January 25, 2007

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 70 15

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 Gaussian mixture (PGM) 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 that implements it.

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 proportional hazards (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 possibly having an impact on the time-to-event $T_{i,l}$. In the remainder of the paper, whenever a subscript i is used, it is understood that $i = 1, \ldots, N$. Further, whenever a double subscript i, l is used, it is understood that $i = 1, \ldots, N$, $l = 1, \ldots, n_i$.

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 $v_{i,l}^L \leq V_{i,l} \leq v_{i,l}^U$. 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).

In Section 2, we describe the data and the research question which motivated the development of our approach. Section 3 describes the assumed model. In Section 4, we specify the model from the Bayesian point of view, derive the posterior distribution of model parameters and describe the estimation procedure. The suggested approach is validated using a simulation study in Section 5 and in Section 6 it is shown how it tackled our research question. The paper ends with conclusions.

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. Additionally, the permanent first molar had to be excluded from the analysis if the adjacent deciduous second molar has not been present in the mouth already at the first examination. Note that for 948 children none of the permanent first molars was included in the analysis due to above mentioned reasons. In total, 3 520 children (12 485 permanent first molars) were included in the analysis of which 187 contributed one tooth, 317 two teeth, 400 three teeth and 2 616 all four 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. (2005a,b) where a relatively simple GEE multivariate log-logistic survival model was used for the caries times. Nevertheless, it must be pointed out that the doubly-interval-censored nature of the response was not taken into account properly in these works and this motivated the developments presented in this paper. We will return to this point in Section 6.3.

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. (2005a) 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) model. 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* (CS) 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 Gaussian mixture (PGM) with an overspecified number of mixture components.

The choice of the modelling approach has been motivated by the following reasons. When analyzing clustered survival data, the frailty Cox proportional hazard (PH) model (e.g., Therneau and Grambsch, 2000) is by far the most common choice. However, compared to the CS AFT model it has several disadvantages. Firstly, the choice of the frailty distribution can have an impact on the estimates of the regression coefficients. On the other hand, in the CS AFT model, estimates of the regression coefficient are robust against misspecification of the random effects distribution (Keiding, Andersen, and Klein, 1997; Lambert et al., 2004).

Secondly, the AFT model, in general, provides for our dental problem a better interpretation of the results than the PH model. As pointed out by Sir David Cox in Reid (1994) or in Hougaard (1999), the regression coefficients in the AFT model have a direct physical interpretation. For the dental application, an AFT model directly indicates how the emergence time (or time-to-caries) is increased or decreased. Whereas, in the PH

model, the regressors express the risk but it is more difficult for practitioners, such as dentists, to evaluate their practical impact.

3.1 Cluster-specific AFT model

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

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

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

where $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ are unknown regression parameter vectors, $\zeta_{i,l}$, are i.i.d. random variables with some density $g_\zeta(\zeta)$. Analogously, the error terms $\varepsilon_{i,l}$, are i.i.d. random variables with density $g_\varepsilon(\varepsilon)$. The random effects d_i, and b_i, 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.

The assumptions outlined above imply that, given the covariates, the *chronological* emergence time $U_{i,l}$ and the time-to-caries $T_{i,l}$ are independent for each i and l. Furthermore, the vectors $\boldsymbol{U}_i = (U_{i,1}, \ldots, U_{i,n_i})'$ and $\boldsymbol{T}_i = (T_{i,1}, \ldots, T_{i,n_i})'$ are independent for each i. 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 (*chronological* emergence time $U_{i,l}$). However, it is useful to stress that our assumption implies neither independence of the *chronological* caries time $V_{i,l}$ and the *chronological* emergence time $U_{i,l}$, nor independence of the vectors $\boldsymbol{V}_i = (V_{i,1}, \ldots, V_{i,n_i})'$ and $\boldsymbol{U}_i = (U_{i,1}, \ldots, U_{i,n_i})'$.

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 Penalized Gaussian mixture

To allow for a distributional flexibility, the univariate densities g_ζ, and g_ε of the random errors and the densities g_d, g_b of the random effects in the **Model I** will be modelled as a penalized Gaussian mixture (PGM).

Let $g(y)$ be a generic density of a random variable Y (later substitute ζ, ε, d, or b for Y). The density $g(y)$ is modelled as a location-and-scale transformed weighted sum of Gaussian densities over a fixed fine grid of knots $\boldsymbol{\mu} = (\mu_{-K}, \ldots, \mu_K)'$ centered around $\mu_0 = 0$. The means of the Gaussian 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. A satisfactory condition for (3) to be a distribution is that the weights satisfy $w_j > 0$ for all j and $\sum_j w_j = 1$. Constraints in the estimation can be avoided if a generalized logit transformation is used. That is, 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

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

Indeed, except an identifiability constraint (i.e., $a_0 = 0$), the elements of the vector $\boldsymbol{a}$ are unconstrained.

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 Gaussian 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 chose an 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 6), 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.

In the following, a super- or subscript ζ, ε, d, or b will be used to distinguish among parameters defining the PGM for different parts of the model.

3.3 Dependence between the onset and time-to-event random effects

It might be useful to relax the assumption of independence of the random effects from the onset and time-to-event parts of the model by assuming that vectors $(d_i,\ b_i)'$, $i = 1, \ldots, N$ are i.i.d. with a bivariate density $g_{d,b}(d,\ b)$. This would also induce possible dependence between the time-to-caries T and the *chronological* emergence time U. As a tool of the sensitivity analysis, we explored the model where the density $g_{d,b}(d,\ b)$ was assumed to be a bivariate normal distribution $\mathcal{N}_2\big(\mathbf{0},\ \mathbb{D}\big)$ with an unknown covariance matrix $\mathbb{D}$.

In a sequel, we shall call the model where independence of d and b is assumed and their densities are modelled using PGM's as **Model I**. The model with bivariate normal random effect vector $(d,\ b)$ will be called **Model D**.

3.4 Likelihood

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

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

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 $g_\zeta(\zeta_{i,l})$.

In **Model I**, $p(d_i,\, b_i) = g_d(d_i)\, g_b(b_i)$, where $g_d(d_i)$ and $g_b(b_i)$ are PGM's (3). Note that it is not possible to distinguish the intercept term of the random effect and the error term in each part of the doubly-censored AFT model, that is α^ζ and α^d and α^ε and α^b. For this reason, we fix the random effect intercepts, α^d and α^b, to zero. Finally, in **Model D**, $p(d_i,\, b_i)$ is a density of $\mathcal{N}_2\big(\mathbf{0},\, \mathbb{D}\big)$ distribution.

4 Estimation

The method of penalized maximum-likelihood was used by Ghidey et al. (2004) and Komárek et al. (2005) to obtain the estimates of the parameters defining the PGM. However, with clustered doubly-censored data, any method requiring optimization of the likelihood (5) is rather cumbersome and computationally intractable. Instead, a Bayesian approach together with the MCMC methodology (see, e.g., Robert and Casella, 2004) will be used here to avoid explicit integration and optimization.

4.1 Prior distribution

To specify the model from the Bayesian point of view, prior distributions for all unknown parameters have to be given. Firstly, we introduce some notation. Let $\mathcal{G}_\varepsilon$ refer 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 hyperparameter λ^ε which will be discussed further. Let $\mathcal{G}_\zeta$, $\mathcal{G}_b$, $\mathcal{G}_d$ be defined in an analogous manner. Finally, let $\mathcal{D}_{d,b} = \{\mathcal{G}_d,\ \mathcal{G}_b\}$ in **Model I** and $\mathcal{D}_{d,b} = \mathbb{D}$ in **Model D**.

Let $\boldsymbol{\theta}$ denote the vector of the unknown quantities. As usual with any Bayesian approach, besides traditional parameters, the vector $\boldsymbol{\theta}$ includes also "true" values of censored onset and event times, latent values of the random effects and latent values of the error terms of the AFT model which simplifies considerably all posterior calculations. Driven partially by the AFT models (1) and (2), the prior distribution of $\boldsymbol{\theta}$ is specified in a hierarchical way, namely

$$p(\boldsymbol{\theta}) = p\big(\{v_{i,l},\ u_{i,l},\ t_{i,l}\}_{i,l},\ \{\zeta_{i,l},\ \varepsilon_{i,l}\}_{i,l},\ \{d_i,\ b_i\}_i,\ \boldsymbol{\delta},\ \boldsymbol{\beta},\ \mathcal{G}_\zeta,\ \mathcal{G}_\varepsilon,\ \mathcal{D}_{d,b}\big) \tag{6}$$

$$= \prod_{i=1}^{N}\prod_{l=1}^{n_i}\Big\{p\big(v_{i,l} \,\big|\, u_{i,l},\ t_{i,l}\big)\ \times\ p\big(u_{i,l} \,\big|\, \boldsymbol{\delta},\ d_i,\ \zeta_{i,l}\big)\ \times\ p\big(t_{i,l} \,\big|\, \boldsymbol{\beta},\ b_i,\ \varepsilon_{i,l}\big)\Big\} \times \tag{7}$$

$$\prod_{i=1}^{N}\prod_{l=1}^{n_i}\Big\{p\big(\zeta_{i,l} \,\big|\, \mathcal{G}_\zeta\big)\ \times\ p\big(\varepsilon_{i,l} \,\big|\, \mathcal{G}_\varepsilon\big)\Big\}\ \times\ \prod_{i=1}^{N} p\big(d_i,\ b_i \,\big|\, \mathcal{D}_{d,b}\big) \times \tag{8}$$

$$p\big(\mathcal{G}_\zeta\big)\ \times\ p\big(\mathcal{G}_\varepsilon\big)\ \times\ p\big(\mathcal{D}_{d,b}\big)\ \times\ p\big(\boldsymbol{\delta}\big)\ \times\ p\big(\boldsymbol{\beta}\big). \tag{9}$$

Some of the terms in (7)–(9) follow directly from the previous model specification. Nevertheless, for clarity, all the terms will be discussed in more detail in the remainder of this section.

4.2 Prior distribution in Model I

Firstly, we concentrate on **Model I** where

$$p\big(d_i,\ b_i \,\big|\, \mathcal{D}_{d,b}\big) = p\big(d_i \,\big|\, \mathcal{G}_d\big)\ \times\ p\big(b_i \,\big|\, \mathcal{G}_b\big), \qquad\qquad p\big(\mathcal{D}_{d,b}\big) = p\big(\mathcal{G}_d\big)\ \times\ p\big(\mathcal{G}_b\big).$$

4.2.1 Prior for the time variables

The prior distributions for the time variables are degenerate and are determined by the AFT expressions from Section 3.1. Namely, the term (7) is obtained from

$$\begin{aligned} p\big(v_{i,l} \,\big|\, u_{i,l},\, t_{i,l}\big) &= I[v_{i,l} = u_{i,l} + t_{i,l}], \\ p\big(u_{i,l} \,\big|\, \boldsymbol{\delta},\, d_i,\, \zeta_{i,l}\big) &= I\big[u_{i,l} = \exp(d_i + \boldsymbol{\delta}'\boldsymbol{z}_{i,l} + \zeta_{i,l})\big], \\ p\big(t_{i,l} \,\big|\, \boldsymbol{\beta},\, b_i,\, \varepsilon_{i,l}\big) &= I\big[t_{i,l} = \exp(b_i + \boldsymbol{\beta}'\boldsymbol{x}_{i,l} + \varepsilon_{i,l})\big]. \end{aligned}$$

4.2.2 Prior for the error terms and for the random effects

Prior distributions $p\big(\zeta_{i,l} \,\big|\, \mathcal{G}_\zeta\big)$, $p\big(\varepsilon_{i,l} \,\big|\, \mathcal{G}_\varepsilon\big)$, $p\big(d_i \,\big|\, \mathcal{G}_d\big)$, and $p\big(b_i \,\big|\, \mathcal{G}_d\big)$, have all the same structure of the PGM (3). Note that the posterior computation can further be simplified by introduction of latent *allocation variables* $r^\zeta_{i,l}$, $r^\varepsilon_{i,l}$, r^d_i, r^b_i, having the following role.

Let y be a generic symbol for $\zeta_{i,l}$, $\varepsilon_{i,l}$, d_i, or b_i, and let $\mathcal{G}$ denote the corresponding set $\mathcal{G}^\zeta$, $\mathcal{G}^\varepsilon$, $\mathcal{G}^d$, or $\mathcal{G}^b$, respectively. Let r be the corresponding allocation variable $r^\zeta_{i,l}$, $r^\varepsilon_{i,l}$, r^d_i, or r^b_i, respectively. From the PGM (3) we have

$$p\big(y \,\big|\, \mathcal{G}\big) = p\big(y \,\big|\, \sigma,\, \boldsymbol{\mu},\, \alpha,\, \tau,\, \boldsymbol{w}\big) = \sum_{j=-K}^{K} w_j\, \varphi\big(y \,\big|\, \alpha + \tau\mu_j,\, (\tau\sigma)^2\big), \tag{10}$$

where $\varphi(\cdot|\mu,\, \sigma^2)$ denotes a density of $\mathcal{N}(\mu,\, \sigma^2)$. Let the allocation variable r have a discrete distribution with

$$p\big(r \,\big|\, \mathcal{G}\big) \equiv \Pr\big(r = j \,\big|\, \mathcal{G}\big) = \Pr\big(r = j \,\big|\, \boldsymbol{w}\big) = w_j, \qquad j = -K, \ldots, K,$$

and let

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

It is easily seen that the marginal density $p(y \,|\, \mathcal{G})$ obtained by integrating the allocation variable r out of the joint distribution $p(y,\, r \,|\, \mathcal{G})$ is indeed equal to the PGM (10).

Let $\boldsymbol{\psi} = \big(\{r^\zeta_{i,l},\, r^\varepsilon_{i,l},\, r^d_i,\, r^b_i\big)$ be the vector of all allocation variables. Employing the idea of Bayesian data augmentation (Tanner and Wong, 1987), we replace the prior

distribution $p(\boldsymbol{\theta})$, see (6), by the prior distribution $p(\boldsymbol{\theta}, \boldsymbol{\psi})$ which is equal to the product of (7), (8), and

$$\prod_{i=1}^{N}\prod_{l=1}^{n_i}\Big\{p\big(\zeta_{i,l} \,\big|\, \mathcal{G}_\zeta,\, r_{i,l}^{\zeta}\big) \,\times\, p\big(r_{i,l}^{\zeta} \,\big|\, \mathcal{G}^{\zeta}\big) \,\times\, p\big(\varepsilon_{i,l} \,\big|\, \mathcal{G}_\varepsilon,\, r_{i,l}^{\varepsilon}\big) \,\times\, p\big(r_{i,l}^{\varepsilon} \,\big|\, \mathcal{G}^{\varepsilon}\big)\Big\} \times \prod_{i=1}^{N}\Big\{p\big(d_i \,\big|\, \mathcal{G}_d,\, r_i^{d}\big) \,\times\, p\big(r_i^{d} \,\big|\, \mathcal{G}^{d}\big) \,\times\, p\big(b_i \,\big|\, \mathcal{G}_b,\, r_i^{b}\big) \,\times\, p\big(r_i^{b} \,\big|\, \mathcal{G}^{b}\big)\Big\}.$$

Since integrating the allocations $\boldsymbol{\psi}$ out of the distribution $p(\boldsymbol{\theta}, \boldsymbol{\psi})$ leads to the original prior (6) all posterior characteristics of the parameter of interest $\boldsymbol{\theta}$ remain the same. However, the joint prior $p(\boldsymbol{\theta}, \boldsymbol{\psi})$ has the advantage that it explicitly does not involve any mixture distributions.

4.2.3 Prior for the penalized Gaussian mixtures

The PGM prior distributions $p\big(\mathcal{G}_\zeta\big)$, $p\big(\mathcal{G}_\varepsilon\big)$, $p\big(\mathcal{G}_d\big)$, and $p\big(\mathcal{G}_b\big)$ are all of the same structure. Let $\mathcal{G} = \{\sigma,\ \boldsymbol{\mu},\ \alpha,\ \tau,\ \boldsymbol{w},\ \boldsymbol{a},\ \lambda\}$ stand for $\mathcal{G}_\zeta$, $\mathcal{G}_\varepsilon$, $\mathcal{G}_d$, or $\mathcal{G}_b$, respectively. Remember that only parameters $\boldsymbol{a}$ (or $\boldsymbol{w}$), λ, α, τ from $\mathcal{G}$ are unknown and their prior is decomposed in the following way

$$p\big(\mathcal{G}\big) = p\big(\boldsymbol{a},\, \lambda,\, \alpha,\, \tau\big) = p\big(\boldsymbol{a} \,\big|\, \lambda\big) \,\times\, p\big(\lambda\big) \,\times\, p\big(\alpha\big) \,\times\, p\big(\tau\big).$$

As mentioned in Section 3.2, we consider the PGM as a spline model for an unknown density function. The weights $\boldsymbol{w}$, or equivalently the transformed weights $\boldsymbol{a}$, are viewed as spline coefficients. Spline smoothing using a spline basis specified over a grid with a relatively high number of equidistant knots, like in our PGM model, became widely used after the seminal work of Eilers and Marx (1996). To avoid overfitting of the data and identifiability problems caused by the large number of spline coefficients to be estimated, they suggested using penalized maximum-likelihood for estimation and call their approach as "smoothing with penalized splines" (P-splines).

In our notation, if the penalized maximum-likelihood estimation had been used, the

log-likelihood would have been penalized by adding the term

$$Q\big(\boldsymbol{a} \mid \lambda\big) = -\frac{\lambda}{2} \sum_{j=-K+s}^{K} \big(\Delta^s a_j\big)^2 = -\frac{\lambda}{2} \boldsymbol{a}' \mathbb{P}_s' \mathbb{P}_s \boldsymbol{a}, \tag{11}$$

where Δ^s denotes a difference operator of order s (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 6) and $\mathbb{P}_s$ the corresponding difference operator matrix. The hyperparameter λ, which has to be estimated, controls the smoothness of the fitted curve, i.e., the unknown density in our case.

For more complex models where a Bayesian estimation using MCMC is a useful alternative to the penalized maximum-likelihood, the penalty term corresponds to the logarithm of the prior density of the spline coefficients. That is,

$$p\big(\boldsymbol{a} \mid \lambda\big) \propto \exp\Big\{Q\big(\boldsymbol{a} \mid \lambda\big)\Big\} = \exp\Big(-\frac{\lambda}{2} \boldsymbol{a}' \mathbb{P}_s' \mathbb{P}_s \boldsymbol{a}\Big). \tag{12}$$

This correspondence between the penalized maximum-likelihood and a Bayesian model specification has been mentioned already in Wahba (1983). Practically, the Bayesian version of the P-splines has been exploited to smooth the regression function in different contexts, e.g., by Fahrmeir and Lang (2001); Lang and Brezger (2004); Hennerfeind et al. (2006); Kneib (2007).

The penalty (11) decreases in its absolute value with decreasing differences between neighboring spline weights, represented by the vector $\mathbb{P}_s \boldsymbol{a}$. Consequently, the prior (12) favors smooth estimates of the estimated functions (densities in our case). Further, the prior (12) is actually a multivariate normal density with zero mean and covariance matrix $\lambda^{-1}\big(\mathbb{P}_s' \mathbb{P}_s\big)^-$, where $\big(\mathbb{P}_s' \mathbb{P}_s\big)^-$ denotes a generalized inverse of the matrix $\mathbb{P}_s' \mathbb{P}_s$. This distribution is known as an intrinsic Gaussian Markov random field (IGMRF) extensively used in spatial statistics (e.g., Rue and Held, 2005).

It is seen from the penalty motivated prior (12) that the smoothing hyperparameter λ scales the prior precision (inverse covariance) matrix of the vector $\boldsymbol{a}$ of transformed spline weights. This also motivates a choice for the prior distribution $p(\lambda)$. The standard

prior for precision parameters is the Gamma distribution. A highly dispersed but proper prior is, e.g., G(h_1^λ, h_2^λ) for $h_1^\lambda = 1$ and h_2 being a small positive quantity, or for $h_1^\lambda = h_2^\lambda$ being small positive quantities. In the analysis in Section 6, G(1, 0.005) is used as a prior for all smoothing hyperparameters involved.

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 $p(\alpha)$. In the analysis in Section 6, $p(\alpha) = \mathcal{N}(0, 100)$.

Finally, for the precision τ^{-2} we have taken a highly dispersed Gamma distribution. In the analysis in Section 6, $p(\tau^{-2}) = \text{Gamma}(1, 0.005)$. Alternatively a uniform distribution on τ which is sometimes preferred for hierarchical models (Gelman et al., 2004, pp. 136, 390) could be taken.

4.2.4 Prior for the regression parameters

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. In the analysis in Section 6, we used the product of $\mathcal{N}(0, 100)$ distributions.

4.3 Prior distribution in Model D

In **Model D** we have

$$p\big(d_i,\, b_i \,\big|\, \mathcal{D}_{d,b}\big) = p\big(d_i,\, b_i \,\big|\, \mathbb{D}\big), \qquad\qquad p\big(\mathcal{D}_{d,b}\big) = p\big(\mathbb{D}^{-1}\big).$$

According to Section 3.3, $p(d_i,\, b_i \,|\, \mathbb{D})$ is a density of $\mathcal{N}_2\big(\mathbf{0},\, \mathbb{D}\big)$. The prior distribution $p(\mathbb{D}^{-1})$ for the inverse of the covariance matrix of the random effects is taken to be a Wishart distribution with *df* degrees of freedom and a scale matrix $\mathbb{S}$. A diffuse prior is obtained, e.g., with $df = 1 + c_1$, $\mathbb{S} = \text{diag}(c_2, c_2)$, where c_1 is a small and c_2 a large positive constant. For the analysis in Section 6 using **Model D**, we used $df = 1.5$ and

$\mathbb{S} = \text{diag}(100,\, 100)$.

4.4 Posterior distribution

Let $\boldsymbol{C}_{i,l}$, denote the censoring process to which the (i,l)-th observation is imposed and let $[\text{data}] = \{u^L_{i,l},\, u^U_{i,l},\, v^L_{i,l},\, v^U_{i,l},\, \boldsymbol{C}_{i,l}\}_{i,l}$. The joint posterior distribution of the parameter vector $\boldsymbol{\theta}$ and the vector of allocations $\boldsymbol{\psi}$ introduced in Section 4.2 is obtained using Bayes' formula as $p\big(\boldsymbol{\theta},\, \boldsymbol{\psi} \,\big|\, [\text{data}]\big) \propto p\big([\text{data}] \,\big|\, \boldsymbol{\theta},\, \boldsymbol{\psi}\big) \times p\big(\boldsymbol{\theta},\, \boldsymbol{\psi}\big)$, where $p\big(\boldsymbol{\theta},\, \boldsymbol{\psi}\big)$ has been described in Sections 4.2 and 4.3.

Due to a hierarchical structure of the model, the data are conditionally independent, given the "true" onset and event times $u_{i,l}$ and $v_{i,l}$ that form a part of $\boldsymbol{\theta}$, of all other components of the vectors $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$. That is $p\big([\text{data}] \,\big|\, \boldsymbol{\theta},\, \boldsymbol{\psi}\big) = p\big([\text{data}] \,\big|\, \{u_{i,l}\}_{i,l},\, \{v_{i,l}\}_{i,l}\big)$. This can further be decomposed as

$$p\big([\text{data}]\big|\{u_{i,l}\}_{i,l},\, \{v_{i,l}\}_{i,l}\big) = \prod_{i=1}^{N}\prod_{l=1}^{n_i}\Big\{p\big(u^L_{i,l},\, u^U_{i,l}|u_{i,l},\, \boldsymbol{C}_{i,l}\big) \times p\big(v^L_{i,l},\, v^U_{i,l}|v_{i,l},\, \boldsymbol{C}_{i,l}\big) \times p\big(\boldsymbol{C}_{i,l}\big)\Big\}. \tag{13}$$

The conditional distributions $p\big(u^L_{i,l},\, u^U_{i,l} \,\big|\, u_{i,l},\, \boldsymbol{C}_{i,l}\big)$ and $p\big(v^L_{i,l},\, v^U_{i,l} \,\big|\, v_{i,l},\, \boldsymbol{C}_{i,l}\big)$ are degenerate. 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 $p\big(u^L_{i,l} = c_{i,l},\, u^U_{i,l} = \infty \,\big|\, u_{i,l},\, C_{i,l} = c_{i,l}\big) = I[u_{i,l} > c_{i,l}]$.

Finally, provided that the censoring is independent and noninformative, the term $p\big(\boldsymbol{C}_{i,l}\big)$ in (13) acts only as a multiplicative constant with respect to the model parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$. So for the purpose of the posterior calculations, we do not have to specify a model for the censoring process.

The hierarchical structure of both the prior and posterior distribution for proposed models may best be seen from the directed acyclic graphs. These can be found in Komárek and Lesaffre (2007).

4.5 Markov chain Monte Carlo

Inference will be based on a sample from the posterior distribution obtained using the MCMC methodology. For (blocks) of model parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ we derived full conditional distributions and used basically the Gibbs algorithm (Gelfand and Smith, 1990) to perform one iteration of the MCMC. When the full conditional distribution was not of a standard form, we exploited either adaptive rejection sampling (Gilks and Wild, 1992) or slice sampling (Neal, 2003). Details can be found in Komárek and Lesaffre (2007).

A software package, called bayesSurv, was written combining the R language (R Development Core Team, 2006) with programs in C++, and is available from the *Comprehensive R Archive Network* on `http://www.R-project.org`.

4.6 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

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

Further,

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

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

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

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 $\mathcal{G}_b^{(m)}$ in **Model I** or from the zero mean normal distribution with the variance equal to the diagonal element of the matrix $\mathbb{D}^{(m)}$ in **Model D**. Analogously, the posterior predictive survivor function for the onset time or posterior predictive hazard functions are computed.

5 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 $\varepsilon_{i,l}$ were 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 were 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 6.

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 $0.4\mathcal{N}(-2.000,\, 0.25) + 0.6\mathcal{N}(1.333,\, 0.36)$ and 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. It shows that the regression parameters are estimated with only a minimal bias 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 1 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). Full results of the simulation study are given in Komárek and Lesaffre (2007).

< Table 1 about here >

< Figure 1 about here >

6 Analysis of the Signal Tandmobiel® data

We started with the **Basic Analysis** 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 Analysis** 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.

Clinically, the permanent teeth cannot emerge during more than the first 5 years after birth (see, e.g., Ekstrand, Christiansen, and Christiansen, 2003). For this reason we subtracted 5 years from all observed times, i.e. $\log(U_{i,l} - 5)$ was used in the left-hand side of the model formula (1). Note that this only changes the interpretation of the emergence time which is now measured from 5 years of age and not from birth. However, the time-to-caries is unchanged.

Based on the results for the **Basic Analysis** (see below) we conducted the **Final Analysis** in which 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*.

Decisions concerning the simplification of the model leading to the **Final Analysis**

were based on (posterior) pseudo-contour probabilities P suggested by Held (2004, Section 2.1). The pseudo-contour probability P is defined as 1 minus the content of the (simultaneous) credible region, see Besag et al. (1995, p. 30), which just covers a vector of zeros. As such, it can be viewed as a Bayesian counterpart of classical two-sided P-values.

Both analyses were performed using **Model I**. To explore possible dependency between emergence and caries child-specific random effects, we also performed the **Final Analysis** using **Model D**. Finally, the question of whether the adjustment of the left-censored emergence times using recorded eruption stages (see Section 2) has significant influence is considered. To address this, we also conducted the **Final Analysis** using **Model I** with the original left-censored emergence times. This will be referred to as **Model I.2.**

Descriptions of these analyses using the R package `bayesSurv` can be found in Komárek and Lesaffre (2007).

6.1 Basic Analysis

For each analysis we ran two chains, both sampling 250 000 values with 1:3 thinning. Simulation of both chains (1.5 million MCMC iterations) took about 24 hours on a dual AMD Opteron 250 processor with 2 GB RAM running Unix OS. The convergence was evaluated by a critical examination of the trace and autocorrelation plots and using the method of Gelman and Rubin (1992). From each chain we kept the last 50 000 values for the inference.

In the **Basic Analysis**, 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. For emergence, the pseudo-contour probability for tooth:gender interaction was higher than 0.5. For the caries part of the model the pseudo-contour

probabilities were higher than 0.5 for the two interactions of tooth with gender and plaque and higher than 0.1 for the three interactions with brushing, sealants and status. It seemed that 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 was a significant difference 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 Analysis**.

6.2 Final Analysis

Table 2 shows posterior medians, 95% highest posterior density (HPD) intervals for the acceleration factors (exponentiated values of the regression parameters) and corresponding pseudo-contour probabilities in the **Final Analysis** using all three considered models (**Model I, I.2, D**). It is seen that neither for the emergence nor for the caries process is there 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. For example, the fact that the neighboring deciduous second molar is either decayed or filled or extracted due to caries accelerates the time to caries with a factor of 0.87 (**Model I**).

< Table 2 about here >

A low value of the correlation coefficient between the emergence and caries random effects has been observed in **Model D**. The posterior median and 95% HPD interval for $\text{cor}(d_i, b_i)$ is -0.133 and $(-0.196, -0.068)$, respectively. As it is seen from Table 2, the assumption of the independence between the emergence and caries random effects did not influence the results for the regression as both the posterior medians and 95% HPD intervals for **Model I** and **Model D** are in the most cases practically the same.

Compared to **Model I**, **Model I.2** led, not surprisingly, to a lower precision in estimation (wider HPD intervals) and to slightly weaker covariate effects on the caries (acceleration factors closer to 1). However, the key findings concerning the covariate effects remain the same. We argue that the adjustment of the left-censored emergence times for **Model I** according to the eruption stages as described in Section 2 was based on scientifically confirmed clinical knowledge. That is why a higher precision of **Model I** is not artificial or incorrect. Consequently, there is no reason to prefer **Model I.2** to **Model I**.

Figure 2 shows the 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 and then remains almost constant. Whereas, when the teeth are not brushed daily and are exposed to other risk factors the hazard starts to increase 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 at approximately 1 year after emergence was also observed by Leroy et al. (2005a) and can be explained by the fact that teeth are most vulnerable for caries soon after 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.

< Figure 2 about here >

6.3 Discussion

One might ask a question why a relatively complex estimation using MCMC is needed for the analysis of doubly-censored data. A necessity of the complex estimation might be illustrated by comparison of our analysis of the Signal Tandmobiel® data to a simpler analysis of the same data performed by Leroy et al. (2005a,b) using a frequentist approach, a GEE method. In the works Leroy et al. (2005a,b), doubly-interval-censored caries times were changed into singly-censored ones using two approaches typical in such situations. Unfortunately, both approaches are not completely appropriate, mainly due to complication in the interpretation of the results.

The simplest approach to avoid doubly-censored data is by taking an exactly known onset time which leads to simply-censored data. This approach was taken by Leroy et al. (2005b) where the time-to-caries is defined as the time between birth and the time of cavity. The risk for caries should be equal to zero during the period that the permanent tooth has not emerged yet. However, this is not guaranteed with this approach. In fact, the estimated model in Leroy et al. (2005b) shows a non-zero risk for caries development before permanent teeth have emerged. Furthermore, any survival model requires that the covariate values are known at each time point starting from the defined onset. Except

for gender, this was not the case in Leroy et al. (2005b).

A second common approach to dealing with doubly-interval-censored data assumes that the time-to-event T is simply-interval-censored with the observed interval $[v^L - u^U,\ v^U - u^L]$. This is indeed the interval where, given the data, the time-to-event T may lie. However, the statistical analysis based on such intervals is incorrect unless the distribution of the onset time U is uniform and U and T are independent (De Gruttola and Lagakos, 1989). This approach was taken in Leroy et al. (2005a).

Finally, given the reported time needed to run the MCMC simulation, one might find our approach impractical. However, our analysis involved more than 12 000 observations whereas most datasets encountered in practice are much smaller, implying much shorter computation time. Furthermore, in the exploratory parts of the analysis, shorter chains or random subsamples of the data can be used which significantly reduces the computation time.

7 Conclusions

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 optimizations unavoidable with classical maximum-likelihood based estimation. We use a penalty-like prior distribution for the transformed mixture weights $\boldsymbol{a}$ and vague priors for all remaining parameters. Taking account of this, 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 unnecessary to perform the classical checks for correct distributional specification. Clearly, this step cannot be avoided

when using fully parametric methods. However, for censored data (let alone doubly-interval-censored data) this is far from trivial. As illustrated in Section 6, important new findings concerning the distribution of the event time, derived, e.g., from the shape of the hazard function, can be discovered without 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 of the first author was performed in the framework of the postdoctoral mandate PDM/06/242 financed by the Research Funds of 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.

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Applied Statistics*, **50**, 201–220.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

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.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulations using multiple sequences (with Discussion). *Statistical Science*, **7**, 457–511.

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 posterior probability statements from Monte Carlo output. *Journal of Computational and Graphical Statistics*, **13**, 20–35.

Hennerfeind, A., Brezger, A., and Fahrmeir, L. (2006). Geoadditive survival models. *Journal of the American Statistical Association*, **101**, 1065–1075.

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

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.

Kneib, T. (2007). Geoadditive hazard regression for interval censored survival times. *Accepted in Computational Statistics and Data Analysis*, **00**, 000–000.

Komárek, A. and Lesaffre, E. (2007). Supplement to "Bayesian accelerated failure time model with multivariate doubly-interval-censored data and flexible distributional assumptions". Technical report, Katholieke Universiteit Leuven, Biostatistical Centre. URL `http://www.med.kuleuven.be/biostat`.

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.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, **13**, 183–212.

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

Leroy, R., Bogaerts, K., Lesaffre, E., and Declerck, D. (2005b). Multivariate survival analysis for the identification of factors associated with cavity formation in permanent first molars. *European Journal of Oral Sciences*, **113**, 145–152.

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 (2006). *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.

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

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC, Boca Raton. ISBN 978-1-58488-432-3.

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.

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

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.

Wahba, G. (1983). Bayesian "confidence intervals" for the cross–validated smoothing spline. *Journal of the Royal Statistical Society, Series B*, **45**, 133–150.

Figure 1: 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.

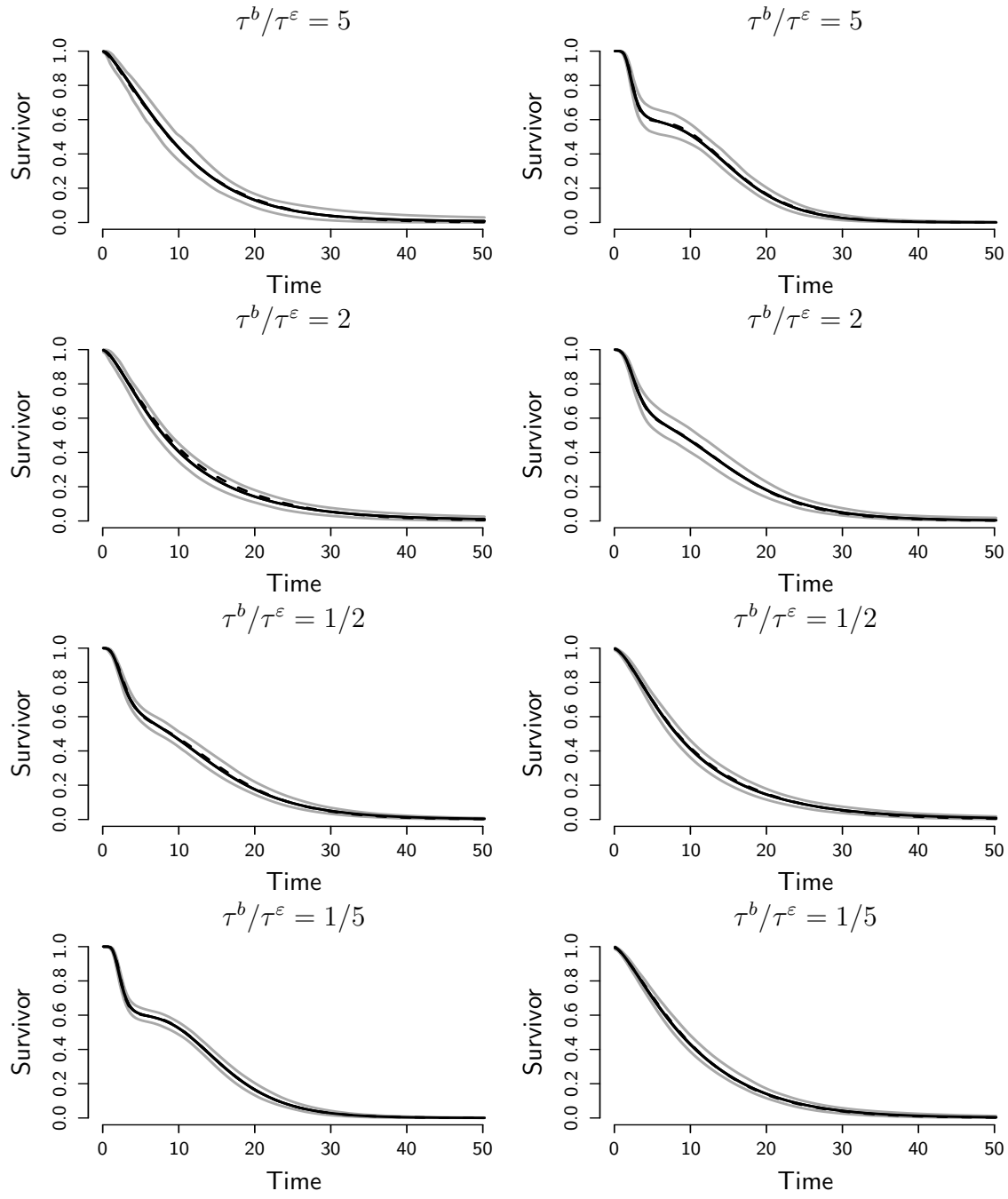
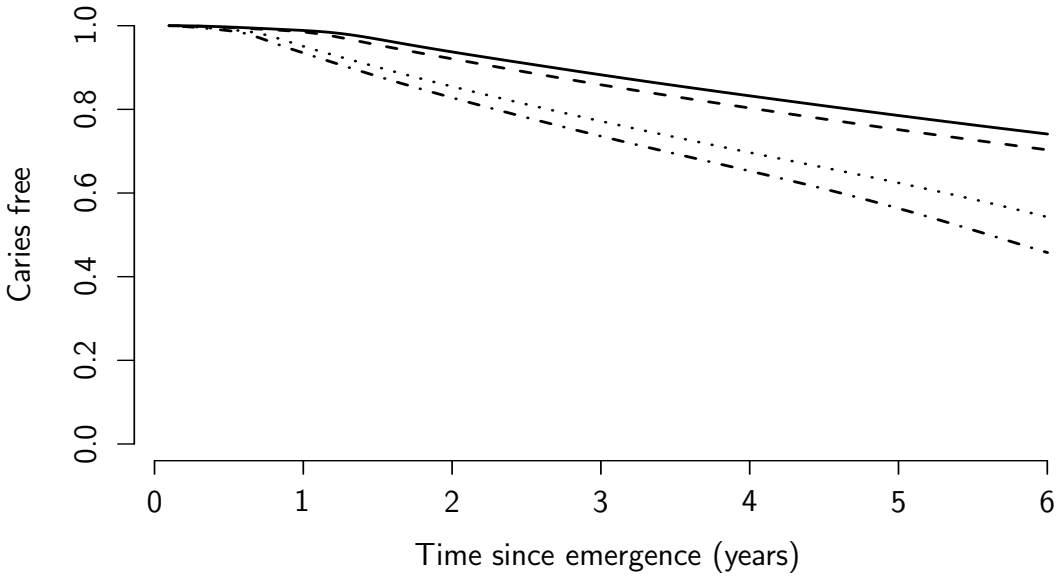
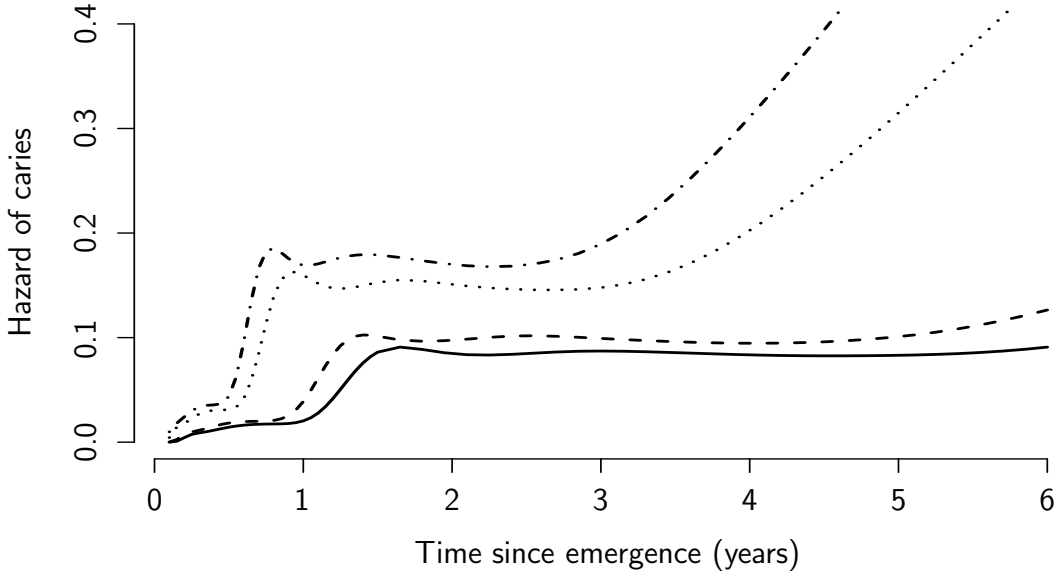


Figure 2: Signal Tandmobiel® data, **Final Analysis, Model I.** Posterior predictive caries free (survivor) and caries hazard curves for tooth 16 of boys and the following combinations of covariates: solid and dashed lines for no plaque, present sealing, daily brushing and sound primary second molar (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).





$\tau^d/\tau^\zeta = \tau^b/\tau^\varepsilon$	$\delta_1 = 0.20$	$\delta_2 = -0.10$	$\beta_1 = 0.30$	$\beta_2 = -0.15$
	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 ($\times 10^{-4}$).

	Model I		**Model I.2**		**Model D**	
Effect	Posterior median	95% HPD interval	Posterior median	95% HPD interval	Posterior median	95% HPD interval
			Emergence			
Tooth	$P > 0.5$		$P > 0.5$		$P > 0.5$	
tooth 26	0.997	(0.989, 1.005)	1.006	(0.869, 1.158)	0.998	(0.991, 1.005)
tooth 36	1.001	(0.993, 1.010)	0.928	(0.806, 1.068)	1.002	(0.994, 1.009)
tooth 46	1.002	(0.994, 1.011)	1.198	(1.039, 1.401)	1.004	(0.996, 1.011)
Gender	$P = 0.008$		$P = 0.039$		$P = 0.005$	
girl	0.977	(0.962, 0.993)	0.897	(0.806, 0.988)	0.976	(0.958, 0.992)
			Caries			
Tooth	$P > 0.5$		$P > 0.5$		$P > 0.5$	
tooth 26	0.994	(0.962, 1.024)	0.991	(0.969, 1.013)	0.995	(0.965, 1.025)
tooth 36	0.991	(0.956, 1.026)	1.000	(0.975, 1.024)	1.000	(0.966, 1.034)
tooth 46	0.984	(0.949, 1.019)	0.978	(0.955, 1.001)	0.989	(0.958, 1.023)
Gender	$P = 0.085$		$P = 0.083$		$P = 0.106$	
girl	0.931	(0.859, 1.011)	0.963	(0.923, 1.005)	0.934	(0.860, 1.011)
Brushing	$P < 0.001$		$P < 0.001$		$P < 0.001$	
daily	1.400	(1.263, 1.547)	1.173	(1.103, 1.244)	1.622	(1.447, 1.796)
Sealants	$P < 0.001$		$P < 0.001$		$P < 0.001$	
present	1.126	(1.059, 1.193)	1.092	(1.057, 1.132)	1.131	(1.067, 1.196)
Plaque	$P < 0.001$		$P < 0.001$		$P < 0.001$	
present	0.892	(0.845, 0.936)	0.935	(0.902, 0.968)	0.900	(0.853, 0.947)
Status	$P < 0.001$		$P < 0.001$		$P < 0.001$	
dmf	0.870	(0.825, 0.913)	0.933	(0.901, 0.962)	0.869	(0.828, 0.910)

Table 2: Signal Tandmobiel® data, **Final Analysis.** Posterior medians and 95% highest posterior density intervals for the acceleration factors ($\exp(\delta)$ and $\exp(\beta)$) parameters.