BAYESIAN SEMIPARAMETRIC

ACCELERATED FAILURE TIME MODEL

FOR PAIRED DOUBLY-INTERVAL-CENSORED DATA

A. KOMÁREK, E. LESAFFRE

Bayesian Semiparametric Accelerated Failure Time Model for Paired Doubly-Interval-Censored Data

ARNOŠT KOMÁREK and EMMANUEL LESAFFRE

Biostatistical Centre, Katholieke Universiteit Leuven,

Kapucijnenvoer 35, B–3000, Leuven, Belgium

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

Tel: +32-16-33 68 96; Fax: +32-16-33 69 00

Summary: In this paper we propose a methodology to evaluate (a) the effect of covariates on paired responses which are doubly-interval-censored and (b) the association between the two responses of the pair. Our methodology tackles two research questions arising from the Signal Tandmobiel® project, a prospective Flemish (Belgian) longitudinal dental study. The research questions are: (1) What is the effect of several baseline covariates on the time to caries of the permanent right first molars? (2) Is the effect of the covariates the same for the upper and lower tooth? Time-to-caries is given as the difference of two interval-censored observations – caries time and emergence time, and hence is a doubly-interval censored response. The accelerated failure time model with a bivariate smooth error distribution is suggested where the error distribution is a mixture of bivariate normal components defined on a fine fixed grid. To deal with the problem of doubly censoring we use

the Bayesian methodology and Markov chain Monte Carlo sampling. For the suggested method we offer also software in the form of an R package.

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

1   Introduction

The motivating example is taken from the Signal Tandmobiel® study, an oral health screening project performed in Flanders (Belgium) in the period 1996–2001. The children (2 315 boys and 2 153 girls), born in 1989 and from randomly selected schools, were annually examined by one of 16 trained dentist-examiners. Additionally, the parents completed questionnaires concerning oral hygiene and dietary habits. Vanobbergen et al. (2000) give full details on the study design and research methods.

Our first aim (research question 1) is to evaluate the impact of several covariates on the *time-to-caries* of the permanent right first molars (teeth 16 and 46 in the Europen dental notation) which are together with the permanent left first molars the teeth most often attacked by caries during childhood. The prevalence of caries experience in the permanent dentition at the end of the Signal Tandmobiel® study, i.e. at the age of about 12 years, was 25.1% for the four first molars whereas it was at most 1.4% for the remaining permanent teeth.

It is also of interest (research question 2) to know whether the covariates have the same effect on both teeth and to evaluate the association between the times-to-caries of the two teeth. Hence the two teeth need to be modelled jointly, resulting thus in *paired (bivariate)* data. Furthermore, time-to-caries is *doubly-interval-censored* since it is the difference between the time that the tooth was scored as decayed *(caries time)* and the

time when it emerged *(emergence time)*. But both were observed in a coarsed manner, i.e. in approximately one year-intervals.

Cox's proportional hazards (PH) model (Cox, 1972) and the accelerated failure time (AFT) model (e.g., Kalbfleisch and Prentice, 2002, Section 2.3.3) are the most frequently used regression models for the analysis of univariate censored data. For the bivariate case with right-censored responses we note that Holt and Prentice (1974) extended the semiparametric Cox's PH regression. Fully parametric models with a simple binary covariate have been suggested by Clayton (1978) and Oakes (1982). An extension to allow for a general covariate vector was proposed by Huster, Brookmeyer, and Self (1989). A Bayesian approach under the assumption of an exponential distribution for the event times is given by Gustafson (1995). The AFT model with bivariate right-censored data estimated via multiple imputation is considered by Pan and Kooperberg (1999).

Though, none of the above approaches treats interval-censored or doubly-interval-censored data. Further, the parametric assumptions used by the approaches of Clayton (1978); Oakes (1982); Huster et al. (1989); Gustafson (1995) are difficult to check in practice and are too restrictive. Recently, Komárek and Lesaffre (2005) suggested a semiparametric AFT model which can be used for general clustered (not only paired) observations allowing for doubly-interval-censored data. To account for within-cluster dependencies they include univariate cluster-specific random effects in the model expression. Conditionally, given these random effects, the observations within each cluster are then assumed to be independent. Distributional parts of the model are specified as penalized *univariate* normal mixtures. They applied their model to the analysis of the caries times from the Signal Tandmobiel® project. However, the within-cluster association is treated as nuisance and, except for the estimated variance of the random effects, their model does not give a direct measure of the within-cluster association.

In this paper, we modify the method of Komárek and Lesaffre (2005) and assume *a bivariate* error distribution as a penalized bivariate normal mixture with a high number of components with equidistant means and constant variance matrices. The Bayesian approach with the MCMC methodology will be used for the inference. Our approach allows to visualize the estimated bivariate distribution and evaluate the association of paired responses.

In Section 2, the regression model of the bivariate doubly-interval-censored response on the covariates is specified. Inference based on the Bayesian paradigm is described in Section 3. The analysis of the Signal Tandmobiel® using the proposed method and the answers on the research questions outlined above are given in Section 4. In Section 5 some concluding remarks are given.

2 Model

2.1 Notation

Let $U_{i,l}$ and $V_{i,l}$, $i = 1, \ldots, N$, $l = 1, 2$ be the *onset* (emergence) time and the *event* (caries) time, respectively for the lth unit (tooth) of the ith cluster (child) in the study. Let $T_{i,l} = V_{i,l} - U_{i,l}$ denote the corresponding *time-to-event* (time-to-caries). The onset time $U_{i,l}$ is only observed in an interval $[u_{i,l}^L, u_{i,l}^U]$, i.e. $u_{i,l}^L \leq U_{i,l} \leq u_{i,l}^U$. Similarly, we only know that the event time $V_{i,l}$ lies in an interval $[v_{i,l}^L, v_{i,l}^U]$, i.e. $v_{i,l}^L \leq V_{i,l} \leq V_{i,l}^U$. We point out that unless the distribution of the onset time $U_{i,l}$ is uniform and $U_{i,l}$ and $T_{i,l}$ are independent, it is incorrect to assume that $T_{i,l}$ is interval-censored in $[v_{i,l}^L - u_{i,l}^U, v_{i,l}^U - u_{i,l}^L]$ (see De Gruttola and Lagakos, 1989).

Further, let $\boldsymbol{z}_{i,l}$ be the vector of covariates which might have an effect on the onset

4

time $U_{i,l}$ and $\boldsymbol{x}_{i,l}$ be the vector of covariates which can possibly influence the time-to-event $T_{i,l}$. Additionally, we assume that the onset time $U_{i,l}$ and the time-to-event $T_{i,l}$ are, given the covariates, for each i and l independent (see Komárek and Lesaffre, 2005 for a detailed discussion of this assumption) and that the interval censoring is independent and noninformative (e.g. pre-scheduled visits).

2.2 Bivariate AFT model for doubly-censored data

The distribution of $(U_{i,1},\ U_{i,2},\ T_{i,1},\ T_{i,2})'$, $i = 1, \ldots, N$, given the covariates, is given by the following accelerated failure time model:

$$\log(U_{i,l}) \;=\; \boldsymbol{\delta}' \boldsymbol{z}_{i,l} + \zeta_{i,l}, \qquad i = 1, \ldots, N, \quad l = 1, 2, \qquad (2.1)$$

$$\log(V_{i,l} - U_{i,l}) = \log(T_{i,l}) \;=\; \boldsymbol{\beta}' \boldsymbol{x}_{i,l} + \varepsilon_{i,l}, \qquad i = 1, \ldots, N, \quad l = 1, 2, \qquad (2.2)$$

where $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ are unknown regression parameter vectors, $\boldsymbol{\zeta}_i = (\zeta_{i,1}, \zeta_{i,2})'$, $i = 1, \ldots, N$ are i.i.d. random vectors with a bivariate density $g_\zeta(\zeta_1, \zeta_2)$ and similarly, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,l}, \varepsilon_{i,l})'$, $i = 1, \ldots, N$ i.i.d. random vectors with a bivariate density $g_\varepsilon(\varepsilon_1, \varepsilon_2)$.

2.3 Semiparametric model for a bivariate distribution

Our model for the unknown bivariate densities $g_\varepsilon(\varepsilon_1, \varepsilon_2)$ and $g_\zeta(\zeta_1, \zeta_2)$ is motivated by a penalized smoothing of unknown functions using B-splines (see, e.g., Eilers and Marx, 1996). Generally, bivariate smoothing can be conducted using the Kronecker products of B-splines. In our context, however, it is advantageous to replace the B-splines by normal densities for which the Kronecker product is simply a density of the bivariate normal distribution with zero correlation. There are two arguments favouring normal densities to B-splines here. Firstly, B-splines, in contrast to the normal densities, have a compact support which might be unnatural for modelling log-survival data. Secondly, there exist

well developed implementations for computation of normal densities, cumulative distribution functions or for the generation of normal pseudo-random numbers which is quite useful in practice. Furthermore, our approach can be viewed as the limiting case of the B-spline smoothing, since under some conditions the B-spline converges to the normal density as its degree tends to infinity (Unser, Aldroubi, and Eden, 1992).

Specifically, we express the unknown density $g_\varepsilon(\boldsymbol{\varepsilon})$ of the bivariate error term $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)'$ from the event part (2.2) of the AFT model as a location-and-scale transformed finite mixture of bivariate normal densities with zero correlation over a *fixed fine* grid with knots $\boldsymbol{\mu}^\varepsilon_{(j_1,j_2)} = (\mu^\varepsilon_{1,j_1}, \mu^\varepsilon_{2,j_2})'$, $j_1 = -K_1, \ldots, K_1$, $j_2 = -K_2, \ldots, K_2$ that are centered around zero, i.e. $\boldsymbol{\mu}^\varepsilon_{(0,0)} = (0, 0)'$. The means of the bivariate normal components are equal to the knots and their covariance matrices are all equal but fixed to $\Sigma^\varepsilon = \mathrm{diag}\left((\sigma^\varepsilon_1)^2, (\sigma^\varepsilon_2)^2\right)$. Thus,

$$\boldsymbol{\varepsilon} \;=\; \boldsymbol{\alpha}^\varepsilon + \mathrm{diag}(\boldsymbol{\tau}^\varepsilon)\,\boldsymbol{\varepsilon}^*, \qquad \boldsymbol{\varepsilon}^* \sim \sum_{j_1=-K_1}^{K_1} \sum_{j_2=-K_2}^{K_2} w^\varepsilon_{j_1 j_2}\, \mathcal{N}_2\left(\boldsymbol{\mu}^\varepsilon_{(j_1,j_2)}, \Sigma^\varepsilon\right). \quad (2.3)$$

In expression (2.3), the intercept term $\boldsymbol{\alpha}^\varepsilon = (\alpha^\varepsilon_1, \alpha^\varepsilon_2)'$ and the scale parameters vector $\boldsymbol{\tau}^\varepsilon = (\tau^\varepsilon_1, \tau^\varepsilon_2)'$ have to be estimated as well as the matrix $\mathbb{W}^\varepsilon = (w^\varepsilon_{j_1,j_2})$, $j_1 = -K_1, \ldots, K_1$, $j_2 = -K_2, \ldots, K_2$ of weights that satisfy $w^\varepsilon_{j_1,j_2} > 0$ for all j_1, j_2 and $\sum_{j_1} \sum_{j_2} w^\varepsilon_{j_1,j_2} = 1$. Note that, although the mixture components have all zero correlation, the covariance matrix of the resulting mixture is, in general, not diagonal. Finally, constrained estimation can be avoided if each element of $\mathbb{W}^\varepsilon$ is expressed as a function of the elements of the matrix $\mathbb{A}^\varepsilon = (a^\varepsilon_{j_1,j_2})$, $j_1 = -K_1, \ldots, K_1$, $j_2 = -K_2, \ldots, K_2$ as follows

$$w^\varepsilon_{j_1,j_2} = \frac{\exp(a^\varepsilon_{j_1,j_2})}{\sum_{k_1=-K_1}^{K_1} \sum_{k_2=-K_2}^{K_2} \exp(a^\varepsilon_{k_1,k_2})}. \quad (2.4)$$

In the following, let $\mathcal{G}_\varepsilon$ refer to the set $\{\Sigma^\varepsilon, \boldsymbol{\mu}^\varepsilon, \boldsymbol{\alpha}^\varepsilon, \boldsymbol{\tau}^\varepsilon, \mathbb{W}^\varepsilon, \mathbb{A}^\varepsilon, \boldsymbol{\lambda}^\varepsilon\}$ which contains the parameters defining the distribution of $\boldsymbol{\varepsilon}$ and a smoothing parameter vector $\boldsymbol{\lambda}^\varepsilon$ which we will discuss in Section 3.3.

The distribution of the bivariate error term $\boldsymbol{\zeta} = (\zeta_1, \zeta_2)'$ for the onset part (2.1) of the AFT model is specified using the parameters $\mathcal{G}_\zeta = \{\Sigma^\zeta,\ \boldsymbol{\mu}^\zeta,\ \boldsymbol{\alpha}^\zeta,\ \boldsymbol{\tau}^\zeta,\ \mathbb{W}^\zeta,\ \mathbb{A}^\zeta,\ \boldsymbol{\lambda}^\zeta\}$ in an analogous manner.

3   Inference

In another context, Ghidey, Lesaffre, and Eilers (2004) used an expression similar to (2.3) to model a density of the random intercept and slope in the linear mixed model with uncensored data. Further, Lesaffre and Bogaerts (2005) used this approach to model a density of bivariate simply-interval-censored data without covariates. In both papers, a penalized maximum likelihood method has been used to estimate unknown parameters. In our context, however, a maximum likelihood procedure is difficult and computationally almost intractable because the likelihood involves several multivariate integrals.

Let p denote a generic density. The likelihood contribution of the ith cluster (child) equals

$$
\begin{aligned}
L_i &= \int_{u_{i,1}^L}^{u_{i,1}^U} \int_{u_{i,2}^L}^{u_{i,2}^U} \int_{v_{i,1}^L - u_{i,1}}^{v_{i,1}^U - u_{i,1}} \int_{v_{i,2}^L - u_{i,2}}^{v_{i,2}^U - u_{i,2}} p(u_{i,1},\ u_{i,2},\ t_{i,1},\ t_{i,2})\, dt_{i,2}\, dt_{i,1}\, du_{i,2}\, du_{i,1} \qquad (3.1) \\
&= \int_{u_{i,1}^L}^{u_{i,1}^U} \int_{u_{i,2}^L}^{u_{i,2}^U} p(u_{i,1},\ u_{i,2}) \left\{ \int_{v_{i,1}^L - u_{i,1}}^{v_{i,1}^U - u_{i,1}} \int_{v_{i,2}^L - u_{i,2}}^{v_{i,2}^U - u_{i,2}} p(t_{i,1},\ t_{i,2})\, dt_{i,2}\, dt_{i,1} \right\} du_{i,2}\, du_{i,1},
\end{aligned}
$$

where the combination of the AFT model (2.1) with the mixture model (2.3) yields an expression for $p(u_{i,1},\ u_{i,2})$, namely $p(u_{i,1},\ u_{i,2}) = (u_{i,1}\, u_{i,2})^{-1}\, g_\zeta \{\log(u_{i,1}) - \boldsymbol{\delta}' \boldsymbol{z}_{i,1},\ \log(u_{i,2}) - \boldsymbol{\delta}' \boldsymbol{z}_{i,2}\}$ and similarly the combination of the AFT model (2.2) with the mixture model (2.3) yields an expression for $p(t_{i,1},\ t_{i,2})$, namely $p(t_{i,1},\ t_{i,2}) = (t_{i,1}\, t_{i,2})^{-1}\, g_\varepsilon \{\log(t_{i,1}) - \boldsymbol{\beta}' \boldsymbol{x}_{i,1},\ \log(t_{i,2}) - \boldsymbol{\beta}' \boldsymbol{x}_{i,2}\}$.

Here we suggest to use the Bayesian methodology employing the idea of data aug-

mentation (Tanner and Wong, 1987) and the MCMC methodology (e.g., Robert and Casella, 2004).

3.1 Bayesian specification of the model

To finalize the specification of the model from the Bayesian point of view, we have to give prior distributions to all unknown parameters. We assume the hierarchical structure represented by the directed acyclic graph (DAG) shown in Figure 1. The DAG child-parent conditional distributions and priors for the parameters residing on the top of the hierarchy are similar to these used by Komárek and Lesaffre (2005). We give a brief overview and highlight the differences with the bivariate model considered here.

3.2 DAG conditional distribution for the error terms ε_i and ζ_i

Analogously to Komárek and Lesaffre (2005) and using the idea of Bayesian data augmentation we introduce latent *allocation vectors* $\boldsymbol{r}_i^\varepsilon = (r_{i,1}^\varepsilon,\, r_{i,2}^\varepsilon)'$ that can take discrete values from $\{-K_1, \ldots, K_1\} \times \{-K_2, \ldots, K_2\}$. Their DAG conditional distribution is given by

$$\Pr(\boldsymbol{r}_i^\varepsilon = (j_1,\, j_2)\,|\,\mathbb{W}^\varepsilon) = w_{j_1, j_2}, \quad i = 1, \ldots, N,\; j_1 = -K_1, \ldots, K_1,\; j_2 = -K_2, \ldots, K_2.$$

The DAG conditional distribution of the error term $\boldsymbol{\varepsilon}_i$ is then simply bivariate normal with independent margins:

$$
\begin{aligned}
p(\boldsymbol{\varepsilon}_i\,|\,\mathcal{G}_\varepsilon,\, \boldsymbol{r}_i^\varepsilon) &= p(\boldsymbol{\varepsilon}_i\,|\,\boldsymbol{\tau}^\varepsilon,\, \boldsymbol{\alpha}^\varepsilon,\, \boldsymbol{\mu}^\varepsilon,\, \Sigma^\varepsilon,\, \boldsymbol{r}_i^\varepsilon) \\
&= \mathcal{N}_2\left(\boldsymbol{\alpha}^\varepsilon \,+\, \mathrm{diag}(\boldsymbol{\tau}^\varepsilon)\,\boldsymbol{\mu}_{\boldsymbol{r}_i^\varepsilon}^\varepsilon,\; \mathrm{diag}(\boldsymbol{\tau}^\varepsilon)\,\Sigma^\varepsilon\,\mathrm{diag}(\boldsymbol{\tau}^\varepsilon)\right), \quad i = 1, \ldots, N.
\end{aligned}
$$

Without introducing the latent allocation vectors we would have to work with $p(\boldsymbol{\varepsilon}_i\,|\,\mathcal{G}_\varepsilon) = p(\boldsymbol{\varepsilon}_i\,|\,\boldsymbol{\tau}^\varepsilon,\, \boldsymbol{\alpha}^\varepsilon,\, \boldsymbol{\mu}^\varepsilon,\, \Sigma^\varepsilon,\, \mathbb{W}_i^\varepsilon)$ which is a bivariate normal *mixture* given by (2.3).

The DAG conditional distribution for the error terms $\boldsymbol{\zeta}_i$, $i = 1, \ldots, N$ is defined in an analogous manner. In the following we omit the super- or subscripts ζ and ε when specifying the DAG conditional distributions for the parameters from $\mathcal{G}_\zeta$ and $\mathcal{G}_\varepsilon$.

3.3 DAG conditional distribution for the transformed mixture weights $\mathbb{A}$

The number of unknown elements of the matrix $\mathbb{A}$ is equal to $(2K_1 + 1) \times (2K_2 + 1)$ and is often quite high (e.g. equal to 961 in the analysis of the Signal Tandmobiel® data). With an uninformative prior for $\mathbb{A}$, this could cause overfitting of the data or identifiability problems. Since the (transformed) mixture weights correspond to spatially located normal components, a Gaussian Markov random field (GMRF) prior (see, e.g., Besag et al., 1995, Section 3), common in spatial statistics, can be exploited here. Such a prior distribution can be defined by specifying the conditional distribution of each a_{j_1,j_2} given remaining a_{k_1,k_2}, $(k_1, k_2) \neq (j_1, j_2)$, here denoted as $\mathbb{A}_{-(j_1,j_2)}$, and the hyperparameter λ that controls the smoothness. Usually, only a few neighboring coefficients are effectively used in the specification of $p(a_{j_1,j_2} \,|\, \mathbb{A}_{-(j_1,j_2)}, \lambda)$. A commonly used conditional distribution is a normal distribution with expectation and variance equal to

$$\mathrm{E}\big(a_{j_1,j_2} \,|\, \mathbb{A}_{-(j_1,j_2)}, \lambda\big) = \frac{a_{j_1-1,j_2} + a_{j_1+1,j_2} + a_{j_1,j_2-1} + a_{j_1,j_2+1}}{2} \tag{3.2}$$
$$- \frac{a_{j_1-1,j_2-1} + a_{j_1-1,j_2+1} + a_{j_1+1,j_2-1} + a_{j_1+1,j_2+1}}{4},$$
$$\mathrm{var}\big(a_{j_1,j_2} \,|\, \mathbb{A}_{-(j_1,j_2)}, \lambda\big) = \frac{1}{4\lambda},$$

respectively, based on the eight nearest neighbors and local quadratic smoothing. Note that the expectation and variance formulas have to be changed appropriately in corners and on edges.

Let $\boldsymbol{a}$ denote the matrix $\mathbb{A}$ stacked into a column vector. Using a bivariate difference

9

operator $\Delta\,a_{j_1,j_2} = a_{j_1,j_2} - a_{j_1+1,j_2} - a_{j_1,j_2+1} + a_{j_1+1,j_2+1}$, and denoting $\mathbb{D}$ the associated difference operator matrix, the joint prior of all transformed weights $\mathbb{A}$ given the smoothing hyperparameter λ can be written as

$$p(\mathbb{A} \mid \lambda) \propto \exp\left\{-\frac{\lambda}{2} \sum_{j_1=-K_1}^{K_1-1} \sum_{j_2=-K_2}^{K_2-1} \left(\Delta\,a_{j_1,j_2}\right)^2\right\} = \exp\left(-\frac{\lambda}{2}\,\boldsymbol{a}'\mathbb{D}'\mathbb{D}\boldsymbol{a}\right) \qquad (3.3)$$

which shows that the DAG conditional distribution $p(\mathbb{A} \mid \lambda)$ specified as a GMRF is multivariate normal with covariance matrix $\lambda^{-1}\left(\mathbb{D}'\mathbb{D}\right)^{-}$, where $\left(\mathbb{D}'\mathbb{D}\right)^{-}$ denotes a generalized inverse of $\mathbb{D}'\mathbb{D}$. Although this distribution is improper (the matrix $\mathbb{D}'\mathbb{D}$ has a deficiency of $2(K_1 + K_2) + 1$ in its rank) the resulting posterior distribution is proper as soon as there is some informative data available, see Besag et al. (1995).

An alternative prior, still belonging to the class of GMRF, corresponding closely to the prior for $\mathbb{A}$ used by Komárek and Lesaffre (2005) is obtained by considering a univariate difference operator for each row and each column of the matrix $\mathbb{A}$ with possibly two different smoothing hyperparameters stacked in a vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$ acting on rows and columns separately. Then

$$\begin{aligned} p(\mathbb{A} \mid \boldsymbol{\lambda}) \;&\propto\; \exp\left\{-\frac{\lambda_1}{2} \sum_{j_1=-K_1}^{K_1} \sum_{j_2=-K_2+m}^{K_2} \left(\Delta_1^m\,a_{j_1,j_2}\right)^2 - \frac{\lambda_2}{2} \sum_{j_2=-K_2}^{K_2} \sum_{j_1=-K_1+m}^{K_1} \left(\Delta_2^m\,a_{j_1,j_2}\right)^2\right\} \\ &=\; \exp\left\{-\frac{1}{2}\,\boldsymbol{a}'\left(\lambda_1\,\mathbb{D}_1'\mathbb{D}_1 + \lambda_2\,\mathbb{D}_2'\mathbb{D}_2\right)\boldsymbol{a}\right\} \qquad\qquad (3.4) \end{aligned}$$

where Δ_l^m, $l = 1, 2$ denotes a difference operator of order m for the lth dimension, e.g. $\Delta_1^3\,a_{j_1,j_2} = a_{j_1,j_2} - 3\,a_{j_1,j_2-1} + 3\,a_{j_1,j_2-2} - a_{j_1,j_2-3}$ and $\mathbb{D}_1$ and $\mathbb{D}_2$ are the corresponding difference operator matrices for each dimension. This prior distribution corresponds to a local polynomial smoothing of degree $m-1$ in each row and each column of the matrix $\mathbb{A}$. For example, the conditional (given the smoothing parameters and the neighboring transformed weights) mean and variance are given (for $m = 3$ and except on the corners

and on edges) by

$$\mathrm{E}\big(a_{j_1,j_2} \mid \mathbb{A}_{-(j_1,j_2)}, \boldsymbol{\lambda}\big) = \frac{\lambda_1 A_{j_2 \mid j_1} + \lambda_2 A_{j_1 \mid j_2}}{\lambda_1 + \lambda_2} \qquad (3.5)$$

$$\mathrm{var}\big(a_{j_1,j_2} \mid \mathbb{A}_{-(j_1,j_2)}, \boldsymbol{\lambda}\big) = \frac{1}{20(\lambda_1 + \lambda_2)},$$

where

$$A_{k \mid j} = \frac{a_{j,k-3} - 6\,a_{j,k-2} + 15\,a_{j,k-1} + 15\,a_{j,k+1} - 6\,a_{j,k+2} + a_{j,k+3}}{20}.$$

In both cases, the prior distribution puts higher probability mass in areas where spatially close coefficients of the matrix $\mathbb{A}$ do not substantially differ. In other words, a priori, we believe that the estimated densities $g_\zeta(\zeta_1, \zeta_2)$ and $g_\varepsilon(\varepsilon_1, \varepsilon_2)$ are smooth. In general, prior (3.4) leads to better fit in our context and hence is preferred.

As pointed out by Komárek and Lesaffre (2005), the GMRF prior corresponds to the penalty term when the penalized maximum-likelihood approach is used for the estimation. For this reason we call the mixture model (2.3) with one of the GMRF prior for the transformed mixture coefficients *penalized Gaussian mixture.*

The λ parameter in the prior (3.3) or the components λ_1, λ_2 of the $\boldsymbol{\lambda}$ parameter in the prior (3.4) determine, together with the fixed difference operator matrix $\mathbb{D}$, the precision of the transformed weights $\mathbb{A}$. We assign these parameters standardly used highly dispersed (but proper) Gamma priors.

3.4   Remaining DAG conditional and prior distributions

Remaining DAG conditional distributions are the same as in the univariate context used by Komárek and Lesaffre (2005). In summary, for the scale parameters τ_1^ε, τ_2^ε, τ_1^ζ, τ_2^ζ we suggest to use either the uniform prior (proposed for scale parameters in the hierarchical models by Gelman et al., 2004, pp. 136, 390) or a highly dispersed inverse-Gamma prior

for the squared scale parameters. The intercept parameters α_1^ε, α_2^ε, α_1^ζ, α_2^ζ as well as all components of the regression parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ can obtain a vague normal prior unless there is some external information available.

The DAG conditional distributions for the time nodes $u_{i,l}$, $v_{i,l}$, and $t_{i,l}$, $i = 1, \ldots, N$, $l = 1, 2$ are all Dirac (degenerated) densities given by the AFT model assumptions (2.1) and (2.2). Finally, the nodes $u_{i,l}^L$, $u_{i,l}^U$ and $v_{i,l}^L$, $v_{i,l}^U$ have, conditionally on their parents, the Dirac distribution driven by the censoring mechanism and the true onset or event time, respectively. Note however that we do not have to specify an exact form of the censoring mechanism as soon as it is noninformative and independent.

3.5 Markov chain Monte Carlo

From the DAG conditional distributions the joint posterior distribution of all model parameters can now be derived. In practice we obtain a sample from the posterior distribution using the Markov chain Monte Carlo method and base our inference on this sample. The basis for the MCMC algorithm is Gibbs sampling (Geman and Geman, 1984) using the full conditional distributions. In the situations when the full conditional distribution was not of standard form we used either slice sampling (Neal, 2003) or adaptive rejection sampling (Gilks and Wild, 1992). For most parameters the full conditionals are identical (with only a slight change in notation) to those given by Komárek and Lesaffre (2005) and we refer the reader thereinto.

Here we mention only the full conditional distribution for the transformed mixture weights which, due to the bivariate nature considered here, differs from that in Komárek

and Lesaffre (2005) and is equal to (for clarity the superscript ε and ζ are omitted):

$$p(a_{j_1,j_2} \mid \cdots) \propto \frac{\exp(N_{j_1,j_2} a_{j_1,j_2})}{\left\{ \sum\limits_{k_1=-K_1}^{K_1} \sum\limits_{k_2=-K_2}^{K_2} \exp(a_{k_1,k_2}) \right\}^N} \times \exp\left[-\frac{\left\{ a_{j_1,j_2} - \mathrm{E}\big(a_{j_1,j_2} \mid \mathbb{A}_{-(j_1,j_2)}, \boldsymbol{\lambda}\big) \right\}^2}{2 \, \mathrm{var}\big(a_{j_1,j_2} \mid \mathbb{A}_{-(j_1,j_2)}, \boldsymbol{\lambda}\big)} \right],$$

where N_{j_1,j_2} denotes the number of latent allocation vectors $\boldsymbol{r}_i$ that are equal to $(j_1, j_2)'$ and $\mathrm{E}\big(a_{j_1,j_2} \mid \mathbb{A}_{-(j_1,j_2)}, \boldsymbol{\lambda}\big)$ and $\mathrm{var}\big(a_{j_1,j_2} \mid \mathbb{A}_{-(j_1,j_2)}, \boldsymbol{\lambda}\big)$ follow from (3.2) or (3.5), respectively.

3.6   Software

To use the suggested methodology in practice we offer a set of functions which are the part of the R (R Development Core Team, 2005) package called `bayesSurv` downloadable from the *Comprehensive R Archive Network*. Specifically, the function `bayesBisurvreg` performs sampling from the posterior distribution, the function `bayesGspline` can be used to compute the estimates of the error densities g_ζ and g_ε and the function `predictive2` computes predictive survivor, hazard or density functions for specified combinations of covariates (see Section 4.4). For a summary of the model parameters and the convergence diagnostics, e.g., the R package `coda` (Plummer et al., 2005), can be used.

4   Signal Tandmobiel® data

The time-to-caries of the permanent first molars has been analyzed by Leroy et al. (2005). However, they used a parametric log-logistic AFT model for the caries time only. In their paper, doubly interval censoring was addressed by the inclusion of a covariate `emergence interval`. In this section, we will show a proper analysis of doubly-interval-censored data using the methodology described in the previous sections.

We started the analysis with the **Basic Model** which resembles closely the model

used by Leroy et al. (2005). Based on the results for the Basic Model we subsequently fitted its simplified version, referred as the **Final Model**.

In the **Basic Model**, we will firstly consider a similar set of covariates as Leroy et al. (2005), namely, the covariate vector $\boldsymbol{x}_{i,l}$ for the caries part of the model (2.2) are gender ($0 = boy$, $1 = girl$), presence of sealants which is a form of protection on the permanent first molar ($0 = absent$, $1 = present$), occlusal plaque accumulation for the permanent first molar (3 levels: *none/in pits and fissures/on total surface*), reported oral brushing habits ($0 = not\ daily$, $1 = daily$), status of the adjacent deciduous second molar (4 levels: *sound/decayed/filled/missing due to caries*). In contrast to Leroy et al. (2005) we did not use the status of the adjacent deciduous first molar as covariate due to its large dependence on the status of the adjacent deciduous second molar. Additionally, to allow for a different effect of the covariates on the upper and lower tooth their interaction term with the covariate jaw ($0 = $ lower, $1 = $ upper) has been included. Finally, the values of the explanatory variables were obtained at the examination where the presence of the permanent first molar was first recorded.

The covariate vector $\boldsymbol{z}_{i,l}$ for the emergence part of model (2.1) includes only gender and its interaction with jaw. In other words, on the log scale, we allow for a shift in the emergence distribution for the upper and lower tooth and also for boys and girls. Note that as well in the caries part as in the emergence part of the model the main effect of jaw is expressed by the intercept terms $\boldsymbol{\alpha}^{\varepsilon}$ and $\boldsymbol{\alpha}^{\zeta}$, respectively.

In the **Final Model**, we excluded all interaction terms with the covariate jaw, i.e. we assumed that the studied factors have the same effect on the emergence and caries for both the upper and lower tooth. Additionally, we binarized the covariates plaque and status such that for plaque: $0 = none$ and $1 = present$ (either in *pits and fissures* or *on total surface*) and for status: $0 = sound$, $1 = dmf$ (*decayed* or *missing due to*

caries or *filled*). Bayesian two-sided *p*-values and for factors with more than two levels simultaneous two-sided Bayesian *p*-values (Held, 2004) were used to arrive at the Final Model.

For reasons stated in Leroy et al. (2005) we excluded the permanent first molar from the analysis if it had experienced caries before the examination where its emergence was recorded. Additionally, about 85% of the permanent first molars had emerged at the first examination. This results in an equal amount of interval-censored onsets of the form $[0, u_{i,l}^U]$ with $u_{i,l}^U$ having a median value of 7.1 and quartiles equal to 6.9 and 7.4. However, at the first examination also the clinical eruption stage was graded using the scale starting with P0 (tooth not visible in the mouth) and ending with P4 (fully erupted tooth with full occlusion). Dental knowledge allowed to approximate the lower limits of the observed intervals as follows: $u_{i,l}^L = u_{i,l}^U - 0.25$ for the teeth with the eruption stage P1, $u_{i,l}^L = u_{i,l}^U - 0.5$ for the teeth with the eruption stage P2, $u_{i,l}^L = u_{i,l}^U - 1.0$ for the teeth with the eruption stage P3 and $u_{i,l}^L = 5.0$ for the teeth with the eruption stage P4. We refer to Leroy et al. (2005) for a motivation of these choices. The clinically minimal emergence time, 5.0 years, was also subtracted from all observed times, i.e. $\log(U_{i,l} - 5.0)$ was used in the left-hand side of the model formula (2.1).

4.1   Choice of prior distribution

To model the bivariate densities g_ζ and g_ε we used in both cases a grid of 31×31 ($K_1 = K_2 = 15$) knots with the distance d between the two knots in each margin equal to 0.3 and the basis standard deviations $\sigma_1^\varepsilon = \sigma_2^\varepsilon = \sigma_1^\zeta = \sigma_2^\zeta = 0.2$. The grid of knots is defined on a square $[-4.5, 4.5] \times [-4.5, 4.5]$ which covers the support of most standardized unimodal distributions (unimodality was checked after the analysis). The choice of the

basis standard deviation which is equal to $2d/3$ is motivated by its correspondence with cubic B-splines where each basis function covers the interval of the length $4d$. The same is nearly true for the basis formed of normal densities if we admit that the $\mathcal{N}(\mu, \sigma^2)$ density is practically zero outside the interval $(\mu - 3\sigma, \mu + 3\sigma)$, see also Ghidey et al. (2004).

For the transformed mixture weights $\mathbb{A}^{\varepsilon}$ and $\mathbb{A}^{\zeta}$ we used the prior (3.4) with the differences of the third order ($m = 3$). The smoothing parameters λ_1^{ε}, λ_2^{ε}, λ_1^{ζ}, λ_2^{ζ} were all assigned dispersed Gamma(1, 0.005) priors. The same priors were used also for the scale parameters τ_1^{ε}, τ_2^{ε}, τ_1^{ζ}, τ_2^{ζ}. The intercept terms α_1^{ε}, α_2^{ε}, α_1^{ζ}, α_2^{ζ} were all assigned dispersed $\mathcal{N}(0, 100)$ priors.

For each model we ran 250 000 MCMC iterations with 1:3 thinning and kept last 25 000 iterations for the inference. Sampling for each model took about 68 hours on a 3 GHz Pentium IV PC with 1 024 MB RAM.

4.2   Results for the Basic Model

Table 1 shows the posterior means, (simultaneous) 95% equal-tail credible intervals and (simultaneous) Bayesian two-sided p-values for the effect of each considered factor on emergence and caries experience, separately for the lower and the upper tooth. It is seen that the results for the lower and the upper tooth are very similar. Indeed, the interaction terms between jaw and the remaining factor variables were all non-significant at 5%, namely, the p-values were > 0.5, > 0.5, > 0.5, 0.262, > 0.5, 0.145, respectively for the interaction with gender in the emergence and the caries part of the model, and for the interaction with brushing, sealants, plaque, and status, respectively.

Additionally, we computed the (simultaneous) Bayesian two-sided p-values for the

two contrasts justifying the simplification of the covariates plaque and status for the Final Model, again separately for the lower and the upper tooth. For the status contrast *decayed* vs. *filled* vs. *missing due to caries*, the p-values were equal to 0.342 and 0.308, respectively for the upper and the lower tooth, respectively. For the plaque contrast *in pits and fissures* vs. *on total surface*, the p-values were equal to 0.262 and 0.301, respectively for the upper and the lower tooth, respectively.

4.3 Results for the Final Model

Results for the Final Model are given in Table 2. We give also the main effect of jaw now. It is seen that the lower tooth 46 emerges slightly later than the upper tooth 16. On the other hand, the emergence occurs slightly earlier for girls than for boys. However, neither the position of the tooth nor gender have a significant effect on the time to caries. The remaining factors do influence significantly the time to caries, namely, daily brushing increases this time with a factor of $\exp(0.250) = 1.28$, presence of sealants with a factor of $\exp(0.109) = 1.115$. The factor for presence of plaque is $\exp(-0.228) = 0.796$ and when the adjacent deciduous second molar was not sound the factor is $\exp(-0.482) = 0.618$.

For the Final Model, we explored the residual association (after adjustment for the effect of covariates) between the upper and lower tooth. For both the emergence and the caries processes, a very low Pearson correlation coefficient was found on the log-scale, namely, $\widehat{\text{corr}}(\zeta_1, \zeta_2) = 0.039$ (0.030, 0.051) and $\widehat{\text{corr}}(\varepsilon_1, \varepsilon_2) = 0.023$ (0.018, 0.028). However, our approach allows easily to compute other measures, like Kendall's tau using expressions as given by Lesaffre and Bogaerts (2005).

Further, Figure 2 shows the estimate of the error density $g_\varepsilon(\varepsilon_1, \varepsilon_2)$ in the caries part of the model and illustrates the smoothing nature of our approach. This figure also

reveals the low correlation between error terms for the upper and lower tooth.

4.4   Predictive survivor and hazard functions

Our approach allows to calculate predictive survivor and hazard functions for either the emergence or caries process and for a specific combination of covariates using the MCMC output. For example, the predictive hazard function $\hbar_1(t \,|\, \text{data}, \boldsymbol{x}_{pred})$ for caries experience on the upper tooth 16 and the covariate combination $\boldsymbol{x}_{pred}$ is given by the relationship

$$\hbar_1(t \,|\, \text{data}, \boldsymbol{x}_{pred}) = \int \hbar_1(t \,|\, \boldsymbol{\theta}, \text{data}, \boldsymbol{x}_{pred}) \, p(\boldsymbol{\theta} \,|\, \text{data}) \, d\boldsymbol{\theta},$$

where $\boldsymbol{\theta}$ denotes the vector of the unknown parameters of the model and $p(\boldsymbol{\theta} \,|\, \text{data})$ its posterior distribution. Further

$$\hbar_1(t \,|\, \boldsymbol{\theta}, \text{data}, \boldsymbol{x}_{pred}) = \hbar_1(t \,|\, \boldsymbol{\theta}, \boldsymbol{x}_{pred})$$

$$= \frac{t^{-1} \sum_{j_1=-K_1}^{K_1} \left(\sum_{j_2=-K_2}^{K_2} w_{j_1,j_2}^{\varepsilon} \right) \varphi\!\left[(\tau_1^{\varepsilon}\, \sigma_1^{\varepsilon})^{-1} \big\{ \log(t) - \alpha_1^{\varepsilon} - \boldsymbol{\beta}' \boldsymbol{x}_{pred} - \tau_1^{\varepsilon} \mu_{1,j_1}^{\varepsilon} \big\} \right]}{\sum_{j_1=-K_1}^{K_1} \left(\sum_{j_2=-K_2}^{K_2} w_{j_1,j_2}^{\varepsilon} \right) \left[1 - \Phi\!\left[(\tau_1^{\varepsilon}\, \sigma_1^{\varepsilon})^{-1} \big\{ \log(t) - \alpha_1^{\varepsilon} - \boldsymbol{\beta}' \boldsymbol{x}_{pred} - \tau_1^{\varepsilon} \mu_{1,j_1}^{\varepsilon} \big\} \right] \right]},$$

where φ denotes density and Φ cumulative distribution function of $\mathcal{N}(0, 1)$. The MCMC estimate of the predictive hazard function is then obtained as

$$\hat{\hbar}_1(t \,|\, \text{data}, \boldsymbol{x}_{pred}) = M^{-1} \sum_{m=1}^{M} \hbar_1(t \,|\, \boldsymbol{\theta}^{(m)}, \boldsymbol{x}_{pred}),$$

where M denotes the number of MCMC iterations and $\boldsymbol{\theta}^{(m)}$ the value of the parameter vector $\boldsymbol{\theta}$ sampled at the mth iteration. The predictive survivor function is computed analogously.

Figure 3 shows the predictive survivor and hazard functions for caries on the upper tooth 16 of boys and 'the best', 'the worst' and two intermediate combinations of covariates. Corresponding curves for the lower tooth 46 or for girls are almost the same

due to the non-significant effect of the covariates gender and jaw on the caries. For teeth that are not brushed daily and are exposed to other risk factors, a high peak in the hazard function is observed already less than 1 year after emergence. A similar peak, however shifted to right and of much lower magnitude is seen also for other covariate combinations. This finding corresponds to the fact that permanent first molars are most vulnerable by caries soon after they emerge, possibly because of not yet fully developed enamel on their surfaces.

4.5 Conclusions

It is shown that daily brushing and the presence of sealants significantly decelerate the time to caries on the right permanent first molars. On the other hand, the presence of plaque or caries on adjacent deciduous second molars accelerate this time. No significant difference has been found between boys and girls or between the upper and lower right permanent first molar with respect to the distribution of the caries times. Also a very low association described by a residual correlation coefficient was observed between the caries times on the upper and lower right permanent first molar.

Detailed description on how to perform the analyses of this section in practice using the R package `bayesSurv` is available in the documentation directory of this package.

5 Discussion

We have suggested and implemented as an R package a semiparametric method to analyze bivariate doubly-interval-censored data in the presence of covariates. The method was applied to the analysis of a dental data set where all covariates were categorical. However, continuous covariates would not cause any difficulties and could have been

used as well. Although the method was presented to deal with doubly-interval-censored data it can be used to analyze also simple interval- or right-censored data.

A disadvantage of the current method is that it requires balanced data, i.e. exactly two observations must be supplied for each cluster and if only one observation of the cluster is missing the whole cluster must be removed from the analysis. Missingness in one event time out of the pair could have been solved using the Bayesian data augmentation in the same way as it solves the problem of censoring. However, if the missingness is caused by a missing covariate value, the Bayesian data augmentation would not help unless a measurement model is set up also for the covariates. With unbalanced data, the cluster specific approach of Komárek and Lesaffre (2005) can be used, however.

Acknowledgments

References

BESAG, J., GREEN, P., HIGDON, D., and MENGERSEN, K. (1995). Bayesian computation and stochastic systems (with Discussion). *Statistical Science*, **10**, 3–66.

CLAYTON, D. G. (1978). A model for association in bivariate life-tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141–151.

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

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.

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.

GUSTAFSON, P. (1995). A Bayesian analysis of bivariate survival data from a multicentre cancer clinical trial. *Statistics in Medicine*, **14**, 2525–2535.

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.

HOLT, J. D. and PRENTICE, R. (1974). Survival analysis in twin studies and matched-pair experiments. *Biometrika*, **61**, 17–30.

HUSTER, W. J., BROOKMEYER, R., and SELF, S. G. (1989). Modelling paired survival data with covariates. *Biometrics*, **45**, 145–156.

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

KOMÁREK, A. and LESAFFRE, E. (2005). Bayesian accelerated failure time model with multivariate doubly-interval-censored data and flexible distributional assumptions. *Submitted*.

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.

LESAFFRE, E. and BOGAERTS, K. (2005). Estimating Kendall's tau for bivariate interval censored data with a smooth estimate of the density. In FRANCIS, A. R., MATAWIE, K. M., OSHLACK, A., and SMYTH, G. K., editors, *Statistical Solutions to Modern Problems: Proceedings of the 20th International Workshop on Statistical Modelling*, pages 325–328, Sydney, 2005. University of Western Sydney.

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

OAKES, D. R. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B*, **44**, 414–428.

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

PLUMMER, M., BEST, N., COWLES, K., and VINES, K. (2005). *coda: Output analysis and diagnostics for MCMC*. URL `http://www-fis.iarc.fr/coda/`. R package version 0.9-5.

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.

ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, New York, Second edition. ISBN 0-387-21239-6.

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.

Table 1: Basic Model. Bayesian simultaneous two-sided p-value, posterior mean and simultaneous 95% equal-tail credible region for each factor variable, separately for the upper tooth 16 and the lower tooth 46.

Effect	Upper tooth 16		Lower tooth 46	
	Posterior mean	95% CR	Posterior mean	95% CR
	Emergence			
Gender	$p = 0.094$		$p = 0.142$	
girl	-0.018	$(-0.039, 0.003)$	-0.016	$(-0.036, 0.005)$
	Caries			
Gender	$p = 0.534$		$p = 0.403$	
girl	-0.034	$(-0.139, 0.073)$	-0.049	$(-0.162, 0.063)$
Brushing	$p = 0.003$		$p < 0.001$	
daily	0.230	$(0.086, 0.386)$	0.265	$(0.097, 0.426)$
Pit and fissure sealing	$p = 0.019$		$p = 0.401$	
present	0.157	$(0.028, 0.283)$	0.054	$(-0.077, 0.180)$
Occlusal plaque	$p = 0.014$		$p = 0.002$	
in pits and fissures	-0.183	$(-0.333, -0.031)$	-0.252	$(-0.404, -0.107)$
on total surface	-0.394	$(-0.819, -0.015)$	-0.479	$(-0.997, -0.038)$
Status second prim. molar	$p < 0.001$		$p < 0.001$	
decayed	-0.451	$(-0.704, -0.224)$	-0.383	$(-0.641, -0.151)$
filled	-0.628	$(-0.844, -0.414)$	-0.377	$(-0.588, -0.175)$
missing due to caries	-0.496	$(-1.377, 0.138)$	-0.739	$(-1.398, -0.208)$

Table 2: Final Model. Bayesian two-sided p-value, posterior mean and 95% equal-tail credible region for each factor variable.

Effect	Posterior mean	95% CR	p-value
		Emergence	
Jaw (*lower*, $\alpha_2^\zeta - \alpha_1^\zeta$)	0.017	(0.003, 0.032)	0.021
Gender (*girl*)	-0.017	$(-0.033, -0.003)$	0.018
		Caries	
Jaw (*lower*, $\alpha_2^\varepsilon - \alpha_1^\varepsilon$)	0.026	$(-0.158, 0.218)$	0.816
Gender (*girl*)	-0.044	$(-0.120, 0.033)$	0.267
Brushing (*daily*)	0.250	(0.139, 0.369)	< 0.001
Pit and fissure sealing (*present*)	0.109	(0.019, 0.195)	0.022
Occlusal plaque (*present*)	-0.228	$(-0.313, -0.141)$	< 0.001
Status second primary molar (*dmf*)	-0.482	$(-0.576, -0.388)$	< 0.001

Figure 1: Directed acyclic graph for the Bayesian specification of the model. Square boxes represent fixed or observed quantities, circles unknown parameters, solid arrows stochastic dependencies and dashed arrows deterministic dependencies.
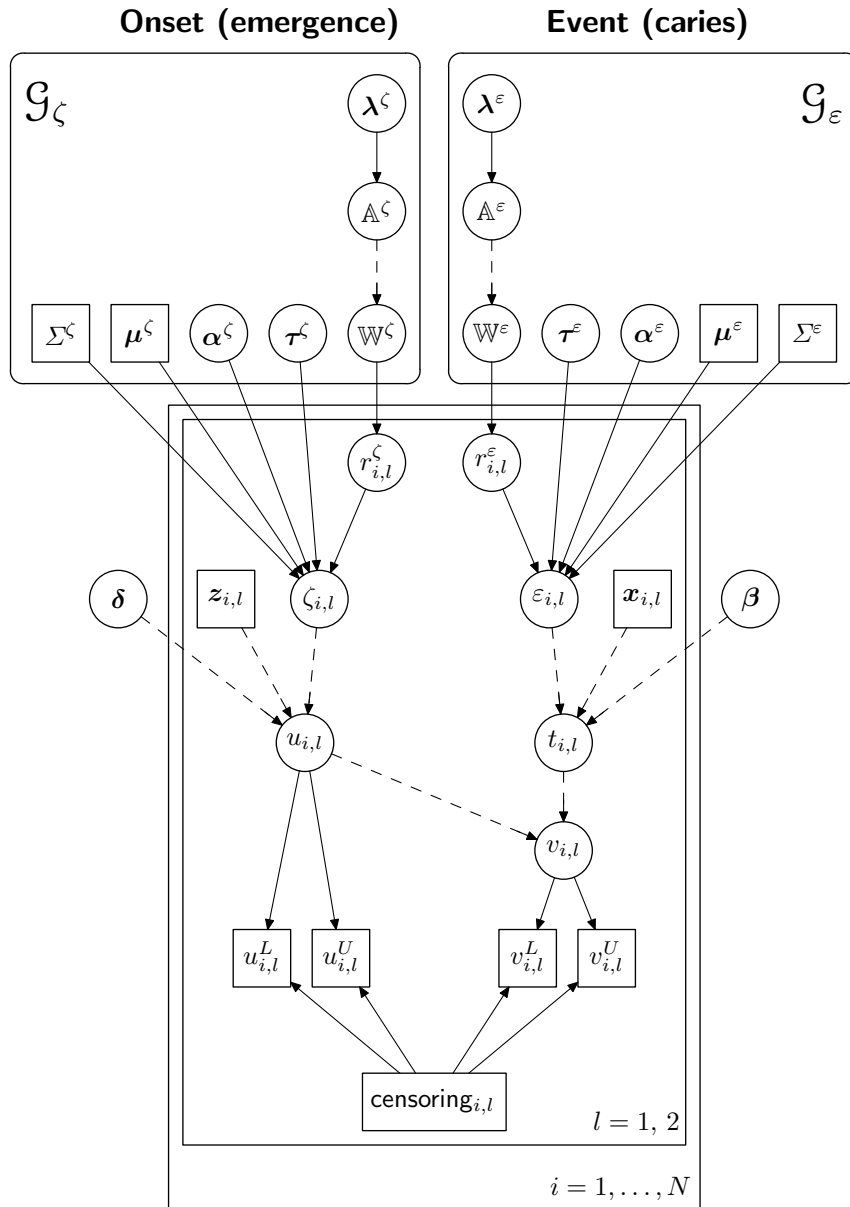
Figure 2: Final Model. Estimate of the density $g_\varepsilon(\varepsilon_1, \varepsilon_2)$ of the error term in the caries part of the model.
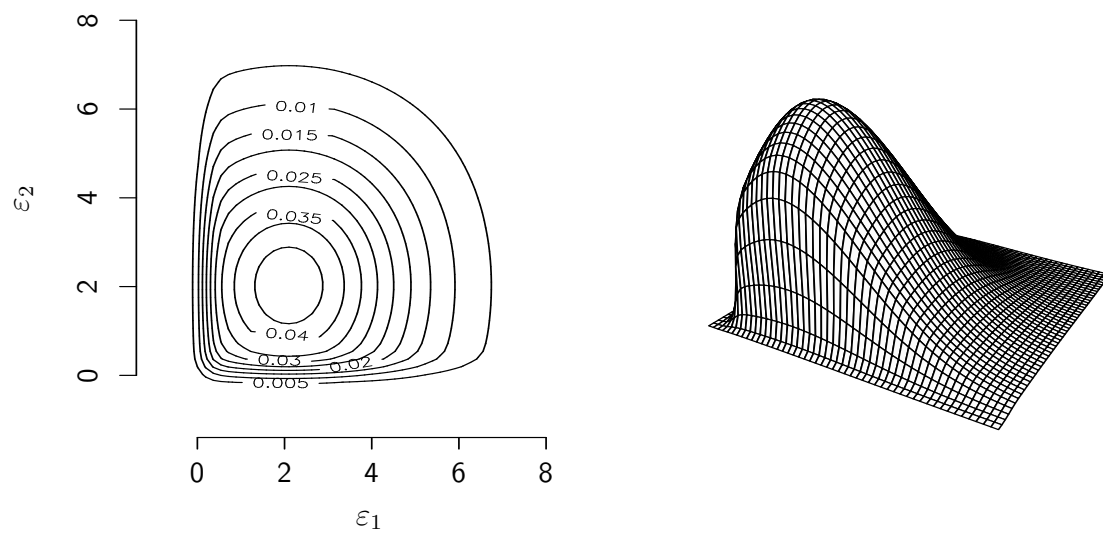
Figure 3: Final Model. Posterior predictive caries free (survivor) and caries hazard curves for tooth 16 of boys and the following combinations of covariates: solid and dashed lines for no plaque, present sealing, daily brushing and sound primary second molar (solid line) or dmf primary second molar (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).