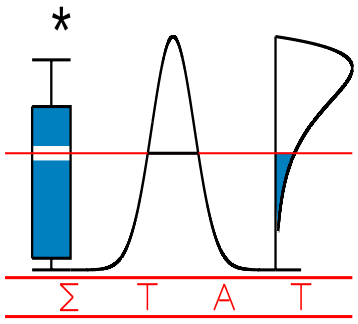
T E C H N I C A L

R E P O R T

0314

A BAYESIAN ANALYSIS OF MULTIVARIATE DOUBLY INTERVAL CENSORED DENTAL DATA

A. KOMÁREK, E. LESAFFRE, T. HÄRKÄNEN, D. DECLERCK AND J. I. VIRTANEN



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

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

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

A Bayesian Analysis of Multivariate Doubly Interval Censored Dental Data

Arnošt Komárek[1], Emmanuel Lesaffre[1], Tommi Härkänen[2], Dominique Declerck[3], Jorma I. Virtanen[4]

[1] Catholic University Leuven, Biostatistical Centre, Kapucijnenvoer 35, B–3000, Leuven, Belgium

[2] National Public Health Institute, Mannerheimintie 166, FIN–00300, Helsinki, Finland

[3] Catholic University Leuven, School of Dentistry, Kapucijnenvoer 33, B–3000, Leuven, Belgium

[4] University of Oulu, Institute of Dentistry, P.O. Box 5281, FIN–90014, Oulu, Finland

Corresponding author: Arnošt Komárek.

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

Tel.: +32 - 16 - 336886.

Fax: +32 - 16 - 336900.

ABSTRACT. The effect of fluoride-intake on caries development on the permanent first molars is examined. Although the research question is fairly standard in dentistry the nature of the dental data implied quite involved statistical analysis. Indeed, the response of interest is the time between the emergence of the tooth and the time when caries is first observed. Most often in dentistry (and also here), both events are only known to lie between two dental examinations which implies doubly interval censored data. Further, dependencies between teeth in the same mouth have to be accounted for, and this implies the analysis of multivariate survival data. Finally, a huge amount of left censored emergence times of the permanent first molars had to be taken into account. A Bayesian analysis based on a model of Härkänen et al. (2000) was performed while addressing all mentioned problems.

KEY WORDS. Bayesian analysis; Intensity models; Multivariate doubly interval censored data;

1. INTRODUCTION

In this paper we present the analysis of a longitudinal dental data set (Signal Tandmobiel® study) to tackle the following research question: does fluoride-intake at a young age have a positive effect on caries development in permanent teeth? Here we will examine this impact on the caries experience of the permanent first molars, i.e. teeth emerging at the age of about 6–7 years. The caries process will be simultaneously modelled for the two maxillary and two mandibular permanent first molars. Since the children from our dental data set were examined annually, this question involves doubly interval-censored data. Indeed, both tooth emergence as well as the onset of caries development can only be measured in a coarse manner, resulting in interval censored data for emergence and for onset of caries. Further, since the time at-risk should be taken into account to measure the impact of fluoride the response should be the time between emergence and the onset of caries development which implies doubly-interval censored data. To improve the precision of the estimation procedure it is advantageous to model several teeth simultaneously and this leads to

multivariate doubly interval censored data. Indeed, it appeared that a considerable number of the permanent first molars had already emerged at the first annual examination implying a relatively high number of left-censored emergence times. The emergence of the other teeth in the mouth can then be used to better estimate the emergence of the permanent first molars. Our research question is not uncommon in dentistry, although implying a quite involved statistical analysis, ruling out the use of any classical statistical package in an elegant way. Bayesian methodology often allows to analyze complex data structures. A very popular in this context is WinBUGS. However, WinBUGS is too restrictive for our problem. Instead, the software package BITE (Härkänen 2001), based on a non-parametric Bayesian survival model developed by Härkänen et al. (2000) was used here.

In Section 2, a description of data will be given and the research question will be explained in more detail. In the third section the Bayesian model suggested by Härkänen et al. (2000) will be explained briefly. The fourth section will give details on the inference. Results are found in Section 5 and we end with a further discussion of our results.

2. DESCRIPTION OF COLLECTED DENTAL DATA AND RESEARCH QUESTION

In the Signal Tandmobiel® study, oral health data from 4 468 Flemish schoolchildren born in 1989 were collected annually. The sample consisted of 2 134 boys and 2 296 girls cluster sampled from randomly chosen Flemish schools, such that the probability of being selected was approximately equal for all children in Flanders. Two stratification factors, i.e. province (5 provinces) and educational system (3 school systems) establishing 15 strata, were also taken into account. The sample represents about 7% of the corresponding Flemish population of schoolchildren. Detailed dental data on tooth and tooth-surface level (caries experience, gingivitis, etc.) were collected by a team of 16 dentists whose examination method was calibrated every half a year. But, also data from a questionnaire (given to the parents) on dietary and oral hygiene habits were collected. The dentists were operating from a dental bus which visited 179 schools in Flanders (North of

Belgium) from 1996 to 2001. Hence our data set consists of a series of at most 6 longitudinal dental observations and reported oral health habits taken from 1996 to 2001. When a child was absent at school at the day of the examination (due to illness, moved to another school, etc.) it was not examined in that particular year. This excluded 38 children not present at any of the six examinations. More information on the setup of the study and on initial research results can be found in Vanobbergen et al. (2000).

Here, we are interested in the question whether fluoride-intake has a protecting effect on permanent teeth for caries experience. Actually, although a considerable amount of research has been done on this topic, there is still some controversion. From our data, we have found that the use of fluoride reduces caries activity in primary teeth, see Vanobbergen et al. (2001). Moreover, there is some evidence that fluoride supplementation delays the emergence of the permanent teeth, see Leroy et al. (2002). In any study on the effect of fluoride on caries experience of permanent teeth, the eventual delay in emergence due to fluorides shortens the time at risk of permanent teeth for caries development in children and adolescent with use of fluoride.

Unfortunately, it is not easy to measure fluoride-intake children accurately, since it can be taken by them in different ways: (1) systemically (fluoride supplements), (2) in toothpaste or (3) from drinking water and all these sources can be recorded only roughly. Therefore it was decided to measure a history of fluoride-intake by the degree of fluorosis present on reference teeth during the fourth, fifth and sixth year of the study. Fluorosis is the most common side-effect of fluoride use and appears as discrete white spots on the enamel of several teeth, in some cases resulting in severe enamel dysplasia. For this analysis, only the presence or absence of fluorosis on four permanent maxillary incisors during the fourth, fifth and sixth year of the study was taken into account. Indeed, a child was considered fluoride-positive if there were observable white spots on at least two permanent maxillary incisors during the fourth year of the study or during both the fifth and sixth years of the study.

In the Signal Tandmobiel® study, fluorosis prevalence is rather low (480 children) and mainly related to the use of fluoride supplements and toothpaste during the period of tooth development, since the fluoride-level in Flemish drinking-water is low in most areas (< 0.1 ppm). For analyzes purposes, we included the 480 fluorosis children and 960 randomly selected fluorosis-free children. To measure the impact of fluoride-intake on caries experience, "fluorosis" will be used as a covariate in a survival model with the time at risk for caries experience for a particular tooth as response. Except for the permanent first molars for which 25% of the children had some caries experience before the end of the observation period, the incidence of caries experience at the age of 12 was negligible for other teeth (at most 1.4%). Hence, we will analyze here only the permanent first molars, see Figure 1. Furthermore, our analyzes will also involve gender as a covariate.

Based on the Signal Tandmobiel® data-base, the incidence of caries experience in fluorosis children is 25.8% compared to 29.4% in fluorosis-free children, with incidences of 23.3% and 27.7%, for boys and 27.9% and 31.2%, for girls respectively. Thus, apparently there seems to be only a slight difference in incidence between the two groups. However, the time-at-risk for caries experience was not taken into account and since the emergence of permanent teeth might be delayed because of fluoride-intake, our analysis should take into account of this imbalance at baseline.

< Figure 1 about here. >

To take the time-at-risk for caries development into account the emergence of that tooth should define time zero. At the onset of the study (i.e. for most children at the age of 7 years) about 86% of the permanent first molars had already emerged (left-censored). This fact makes it difficult to estimate efficiently their emergence times and hence will also have an effect on the efficiency with which the effect of fluoride-intake is estimated. We tried two strategies to improve the efficiency of our estimation procedure. Firstly, we included in our analysis also the emergence times of 8 other permanent teeth (the first premolars, the upper lateral incisors and the lower canines). By incorporating teeth which emerged during our observation period and using the association among

teeth of the same subject (via the concept of "the birth time of dentition", see next section), it was hoped to better estimate the true emergence time of the permanent first molars. Secondly, a Finnish longitudinal dataset, see Virtanen (2001) involving 235 boys and 223 girls born in 1980–1981 with follow-up from 6 to 18 years was added to our Flemish data. For these data the emergence of almost all 28 permanent teeth happened during the study period.

Since the emergence times are left-, right- or interval censored and the onset of caries experience of the four first permanent molars is right- or interval censored, we are dealing with multivariate doubly interval censored data. Hence, some teeth were used only to improve the estimation of the emergence times of the permanent first molars. Thus our Bayesian model needs to accommodate a high amount of left-censoring among otherwise interval-censored data in a multivariate manner.

3. THE BAYESIAN SURVIVAL MODEL FOR INTERVAL CENSORED DATA

The non-parametric Bayesian intensity model of Härkänen et al. (2000) provides a flexible tool for analyzing multivariate survival data. Left-, right- and interval censored survival times can be modelled. Further, a software package written in C, called BITE, allows to perform the analyzes in practice. Here, we used it to model the information about the tooth emergence and caries experience to assess the dependence of risk for caries experience with fluoride-intake. Furthermore, the model and the software allow to combine data from two sources, i.e. from the Signal Tandmobiel® study and the Finnish study (Virtanen 2001).

We now describe the model of Härkänen et al. (2000) adapted to solve our research question. We start with the notation and then move on to the specification of the survival intensity models for emergence and caries experience, respectively.

3.1. **Notation.** Let i denote the child ($i = 1, \ldots, N$) and let $j = (\kappa, \nu)$ be the double index indicating the position of the tooth in the mouth, following the standard European dentist notation (see Figure 1). Note that $\kappa \in \{1, 2, 3, 4\}$ denotes the quadrant of the mouth (upper right and left,

and lower left and right, respectively), and $\nu \in \{1, \ldots, 8\}$ indexes the teeth from front to back. Let a_{ij} be the age at which tooth j of subject i emerged and b_{ij} the age at which it experienced caries for the first time. Remember that none of the times a_{ij} and b_{ij} were observed exactly. Indeed the data are interval, left or right censored relative to the examinations performed at ages $u_{i1} < \cdots < u_{ik_i}$, $k_i \leq 6$.

3.2. **Model for emergence.** The (survival) model for emergence of permanent teeth is obtained from specifying the conditional intensity of emergence. Namely, here it is assumed that the intensity of emergence can be expressed as follows:

$$\lambda_{ij}^{(e)}(t) = f(t - \eta_i | \mathsf{tooth}_j, \mathsf{gender}_i) \times I[\eta_i < t \leq a_{ij}], \tag{3.1}$$

where gender_i is a dichotomous variate ($\mathsf{gender}_i = 0$ for boys, 1 for girls) and tooth_j is a tooth-specifying covariate – label (see further). The dependence between emergence times of one child is accounted for by using a subject specific variable η_i called *birth time of dentition*. This is a latent variable which represents the common time which marks the onset of the tooth eruption process and hereby "explains" the positive correlation between eruption times a_{ij} within a subject. Note that η_i almost surely does not equal to the first emergence time of the permanent teeth. We do admit however that the "birth time of dentition" concept is not commonly used by the dentists. It is merely a term pertaining to a subject specific property. Of course, the emergence of the different teeth j will in general be different, but always occurring after η_i. The intensity of emergence for a particular child is zero before that time. The indicator $I[\eta_i < t \leq a_{ij}]$ expresses the fact that a tooth is at risk for eruption only before it emerges. Based on some preliminary modelling we assume so called *horizontal symmetry*, i.e. we do not make distinction between left and right teeth coming from the same position of one jaw. So that $\mathsf{tooth}_{(1,\nu)} = \mathsf{tooth}_{(2,\nu)}$ and $\mathsf{tooth}_{(3,\nu)} = \mathsf{tooth}_{(4,\nu)}$ with $\mathsf{tooth}_{(1,6)} = \mathsf{tooth}_{(2,6)} = 0$, $\mathsf{tooth}_{(3,6)} = \mathsf{tooth}_{(4,6)} = 1$, $\mathsf{tooth}_{(1,4)} = \mathsf{tooth}_{(2,4)} = 2$ etc. However, we do make distinction between maxillary and mandibular teeth and also between teeth from

different positions of a quadrant of the mouth. No specific parametric expression is assumed for the (baseline) hazard functions $f(\cdot|\mathsf{tooth}_j, \mathsf{gender}_i)$, but for estimation purposes it will be defined as piecewise constant with support on the subset of non-negative real numbers.

3.3. **Model for caries experience.** The intensity for the caries process is given by

$$\lambda_{ij}^{(c)}(t) = Z_i \times h(t - a_{ij}|\mathsf{tooth}_j, \mathsf{gender}_i, \mathsf{fluor}_i) \times I[a_{ij} < t \leq b_{ij}], \tag{3.2}$$

where fluor_i takes value 1 for children with diagnosed fluorosis and 0 otherwise. Variable Z_i is an unknown subject-specific frailty coefficient which modulates the hazard function and which reflects a common environment shared by teeth of one person. Again no parametric assumptions are made for the functions $h(\cdot|\mathsf{tooth}_j, \mathsf{gender}_i, \mathsf{fluor}_i)$ while assuming them piecewise constant for the calculations. Note that the difference $b_{ij} - a_{ij}$ is our quantity of primary interest and we can call it *the caries-free time.*

Our statistical model will involve the above two measurement models. Hence the possible dependencies among times of interest are taken into account by involving subject-specific parameters η_i and Z_i. The first subject-specific parameter η_i is included in the model for the emergence and will shift the hazard function in time, reflecting the idea that evolution of a particular tooth type is similar for all children and different children only vary in the onset of emergence process. On the other hand, a multiplicative frailty Z_i included in the model for caries experience enables teeth of one child to be more sensitive to caries (failure) than corresponding teeth of another child reflecting different dietetic behavior, brushing habits, etc.

The covariate "fluorosis" will be incorporated in the caries model in two ways. In the first step, for each combination of values of fluor, gender and tooth a hazard function is specified and fitted. Secondly, a proportional hazards assumption is made regarding the effect of the covariates on the caries experience by replacing the term $h(\cdot|\mathsf{tooth}_j, \mathsf{gender}_i, \mathsf{fluor}_i)$ in the measurement model for

caries by $h(\cdot) \times \exp(\boldsymbol{\beta}^T \text{covariate vector}_{ij})$ thus assuming a proportional hazards model for caries. To this end we verified the proportional hazards assumption.

3.4. **Priors for baseline hazard functions.** In BITE the working assumption is that the hazard functions are piecewise constant. Further, for the emergence hazard functions $f(\cdot|\mathsf{tooth}_j, \mathsf{gender}_i)$ the first level of the piecewise constant hazard function is assumed to have a priori a Gamma distribution and the increments of the levels also. This will ensure a priori an increasing hazard function for emergence. In the case of caries experience the first level of the hazard functions $h(\cdot|\mathsf{tooth}_j, \mathsf{gender}_i, \mathsf{fluor}_i)$ (let say h_0) is assumed to have a Gamma prior and conditional on the previous levels $h_0, \ldots, h_{m-1}$ the subsequent level h_m follows a priori a Gamma$(\alpha, \alpha/h_{m-1})$ distribution. This gives a priori $E[h_m|h_{m-1}, \ldots, h_0] = h_{m-1}$ and assures that there is no built-in prior assumption of trend of the hazard rate. Finally, the prior distribution of the jump points of the piecewise constant functions are assumed to arise as a result of a homogeneous Poisson process. A more detailed description of these prior distributions can be found in Arjas and Gasbarra (1994).

3.5. **Priors for the random effect terms.** The prior distribution for the birth time of dentition η_i illustrates how we have combined the Signal Tandmobiel® data and the Finnish data. It also shows how the information about the emergence of the Finnish data is included in our analysis. Indeed, we assume that the shapes of the emergence hazard functions f for Finland and Flanders are the same, but we do allow for a shift in emergence times by assuming different means for the birth time of dentition in the two countries. More precisely, the prior distribution of η_i is assumed normal $\mathrm{N}(\xi_0, \tau^{-2})$ when the ith child comes from Finland and is assumed normal $\mathrm{N}(\xi_1, \tau^{-2})$ for a Flemish child. This prior assumption was partly driven by some earlier work on emergence times of permanent teeth (Nanda 1960).

The Bayesian approach allows to include dentist's knowledge on the problem at hand in the form of a prior distribution for the parameters ξ_0 and ξ_1 which are assumed to have a priori a normal distribution with mean 5.2 years and standard deviation 1 year. This choice is motivated by the

results found in the literature on the earliest emergence of permanent teeth (see for instance Nanda 1960) and reflects the dentist's belief that permanent teeth on average emerge slightly after 5 years of age. The parameter τ^2 is assumed a priori to have a Gamma(2,2) distribution.

The individual frailties Z_i in the model for caries are a priori assumed to be conditional on the hyper-parameter ϕ, independent and identically Gamma distributed with both shape and inverse scale equal to that hyper-parameter. The hyper-parameter itself is then given a prior distribution Gamma(2,2).

3.6. **Treatment of censored data.** Left- and interval censoring are treated by data augmentation, see Tanner and Wong (1987). First of all, the left censored emergence times of all our teeth are changed into interval censored emergence times with a lower limit equal to 4 years. This reflects our prior belief that no permanent teeth can emerge before that age. Further, all interval censored event times are sampled within each iteration of the Monte Carlo Markov Chain simulation from the current state of the conditional distribution of the particular event conditioned further on the interval where the event was observed. If either of event times is right censored it is assumed to be fixed, it is not augmented and the likelihood contribution of such observation is of the common form for right censored observations.

4. BAYES INFERENCE OF MODEL COMPONENTS

The model in the previous section involves many unobservables and does not use conjugate priors implying the need to calculate the posterior distributions using Markov Chain Monte Carlo sampling techniques, see for example Gilks et al. (1996). Numerical work is done using versions of the Metropolis-Hastings-Green algorithm, see Green (1995) and the adaptive Metropolis algorithm, see Haario et al. (2001).

4.1. **Posterior distribution.** The form of the joint posterior distribution of all parameters involved in the model can be written using successive conditioning since the model has a hierarchical

structure. Let us denote $\boldsymbol{\nu}$ all parameters involved in a definition of hazard functions f and h (i.e. the times where these functions jump and their increments or levels respectively), $\boldsymbol{\theta}$ hyper-parameters $\xi_0, \xi_1, \tau^2, \phi$ and $\boldsymbol{\zeta}$ the random effects Z_i and η_i. Let further $\boldsymbol{u}_{ij} = \{u_{ij}^1, u_{ij}^2, u_{ij}^3, u_{ij}^4\}$ denote the endpoints of the intervals where emergence (u_{ij}^1, u_{ij}^2) and caries (u_{ij}^3, u_{ij}^4) of the jth tooth of the ith child were observed. Assign $u_{ij}^2 = \infty$, respectively $u_{ij}^4 = \infty$ if either of the event times was right censored.

Likelihood contributions of augmented or right censored emergence times a_{ij} are then given by

$$\begin{aligned} L_{ij}^{(e)} &= p(a_{ij}|\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\zeta}) = p(a_{ij}|\boldsymbol{\nu}, \boldsymbol{\zeta}) = \\ &= \big\{ f(a_{ij} - \eta_i | \mathsf{tooth}_j, \mathsf{gender}_i) \big\}^{\delta_{ij}} \exp\Big\{ - \int_{\eta_i}^{a_{ij}} f(s - \eta_i | \mathsf{tooth}_j, \mathsf{gender}_i)\, ds \Big\}, \end{aligned} \tag{4.1}$$

the likelihood contributions of the (augmented) observations from the caries process are

$$\begin{aligned} L_{ij}^{(c)} &= p(b_{ij}|\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\zeta}) = p(b_{ij}|\boldsymbol{\nu}, \boldsymbol{\zeta}) = \\ &= \big\{ Z_i h(b_{ij} - a_{ij} | \mathsf{tooth}_j, \mathsf{gender}_i, \mathsf{fluor}_i) \big\}^{\omega_{ij}} \exp\Big\{ - Z_i \int_{a_{ij}}^{b_{ij}} h(s - a_{ij} | \mathsf{tooth}_j, \mathsf{gender}_i, \mathsf{fluor}_i)\, ds \Big\}, \end{aligned} \tag{4.2}$$

with δ_{ij} (ω_{ij}) equal to one if the emergence (caries) has been observed and equal to zero if the emergence time a_{ij} (caries time b_{ij}) is right censored. Note that right censored events are kept fixed during MCMC iterations and satisfy $a_{ij} = u_{ij}^1$, $b_{ij} = u_{ij}^3$ respectively. For interval censored emergence times a_{ij} the posterior distribution given the hazard function, the random effects, the hyper-parameters and the observed interval (u_{ij}^1, u_{ij}^2) is obtained by Härkänen et al. (2000) and is proportional to

$$p(a_{ij}|\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{u}_{ij}) = p(a_{ij}|\boldsymbol{\nu}, \boldsymbol{\zeta}, \boldsymbol{u}_{ij}) \propto L_{ij}^{(e)} \cdot I[u_{ij}^1 < a_{ij} \leq u_{ij}^2]. \tag{4.3}$$

Similarly, the posterior distribution of the interval censored caries times is proportional to

$$p(b_{ij}|\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{u}_{ij}) = p(b_{ij}|\boldsymbol{\nu}, \boldsymbol{\zeta}, \boldsymbol{u}_{ij}) \propto L_{ij}^{(c)} \cdot I[u_{ij}^3 < b_{ij} \leq u_{ij}^4]. \tag{4.4}$$

The joint posterior distribution of all random quantities in the model given the observed intervals is then proportional to the expression

$$p\big(\{a_{ij}\},\{b_{ij}\},\boldsymbol{\nu},\boldsymbol{\theta},\boldsymbol{\zeta}\big|\{\boldsymbol{u}_{ij}\}\big) \propto \tag{4.5}$$

$$\propto \prod_{i,j}\big[p(a_{ij}|\boldsymbol{\nu},\boldsymbol{\zeta},\boldsymbol{u}_{ij})\,p(b_{ij}|\boldsymbol{\nu},\boldsymbol{\zeta},\boldsymbol{u}_{ij})\big] \times$$

$$\times\, p(\boldsymbol{\nu}) \times \prod_{i}\big[p(Z_i|\phi)\,p(\eta_i|\xi_0,\xi_1,\tau^2)\big] \times p(\phi)\,p(\xi_0)\,p(\xi_1)\,p(\tau^2).$$

5. RESULTS

We ran twice 20 000 iterations of burn-in, and, in addition, 14 000 iterations with a 1:4 thinning to obtain a sample from the posterior distribution. The Gelman and Rubin (1992) test and autocorrelation plots have been used to check for convergence.

< Figure 2 about here. >

To evaluate the effect of fluoride-intake on the development of caries on the permanent first molars we have calculated the posterior expectations of hazard ratios $\frac{h(t|\mathsf{tooth},\mathsf{gender},\text{fluorosis})}{h(t|\mathsf{tooth},\mathsf{gender},\text{fluorosis free})}$. These hazard ratios together with their 95% equal tail pointwise credibility intervals can be found in Figure 2. Proportional hazards assumption with respect to covariate fluor seems to be satisfied since credibility intervals cover in all cases a horizontal line. In three cases, this horizontal line is close to the line $y = 1$ implying no effect of fluoride-intake on caries development. A positive effect of fluoride-intake seems to be present only for mandibular permanent first molars of boys. There are no deviations from the proportional hazards assumption also with respect to covariates gender and tooth (plots are not shown). Hence, there are no serious deviations from the (conditional) proportional hazards assumption with respect to all three covariates. This allowed us to assume for the caries model a proportional hazards effect of the three covariates, possibly including some interaction terms. By this semi-parametric assumption it was hoped to see more clearly the effect of fluoride-intake on caries experience.

That is why we have fitted a model where the caries hazard function (3.2) was changed into

$$h(\cdot) \times \exp(\beta_1 \mathsf{fluor}_i + \beta_2 \mathsf{gender}_i + \beta_3 \mathsf{tooth}_j + \beta_4 \mathsf{fluor}_i \times \mathsf{gender}_i + \beta_5 \mathsf{fluor}_i \times \mathsf{tooth}_j) \times \tag{5.1}$$

$$\times I[a_{ij} < t \leq b_{ij}].$$

The additional β-parameters were given a flat normal prior with zero mean and variance equal to 100. Observe though that the hazard function for emergence is still non-parametric and defined by (3.1). This model leads to the posterior expectations of β parameters given in Table 1. Posterior expectations of the hazard ratios between the fluorosis and the fluorosis-free group of children while fixing other covariates are given in the left columns of Table 2.

< Table 1 about here. >

The semi-parametric analysis using the Cox proportional model for caries gives similar conclusions to the completely non-parametric model. A positive effect of fluoride-intake is clearly seen for the mandibular permanent first molars of boys and has a borderline positive effect for the maxillary permanent first molars of boys. However, no effect of fluoride-intake was seen for girls.

We have also computed posterior predictive caries-free curves, see Arjas and Gasbarra (1996) for a new tooth of a child of a particular gender and with a given value of fluorosis. Suppose that T denotes the tooth lifetime, the posterior predictive caries-free function is then given by

$$\hat{S}(t; \mathsf{tooth}_j, \mathsf{gender}_i, \mathsf{fluor}_i) = P(T > t | \mathsf{tooth}_j, \mathsf{gender}_i, \mathsf{fluor}_i, \text{data}). \tag{5.2}$$

Its values can be approximated from the resulting Monte Carlo Markov chain of length M by the expression

$$\begin{aligned}\hat{S}(t;\mathsf{tooth}_j,\mathsf{gender}_i,\mathsf{fluor}_i) &\approx \frac{1}{M}\sum_{m=1}^{M} P(T>t|\boldsymbol{\nu}^{(m)},\boldsymbol{\theta}^{(m)},\boldsymbol{\zeta}^{(m)},\mathsf{tooth}_j,\mathsf{gender}_i,\mathsf{fluor}_i)\\ &= \frac{1}{M}\sum_{m=1}^{M}\exp\Big(-Z^{(m)}e^{\beta_1^{(m)}\mathsf{fluor}_i+\beta_2^{(m)}\mathsf{gender}_i+\beta_3^{(m)}\mathsf{tooth}_j+\beta_4^{(m)}\mathsf{fluor}_i\mathsf{gender}_i+\beta_5^{(m)}\mathsf{fluor}_i\mathsf{tooth}_j}\times \\ &\qquad \times\int_0^t h^{(m)}(u)\,du\Big).\end{aligned} \tag{5.3}$$

Note that the posterior predictive caries-free function takes also the distribution of the frailty parameter into account. The corresponding plots can be found in Figure 3 and the difference between the posterior predictive caries-free functions of fluorosis and fluorosis free children in Figure 4.

< Figures 3 and 4 about here. >

Further, the posterior expectations and 95% credibility intervals of the hyper-parameters related to the birth times of dentition η_i and frailties Z_i are given in Table 3. The non-parametric solution and Cox regression model for caries give similar results. The emergence process starts slightly earlier in Finland (by approx. 2.5 months) than in Belgium, this conclusion is driven by the difference of the posterior expectations of the means of birth time of dentition which is about 0.2 years. Further, we can say that results are not sensitive to the choice of the prior mean for hyper-parameters μ_1 and μ_0 (means of birth time of dentition in Belgium and Finland). We have tried several values with a negligible effect to results. A value of 5.2 presented here wanted to highlight the fact that we have included a prior belief given by dentists that first permanent teeth may not emerge earlier than shortly after 5 years of age.

< Table 3 about here. >

Finally, to see the influence of including the Finnish data set we have also fitted the Cox regression model for caries using only the Belgian data. Figure 5 gives a comparison of 95% pointwise equal tail credibility regions for the emergence hazard functions of the permanent first molars based on

the analysis with both datasets and the Belgian dataset only. The credibility regions are somewhat narrower when both databases are used. However, posterior means and credibility intervals for the caries hazard ratios are almost unchanged as can be seen from Table 2.

< Table 2 about here. >

6. DISCUSSION

We have tackled a relatively simple dental question implying however quite complex Bayesian analysis taking into account multivariate doubly interval censored data with a huge amount of left censoring times. The approach of Härkänen et al. (2000) employing data augmentation in a Bayesian way can deal with the problem of interval censored data in an elegant way. The multivariate character of the data was accounted for by using subject-specific parameters in a model. Finally, the high degree of left censoring of the emergence of the permanent first molars was tackled by incorporating the additional Finnish data set in the estimation procedure.

The use of advanced Bayesian methods is often hampered by the lack of appropriate software. The WinBUGS package, see Spiegelhalter et al. (2000) is often the choice in practice. The WinBUGS software could also be used here but only for a simplified version of our problem. Namely, a particular parametric distribution needs to be specified for the events of interest. Additionally, the effect of covariates needs to be included in a parametric way.

The BITE package of Härkänen (2001) gave us the possibility to avoid the specification of a parametric model, while keeping the opportunity to check e.g. the proportional hazards assumption. Random effects accounting for dependencies between teeth of one child were easily included and it was possible to change the non-parametric model into a semi-parametric one with a proportional hazards assumption.

ACKNOWLEDGMENTS

The study was financially supported by Research Grant OT/00/35, Catholic University Leuven. 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). The first two authors acknowledge support from the Interuniversity Attraction Poles Program P5/24 – Belgian State – Federal Office for Scientific, Technical and Cultural Affairs.

REFERENCES

Arjas, E., and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica,* **4,** 505–524.

Arjas, E., and Gasbarra, D. (1996). Bayesian inference of survival probabilities, under stochastic ordering constraints. *Journal of the American Statistical Association,* **91,** 1101–1109.

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

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice.* London: Chapmann & Hall.

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

Haario, H., Saksman, E. and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli,* **7,** 223–242.

Härkänen, T. (2001). BITE: A Bayesian intensity estimator. *Submitted to publication.* `http://www.rni.helsinki.fi/~tth`

Härkänen, T., Virtanen, J. I., and Arjas, E. (2000). Caries on permanent teeth: A nonparametric Bayesian analysis. *Scandinavian Journal of Statistics,* **27,** 577–588.

Leroy, R., Bogaerts, K., Lesaffre, E., and Declerck, D. (2002). Correlation between fluorides, caries and tooth emergence. Fact or fiction? *Submitted to publication in Community Dent. Oral Epidemiol.*

Nanda, R. S. (1960). Eruption of human teeth. *American Journal of Orthodontics,* **46,** 363–378.

Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. R. (2000). *WinBUGS.* Medical Research Council Biostatistics Unit, Cambridge.

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.

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.

Vanobbergen, J., Martens, L., Lesaffre, E., Bogaerts, K., and Declerck, D. (2001). Assessing risk indicators for dental caries in the primary dentition. *Community Dentistry and Oral Epidemiology,* **29,** 424–434.

Virtanen, J. I. (2001). Changes and trends in attack distributions and progression of dental caries in three age cohorts in Finland. *Journal of Epidemiology and Biostatistics,* **6,** 325–329.

TABLE 1. Posterior means and 95% equal tail credibility intervals for the regression parameters based on the Cox regression model.

Effect	Posterior mean	95% equal tail credibility interval
Fluorosis (β_1)	-0.41	(-0.71, 0.01)
Gender (β_2)	0.11	(-0.11, 0.38)
Jaw (β_3)	0.23	(0.13, 0.32)
Fluorosis*gender (β_4)	0.39	(-0.17, 0.81)
Fluorosis*jaw (β_5)	-0.17	(-0.33, 0.00)

TABLE 2. Posterior means and 95% equal tail credibility intervals of hazard ratios between a fluorosis and fluorosis free group of children while fixing all other covariates based on the Cox regression model. Comparison of the analysis with both Belgium and Finnish datasets (before slash) and Belgium dataset only (after slash).

Group	Posterior mean	95% equal tail credibility interval
Boys, maxilla	0.67/0.65	(0.49, 1.01)/(0.46, 0.96)
Boys, mandible	0.57/0.55	(0.41, 0.85)/(0.39, 0.78)
Girls, maxilla	0.99/1.00	(0.72, 1.36)/(0.70, 1.33)
Girls, mandible	0.84/0.84	(0.61, 1.14)/(0.60, 1.13)

TABLE 3. Posterior means and 95% equal tail credibility intervals for the hyperparameters.

Hyper-parameter	Non-parametric model	Cox regression model
Expectation of η_i for Finland (μ_0)	5.47 (5.40, 5.54)	5.45 (5.38, 5.52)
Expectation of η_i for Belgium (μ_1)	5.69 (5.64, 5.73)	5.68 (5.64, 5.73)
Variance of η_i (τ^{-2})	0.48 (0.45, 0.52)	0.49 (0.45, 0.52)
Variance of frailties Z_i (ϕ^{-1})	3.85 (3.57, 4.17)	4.00 (3.57, 4.35)

FIGURE 1. Permanent teeth. Terminology and notation. Maxillary permanent first molars are labeled as 16, 26, mandibular permanent first molars as 36, 46. Other permanent teeth used in the analysis were: maxillary lateral incisors (12, 22), maxillary first premolars (14, 24), mandibular canines (33, 43) and mandibular first premolars (34, 44).

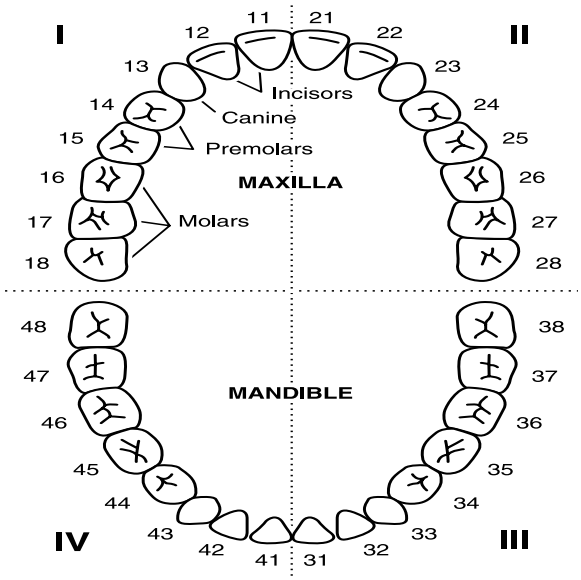


FIGURE 2. Posterior means of the hazard ratios between the fluorosis free and fluorosis groups (solid line), 95% point-wise equal tail probability region (dashed line) and a horizontal line indicating no effect of fluorosis on caries development (dotted-dashed line). Results based on the non-parametric model with Belgium + Finnish data.

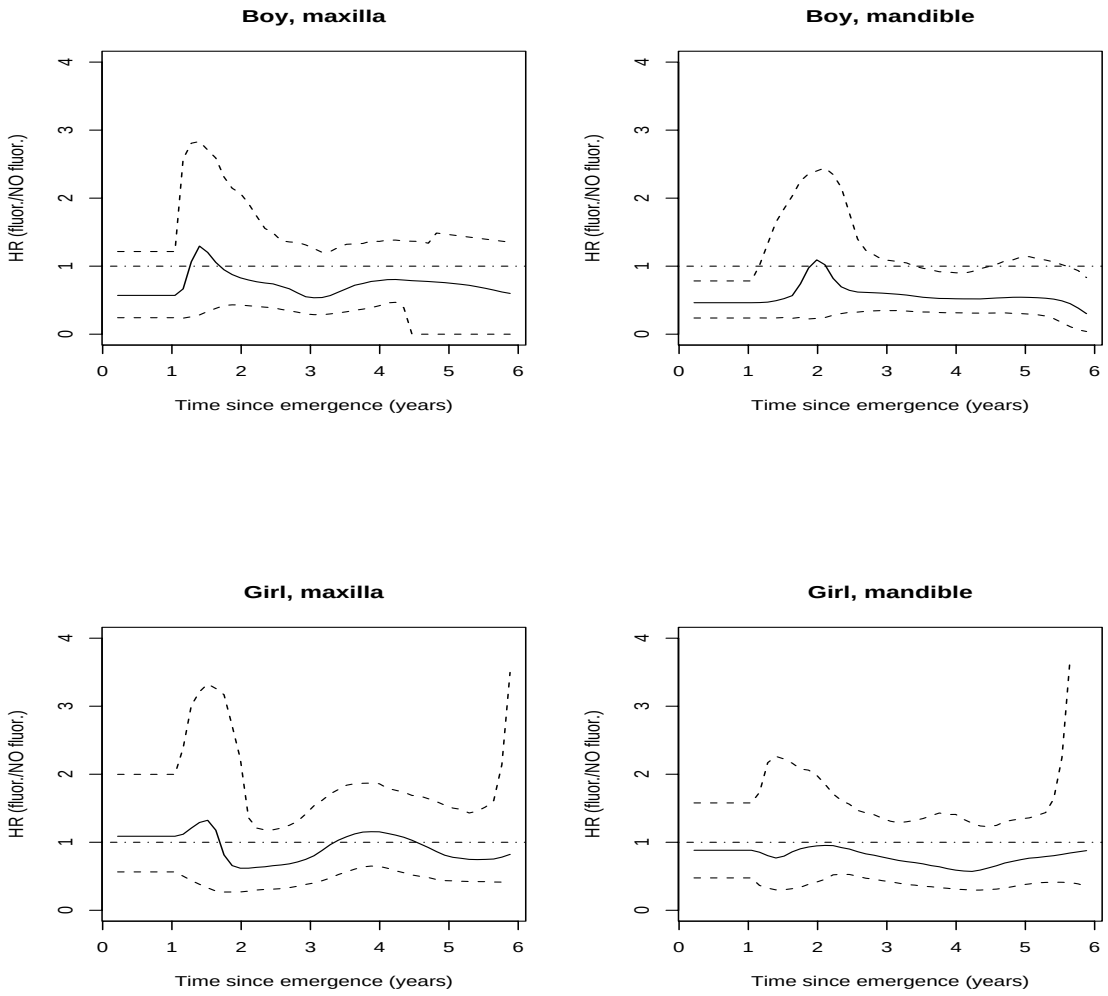


FIGURE 3. Posterior predictive caries-free functions for the four groups of children and the two jaws based on the Cox regression model with Belgium + Finnish data.

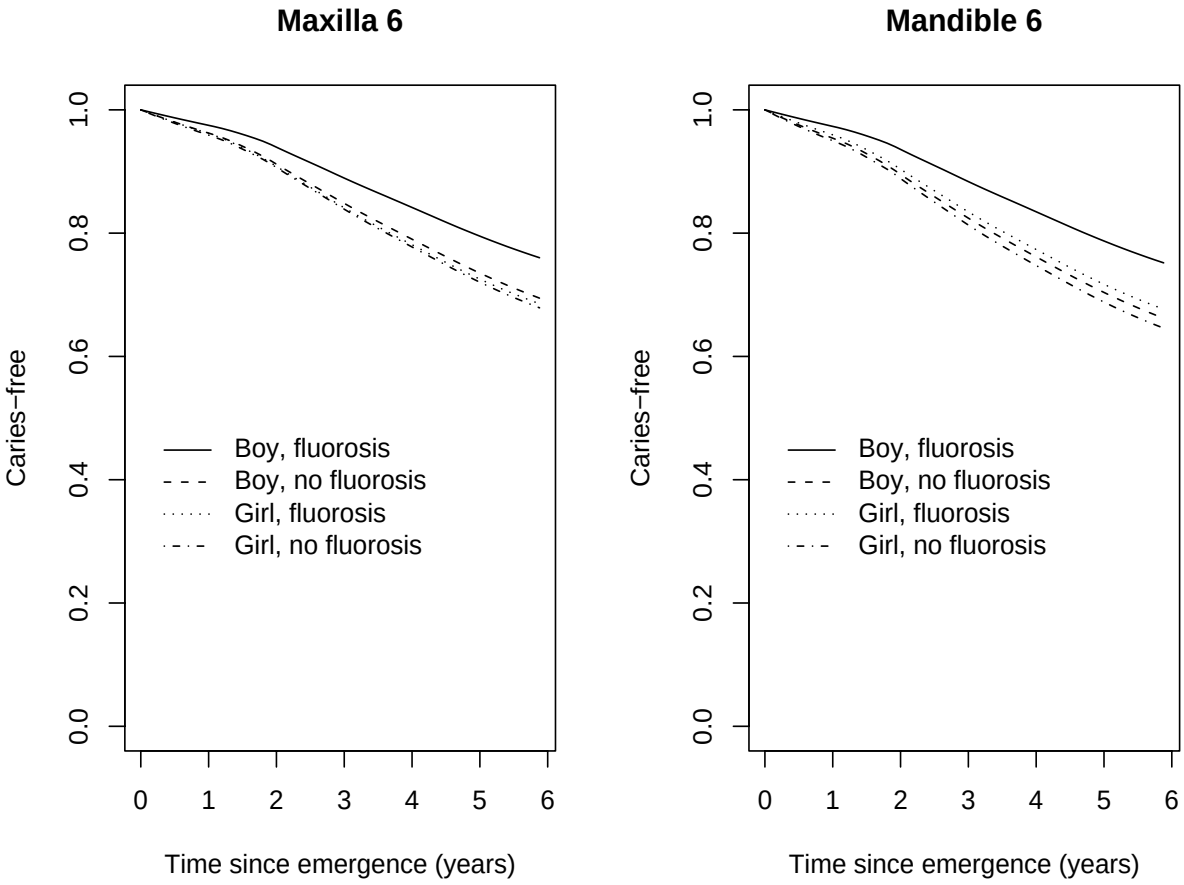


FIGURE 4. Difference between the two posterior predictive caries-free functions based on the Cox regression model with Belgium + Finnish data.

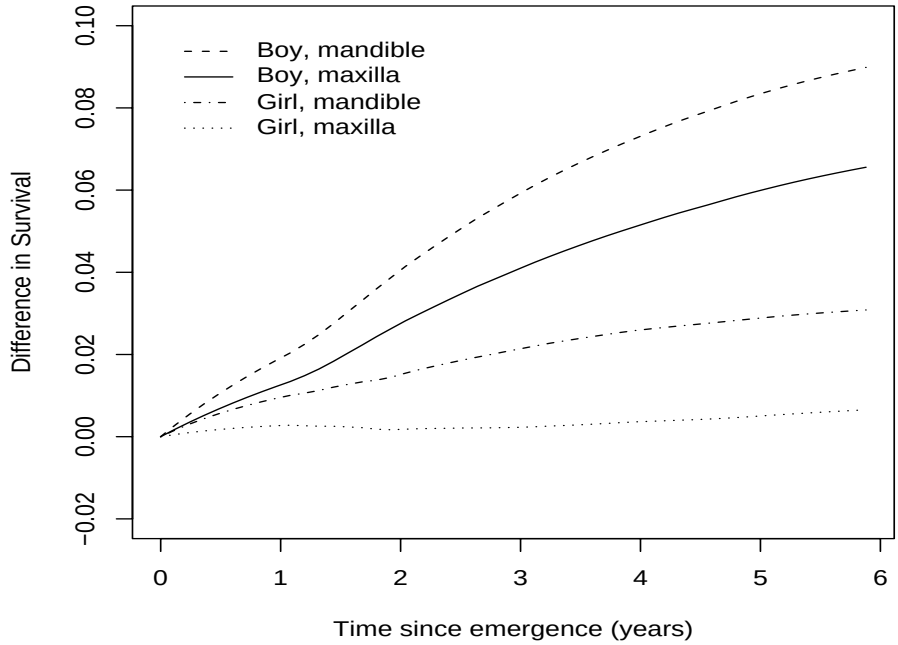


FIGURE 5. Posterior meanss of the emergence hazard functions $f(\cdot|\mathsf{tooth}, \mathsf{gender})$ for the permanent first molars together with their 95% pointwise equal tail probability regions. Comparison of the analyzes with and without additional Finnish data. Results based on the Cox regression model. Solid line for post. mean based on Belgium + Finnish data, dashed line for post. mean based on Belgium data only, dotted-dashed line for 95% prob. region based on Belgium + Finnish data and dotted line for 95% prob. region based on Belgium data only.