fitted survivor functions of the time-to-event $T_{i,l}$ and selected simulation patterns and combinations of covariates (results for the other simulation patterns or for the onset time $U_{i,l}$ were similar).

< Table 1 about here >

< Figure 4 about here >

7 Analysis of the Signal Tandmobiel® data

The analysis starts with the **Basic Model** where we allowed for a different effect of the covariates on both emergence and caries experience for the four permanent first molars. Namely, the Basic Model was based on the AFT models (1) and (2) with the covariate vector $\boldsymbol{z}_{i,l}$ for emergence composed of gender, three dummies for tooth and interaction terms between gender and tooth. The covariate vector $\boldsymbol{x}_{i,l}$ for the caries part of the model was equal to the covariate vector $\boldsymbol{z}_{i,l}$ extended by three dummy variables expressing the status of the adjacent deciduous second molar: *decayed, filled, missing due to caries* with *sound* being the baseline, by two binary covariates brushing ($1 = daily$, $0 = not\ daily$) and sealants ($1 = present$, $0 = not\ present$) and by two dummy variables: *pits and fissures, total surface* for plaque with *no plaque* as the baseline. Additionally, two-way interaction terms between tooth and all remaining factors included in the model were involved. For a better fit we subtracted 5 years (which is clinically minimal emergence time for the permanent first molars, see, e.g. Ekstrand, Christiansen, and Christiansen, 2003) from all observed times, i.e. $\log(U_{i,l} - 5)$ was used in the left-hand side of the model formula (1). The Basic Model corresponds, for comparison purposes, as closely as possible to the model used by Leroy et al. (2005). The differences were outlined in Section 2. The most important one is the use here of the flexible and cluster-specific (conditional) model fitted in the Bayesian way, whereas in Leroy et al. (2005) a parametric and population-averaged (marginal) model fitted using a frequentist method.

Based on the results for the Basic Model (see below) we fitted the **Final Model** where we omitted all two-way interactions with the covariate tooth and additionally, we binarized the covariates status and plaque. More specifically, for the covariate status we put together the groups of *decayed, filled* and *missing* (*dmf*) deciduous molars, and for the covariate plaque we joined the groups with the plaque present *in pits and fissures*

and *on total surface*.

7.1 Results

For each considered model we ran 500 000 iterations with 1:3 thinning which took about 44 hours on a 3 GHz Pentium IV PC with 1 GB RAM. We kept last 100 000 iterations for the inference.

In the **Basic Model**, we found out that all interaction terms with tooth are redundant implying that the effect of all considered covariates is the same for all four permanent first molars. To evaluate this we used simultaneous Bayesian p-values computed using the method of Held (2004). For emergence, the p-value for tooth:gender interaction was higher than 0.5. For the caries part of the model the p-values were higher than 0.5 for the interactions of tooth with gender and plaque and higher than 0.1 for the interactions with brushing, sealants and status. It seemed that also the effect of the covariate tooth itself is not significant however we kept it in the model as the question was also whether the emergence and caries timing are the same for the four permanent first molars.

Further, for none of the four permanent first molars a significant difference was found between the status groups *decayed, filled* or *missing*, and between the plaque groups *present in pits and fissures* or *present on total surface*. This finding, together with the fact that the group with *extracted* deciduous molar and the group with the plaque *present on total surface* had very low prevalence (1.45% and 3.13%, respectively) led to the simplification of these two covariates in the Final Model.

Table 2 shows posterior means, 95% equal-tail credibility intervals and Bayesian two-sided p-values for the regression parameters in the **Final Model**. It is seen that neither for the emergence and nor for the caries process there is a significant difference between the four permanent first molars. However, the molars of girls emerge significantly earlier than those of boys. With respect to caries experience, the difference between boys and

girls is not significant at 5%, however all remaining covariates have a significant impact on the caries process. Namely, daily brushing increases the time to caries with a factor of $\exp(0.337) = 1.40$ compared to less frequent brushing. Presence of sealants increases the time to caries with a factor of $\exp(0.118) = 1.13$. On the other hand, the presence of the plaque decreases the time to caries with a factor of $\exp(-0.115) = 0.89$ and the fact that the neighboring deciduous second molar is either decayed or filled or extracted due to caries decreases the time time to caries with a factor of $\exp(-0.140) = 0.87$.

The results for the regression parameters of the caries part of the model correspond quite closely to the earlier findings of Leroy et al. (2005) where, however, no attempts were done to simplify the model and where a marginal (population-averaged) specification of the model was used. Nevertheless, our results largely confirmed their findings. Namely, they found the overall effect (on all four teeth) of all factors except gender to be significant with p-value < 0.001. For the effect of gender they observed a p-value of 0.060 compared to 0.085 found by us.

< Table 2 about here >

Figure 5 shows posterior predictive survivor and hazard functions for the time-to-caries on the upper right permanent first molar of boys and 'the best', 'the worst' and two intermediate combinations of covariates (the curves for the remaining teeth and girls are similar). It is seen that when the teeth are daily brushed, plaque-free and sealed the hazard for caries starts to increase approximately 1 year after emergence however then remains almost constant. Whereas, when the teeth are not brushed daily and are exposed to other risk factors the hazard starts to increase already approximately 6 months after emergence. After a period of constant risk then the hazard starts to increase again.

The peak in the hazard for caries approximately 1 year after emergence was observed also by Leroy et al. (2005) and can be explained by the fact that teeth are most vulnerable

for caries soon after the emergence when the enamel is not yet fully developed. This peak is also present, although with a different size and with a slight shift, for all covariate combinations. On the other hand, for covariate combinations reflecting good oral health and hygiene habits, the hazard remains almost constant after the initial period of highly increasing risk whereas for combinations of covariates reflecting bad oral conditions the hazard starts to increase again approximately 3 years after emergence. This shows clearly the relationship between caries experience and oral health and hygiene habits. Due to the fact that Leroy et al. (2005) used a parametric log-logistic AFT model, they could not reveal the second period of increased hazard found here.

< `Figure 5 about here` >

Description of the performed analyses using the R package `bayesSurv` can be found in the documentation of the package.

8 Discussion

A semiparametric method to perform a regression analysis with clustered doubly-interval-censored data was suggested in this paper. We opted for a fully Bayesian approach and MCMC methodology. Note however, that the Bayesian approach is used only for technical convenience to avoid difficult optimization unavoidable with more classical maximum-likelihood based estimation. Remember that we use a penalty-like prior distribution for the transformed mixture weights $\boldsymbol{a}$ and vague priors for all remaining parameters. We did not make any attempt to use any prior information although it could have been utilized. Taking into account the above reasoning, we conclude that similar results would have been obtained if the penalized maximum-likelihood estimation had been used.

Owing to flexible distributional assumptions it was not here necessary to perform the classical checks for correct distributional specification. Clearly, this step cannot

be avoided when using fully parametric methods. However, for censored, or let alone doubly-interval-censored data, this is far from trivial. As was illustrated in Section 7 new important findings concerning the distribution of the event time, derived e.g. from the shape of the hazard function, can be discovered when leaving conventional parametric assumptions.

Finally, we have to admit that some covariates used in our dental application should actually be treated as time-dependent. Unfortunately, with our and any other method where the distribution of the event time is specified using a density and not using an instantaneous quantity like the hazard function, inclusion of time-dependent covariates is difficult.

Acknowledgments

The research was primarily supported by Research Grant 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 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

Bacchetti, P. and Jewell, N. P. (1991). Nonparametric estimation of the incubation period of AIDS based on a prevalent cohort with unknown infection times. *Biometrics*, **47**, 947–960.

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

Carvalho, J. C., Ekstrand, K. R., and Thylstrup, A. (1989). Dental plaque and caries on occlusal surfaces of first permanent molars in relation to stage of eruption. *Journal of Dental Research*, **68**, 773–779.

De Gruttola, V. and Lagakos, S. W. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics*, **45**, 1–11.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties (with Discussion). *Statistical Science*, **11**, 89–121.

Ekstrand, K. R., Christiansen, J., and Christiansen, M. E. (2003). Time and duration of eruption of first and second permanent molars: a longitudinal investigation. *Community Dentistry and Oral Epidemiology*, **31**, 344–350.

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

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.

Ghidey, W., Lesaffre, E., and Eilers, P. (2004). Smooth random effects distribution in a linear mixed model. *Biometrics*, **60**, 945–953.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, **41**, 337–348.

Goggins, W. B., Finkelstein, D. M., and Zaslavsky, A. M. (1999). Applying the Cox proportional hazards model for analysis of latency data with interval censoring. *Statistics in Medicine*, **18**, 2737–2747.

Gómez, G. and Calle, M. L. (1999). Non-parametric estimation with doubly censored data. *Journal of Applied Statistics*, **26**, 45–58.

Gómez, G. and Lagakos, S. W. (1994). Estimation of the infection time and latency distribution of AIDS with doubly censored data. *Biometrics*, **50**, 204–212.

Held, L. (2004). Simultaneous inference in risk assessment; A Bayesian perspective. In Antoch, J., editor, *Compstat 2004: Proceedings in Computational Statistics*, pages 213–222, Heidelberg, 2004. Physica-Verlag.

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

Kauermann, G. (2005). A note on smoothing parameter selection for penalised spline smoothing. *Journal of Statistical Planning and Inference*, **127**, 53–69.

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.

Kim, M. Y., De Gruttola, V. G., and Lagakos, S. W. (1993). Analyzing doubly censored data with covariates, with application to AIDS. *Biometrics*, **49**, 13–22.

Komárek, A., Lesaffre, E., and Hilton, J. F. (2005). Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *Journal of Computational and Graphical Statistics*, **14**, 726–745.

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*, **23**, 3177–3192.

Leroy, R., Bogaerts, K., Lesaffre, E., and Declerck, D. (2005). Effect of caries experience in primary molars on cavity formation in the adjacent permanent first molar. *Caries Research*, **39**, 342–349.

Neal, R. M. (2003). Slice sampling (with Discussion). *The Annals of Statistics*, **31**, 705–767.

Pan, W. (2001). A multiple imputation approach to regression analysis for doubly censored data with application to AIDS studies. *Biometrics*, **57**, 1245–1250.

R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

Reid, N. (1994). A conversation with Sir David Cox. *Statistical Science*, **9**, 439–455.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non–parametric regression curve fitting. *Journal of the Royal Statistical Society, Series B*, **47**, 1–52.

Sun, J. (1995). Empirical estimation of a distribution function with truncated and doubly interval-censored data and its application to AIDS studies. *Biometrics*, **51**, 1096–1104.

Sun, J., Liao, Q., and Pagano, M. (1999). Regression analysis of doubly censored failure time data with application to AIDS studies. *Biometrics*, **55**, 909–914.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**, 528–550.

Unser, M., Aldroubi, A., and Eden, M. (1992). On the asymptotic convergence of B-spline wavelets to Gabor functions. *IEEE Transactions on Information Theory*, **38**, 864–872.

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.

$\tau^d/\tau^\zeta = \tau^b/\tau^\varepsilon$	$\delta_1 = 0.20$	$\delta_2 = -0.10$	$\beta_1 = 0.30$	$\beta_2 = -0.15$
	Scenario I			
5	0.1995 (0.56)	−0.1008 (0.17)	0.3020 (2.12)	−0.1493 (0.64)
3	0.2001 (0.56)	−0.0998 (0.27)	0.3138 (12.51)	−0.1491 (3.20)
2	0.1976 (1.30)	−0.1000 (0.37)	0.2982 (30.55)	−0.1504 (11.75)
1	0.1988 (1.84)	−0.0998 (0.76)	0.3043 (29.04)	−0.1478 (7.55)
1/2	0.1996 (3.14)	−0.1000 (0.92)	0.3015 (18.07)	−0.1475 (9.66)
1/3	0.2010 (3.74)	−0.1006 (1.02)	0.3111 (34.67)	−0.1498 (11.88)
1/5	0.1997 (3.35)	−0.1017 (1.14)	0.3036 (33.00)	−0.1493 (9.03)
	Scenario II			
5	0.1996 (0.93)	−0.1005 (0.30)	0.2983 (9.40)	−0.1477 (2.74)
3	0.2008 (2.10)	−0.1013 (0.76)	0.2950 (22.85)	−0.1526 (7.55)
2	0.2003 (3.44)	−0.0990 (1.27)	0.3060 (42.02)	−0.1458 (13.01)
1	0.1963 (8.73)	−0.0991 (3.45)	0.2988 (105.54)	−0.1487 (32.60)
1/2	0.1945 (14.46)	−0.0973 (6.30)	0.3035 (144.59)	−0.1507 (48.40)
1/3	0.2010 (16.73)	−0.0986 (5.90)	0.2963 (157.36)	−0.1456 (50.06)
1/5	0.2029 (18.12)	−0.1001 (4.12)	0.3082 (125.51)	−0.1421 (42.10)

Table 1: Simulation study. Results for the regression parameters. Mean of the estimates over the simulations and MSE ($\times 10^{-4}$).

	Emergence		Caries	
Parameter	Poster. mean	95% CR	Poster. mean	95% CR
Tooth	$p > 0.5$		$p > 0.5$	
tooth 26	−0.003	(−0.013, 0.007)	−0.006	(−0.045, 0.031)
tooth 36	0.001	(−0.008, 0.011)	−0.009	(−0.051, 0.034)
tooth 46	0.002	(−0.008, 0.012)	−0.016	(−0.059, 0.026)
Gender	$p = 0.008$		$p = 0.085$	
girl	−0.023	(−0.039, −0.007)	−0.071	(−0.155, 0.009)
Brushing			$p < 0.001$	
daily			0.337	(0.233, 0.436)
Sealants			$p < 0.001$	
present			0.118	(0.060, 0.178)
Plaque			$p < 0.001$	
present			−0.115	(−0.171, −0.067)
Status			$p < 0.001$	
dmf			−0.140	(−0.193, −0.091)

Table 2: Signal Tandmobiel® data. Posterior means, 95% equal-tail credible regions (CR) and Bayesian two-sided p-values. For the covariate tooth both the CR and the p-value are simultaneous.

Figure 1: Scheme of doubly interval censoring.

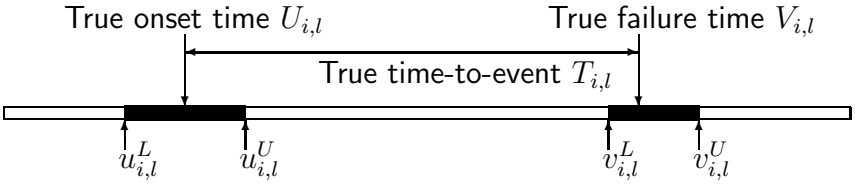

Figure 2: Directed acyclic graph for the cluster-specific AFT model.

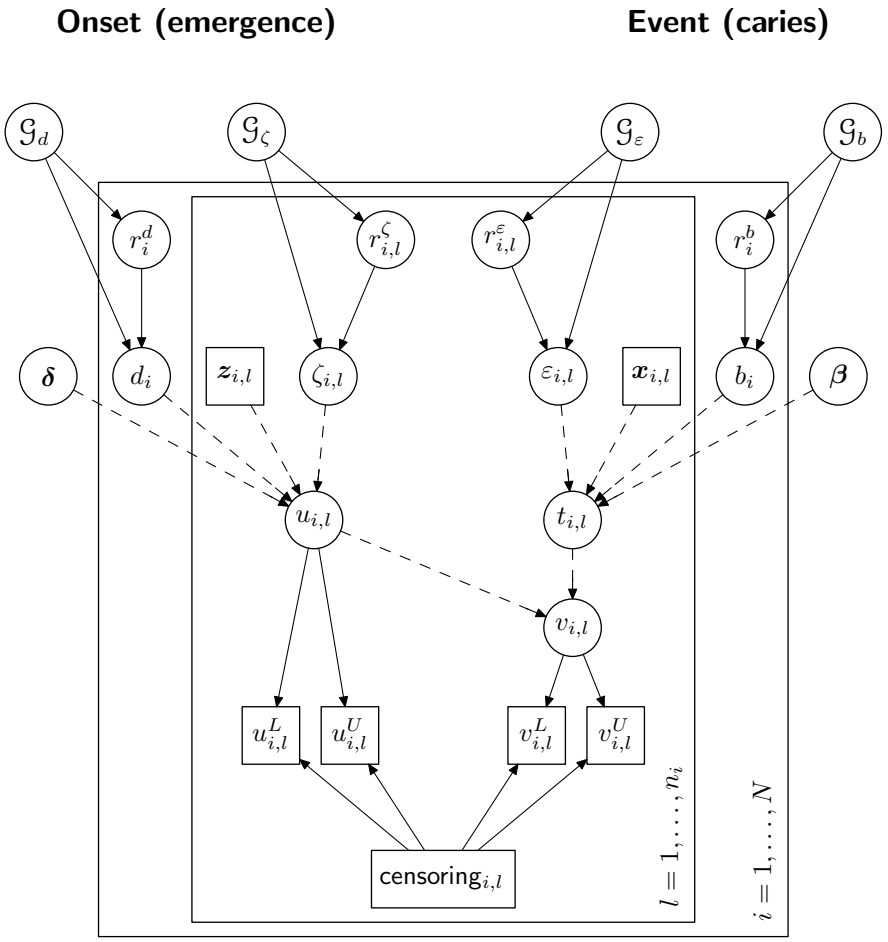

Figure 3: Directed acyclic graph for the penalized Gaussian mixture.

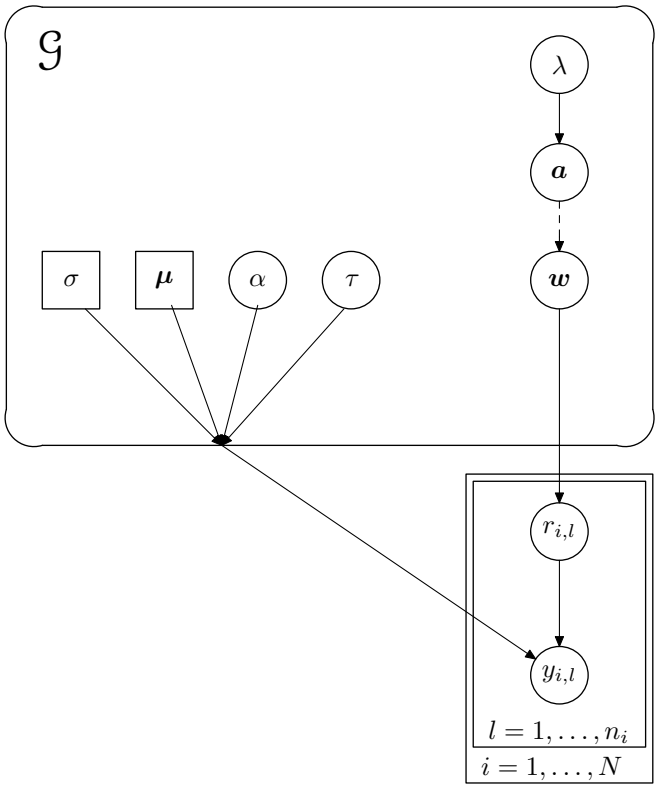

Figure 4: Simulation study. Results for the survivor functions of the time-to-event part of the model for the combination of covariates $\boldsymbol{x}_{i,l} = (0.5,\ 1)'$. Solid line: pointwise average over the predictive survivor functions at each simulation, dashed line: true survivor function (often superimposed by the solid line), grey lines: simulation based pointwise equal-tail 95% confidence interval. Scenario I is found in the left part, scenario II in the right part of the figure.

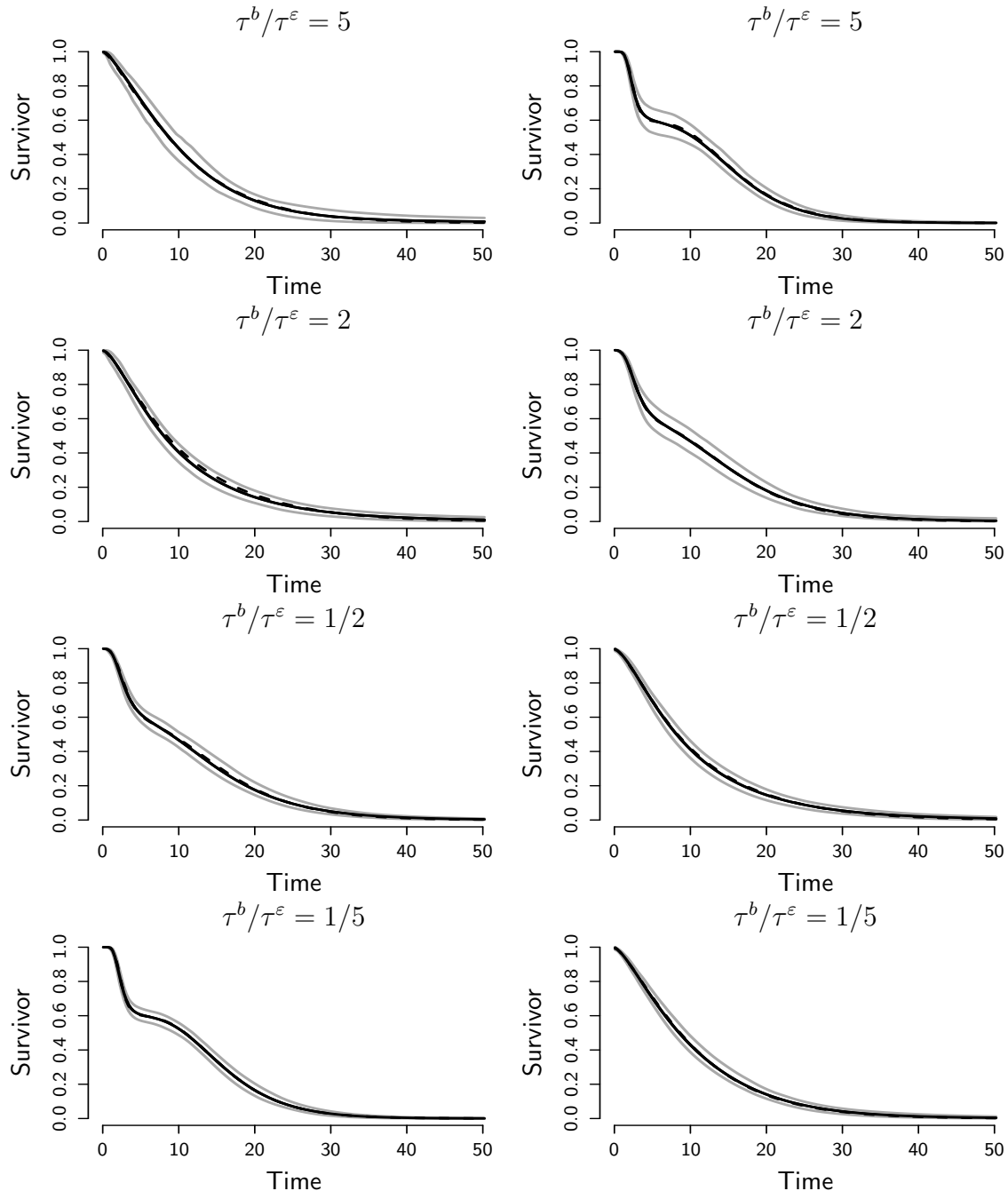

Figure 5: Signal Tandmobiel® data. Posterior predictive caries free (survivor) and caries hazard curves for tooth 16 of boys and the following combinations of covariates: solid and dashed lines for no plaque, present sealing, daily brushing and sound primary second molar (solid line) or dmf primary second molar (dashed line), dotted and dotted-dashed lines for present plaque, no sealing, not daily brushing and sound primary second molar (dotted line) or dmf primary second molar (dotted-dashed line).