© Board of the Foundation of the Scandinavian Journal of Statistics 2000. Published by Blackwell Publishers Ltd, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA Vol 27: 577–588, 2000

Caries on Permanent Teeth: A Non-parametric Bayesian Analysis*

TOMMI HÄRKÄNEN

University of Helsinki

JORMA I. VIRTANEN

University of Oulu

ELJA ARJAS

University of Helsinki

ABSTRACT. Most earlier epidemiological investigations of dental caries have been based on cross-sectional data. Subject-specific information of dental caries in the past, and the duration of exposure of each tooth to the oral environment, are obviously important factors also influencing the presence of dental caries in the future. This has led us to consider multivariate survival models in which the information about the tooth eruption and failure times are combined to assess caries risk. A non-parametric Bayesian intensity model is presented, reflecting, on the one hand, the within subject and between subject sources of variability, and a corresponding split of variability when considering the 28 permanent teeth. We analyse a data set consisting of the dental history of 240 boys, where the observations are based on predetermined dental examinations taking place approximately once every year. Markov chain Monte Carlo integration techniques are applied in the numerical work.

Key words: frailty model, intensity model, Markov chain Monte Carlo method, measurement model, multivariate survival analysis

1. Introduction

Epidemiological investigations of dental caries are commonly based on cross-sectional data from different age groups. Such data do not provide direct means for determining the length of the time during which each individual tooth was exposed to risk, which would be a necessary prerequisite for understanding how the cariotic process of children and adolescents develops in time. From a statistical perspective, study of dental caries gives rise to a number of interesting issues. First, even if there is longitudinal data, it is in practice always interval censored, providing only tooth-by-tooth current status information as determined in a sequence of dental examinations. Second, both the eruption and cariotic processes are likely to lead to correlated incidence times across different teeth of the same subject. And third, the incidence times are strongly dependent on the tooth. These considerations lead us to expressing the tooth eruption and caries attack times in terms of two nested multivariate intensity models.

The dependence of the lifetimes between the teeth of a subject are here accounted for by introducing subject-specific frailty parameters (see for example, Clayton (1978, 1991). Given such frailty parameters, the tooth lifetimes across different subjects are assumed to be independent, with intensities which are defined non-parametrically by using piecewise constant functions as approximations (see Arjas & Gasbarra, 1994), and which are generally allowed to

*This paper was presented as an Invited Lecture at the 17th Nordic Conference on Mathematical Statistics, Helsingor, Denmark, June 1998.

depend on the tooth. Finally, the interval censoring is handled by augmenting the data with (unobserved) exact incidence times and by introducing a corresponding simple current status measurement model.

The model parameters are estimated by using Bayesian inferential methods. Since the model involves very many unobservables, and we cannot use conjugate priors, we calculate the (marginal) expectations numerically by using Markov chain Monte Carlo integration techniques (see for example Gilks *et al*., 1996).

The contents of this paper are as follows: in section 2, the data set and the relevant notions of dentistry are described. In section 3, the hierarchical model, the prior distributions, and the estimation procedure are explained. The empirical results are presented in section 4, and the paper ends with discussion in section 5.

2. The data

The data set consists of a series of longitudinal observations on 240 boys, born in 1970 and 1971 in three rural municipalities in Finland. They were examined annually by municipal dentists according to a prespecified protocol, and followed approximately from age 6 to age 18 (Virtanen *et al*., 1996).

In the check-ups a tooth-by-tooth examination was made and the status of each tooth (not erupted, erupted but not needing restoration, erupted and needing restoration, previously restored or extracted) was recorded. In this study, a permanent tooth is considered to have *erupted* if any part of the crown has penetrated the gingiva. The time at which dental caries has progressed through dental enamel (after which a dentist would normally deem the tooth as needing restoration) is called here the failure time of the tooth. Further, the time interval between eruption and failure is called here the *lifetime of a tooth*. The term *dental age* of a tooth refers to its developmental age, that is, the time it has been exposed to risk for caries in the oral cavity.

Let i index the 240 subjects, $i = 1, 2, \ldots, 240$. We follow the standard practice in dentistry indicating each of the 28 teeth by a double index $j = (\kappa, \nu)$, where $\kappa \in \{1, 2, 3, 4\}$ denotes the (upper right and left, and lower left and right, respectively) "quadrant" of the mouth, and $\nu \in \{1, 2, \ldots, 7\}$ indexes the teeth from front to back. The first two teeth in the front ($\nu = 1, 2$) are called incisors (I1 and I2), the next ($\nu = 3$) canines, then ($\nu = 4, 5$) premolars (B1 and B2), and finally ($\nu = 6, 7$) molars (M1 and M2). We use t as the generic notation for age in years and denote by c_i the age at which subject i was last examined (right censored).

Table 1. *The proportions of teeth* (κ, ν) *with at least one failure at the end of the follow-up. Columns* $(^*)$ *give the mean age at the time of examination in which tooth eruption was recorded.*

			Upper jaw			Lower jaw		
				Failure frequency			Failure frequency	
Tooth		ν	*	Right $\kappa = 1$	Left 2	*	Left $\kappa = 3$	Right 4
Incisors	I1	1	7.6	0.167	0.188	6.8	0.017	0.025
	I2	2	8.6	0.229	0.208	7.8	0.004	0.008
Canines	C	3	11.9	0.063	0.046	11.1	0.029	0.008
Premolars	B1	4	11.5	0.108	0.104	11.4	0.042	0.058
	B2	5	12.4	0.121	0.117	12.3	0.108	0.138
Molars	M1	6	7.1	0.596	0.596	6.9	0.700	0.671
	M2	7	12.9	0.358	0.412	12.4	0.454	0.467

Let a_{ij} be the age at which tooth j of subject i, henceforth indexed by (i, j), erupted, and b_{ij} the age at which it failed. The lifetime of a tooth (i, j) is the difference $d_{ij} := b_{ij} - a_{ij} > 0$. A sketch of such a tooth history is presented in Fig. 1. Terms "Model e", "Model c", and notation η_i will be defined in section 3.

In our data set, none of these variables were observed exactly. Instead, the data consist of the findings made by dentists in a series of examinations. Denote the ages at which subject i was examined by $u_{i1} < \cdots < u_{ik_i} := c_i$. These examinations took place approximately once every year, the average age at the time of the first examination being 5.7 years. Making the convention that $u_{i0} \equiv 0$ and $u_{i,k_i+1} \equiv \infty$, we see that all true eruption and failure times are filtered into the data through the endpoints of the intervals $(u_{i,k-1}, u_{ik}]$. This gives rise to the following notation: the eruption of tooth (i, j) was recorded at age $u_{i,k(i,j)}$ (the first examination after eruption), where $k(i, j) := \min\{k\colon u_{ik} \geqslant a_{ij}\}$. Similarly, tooth (i, j) was recorded for the first time as having failed at age $u_{i,l(i,j)}$, with $l(i, j)$ defined by $l(i, j) := \min\{k\colon u_{ik} \geqslant b_{ij}\}$. Most of the failure times were right censored, that is, $b_{ij} > c_i$. Table 1 gives a brief summary of the data.

3. The model

3.1. Conditional intensities

The tooth history data can be grouped in two obvious ways, either according to the subject (index i) or according to the tooth in question (index j). Therefore, it is a natural requirement for a statistical model of tooth histories that there should be some control for the smaller "within subject" and "within tooth (type)" variances. In the following we propose a model which reflects such dependencies. In addition, we must take into account, through a simple measurement model, the fact that the true eruption and failure times were not recorded in the data.

There are several reasons contributing to the level of caries activity and the resulting tooth decay. Factors such as possible bacterial infections, aspects related to nutrition, or general oral hygiene, could be viewed as properties of the oral environment. Some risk factors are no doubt of a genetic origin. The genetic factors remain constant over time, whereas the environmental factors could vary considerably. These factors vary between subjects, but are shared to some degree by all teeth of the same subject.

It is not clear whether some of these genetic and environmental factors could actually be measured, either directly, or indirectly by using suitable surrogates. In the absence of such information in the data, however, we assume simply that all these characteristics are represented by a subject specific real valued (positive) frailty coefficient Z_i modulating all 28 tooth-specific failure intensities of the ith subject. Variability in the Z_i values across different subjects can then be understood as giving rise to the "extra-binomial" variance in the number of tooth restorations found at the end of the follow-up period.

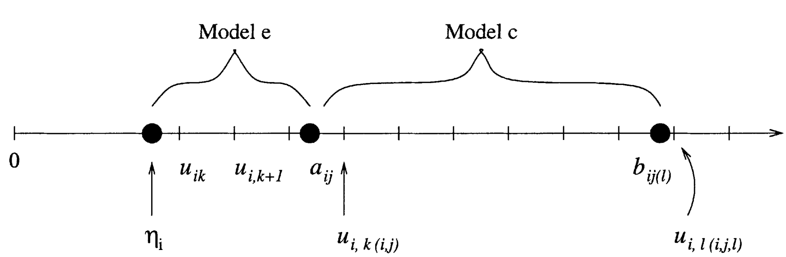

Fig. 1. An example of a tooth history, and the corresponding dental examinations.

But there are also considerable differences in the caries proneness of different teeth j, and consequently, in their lifetimes. The caries risk is well known to be high, particularly in the molar teeth, whereas, for example, canines and lower incisors would need restoration only rarely. These considerations lead us to postulate the following intensity model (Model c): Considering subject i at age t and the tooth indexed by j, we assume that the risk of failure can be expressed in the form

$$\lambda_{ij}^{(c)}(t) := h_j(t - a_{ij}) \cdot Z_i \cdot 1\{a_{ij} < t \leqslant b_{ij}\}. \tag{1}$$

Here $t - a_{ij}$ is the dental age, and h_j is a corresponding tooth-specific non-negative baseline hazard rate function. In the absence of clear-cut justification for parametric hazard models, we let the function h_j be defined non-parametrically. As stated above, Z_i is an unknown subject-specific frailty coefficient which modulates the baseline hazard, but which itself is assumed to remain constant over time. Finally, the indicator $1\{a_{ij} < t \leqslant b_{ij}\}$ simply expresses the fact that a tooth is at risk only after it has erupted and before it failed.

Subject matter considerations suggest a right-left symmetry between the baseline hazards. Hence, considering the indices $j = (1, 1), (1, 2), \ldots, (4, 7)$, we assume that

$$h_{(1,\nu)} \equiv h_{(2,\nu)} \text{ and } h_{(3,\nu)} \equiv h_{(4,\nu)}, \quad \nu = 1, 2, \ldots, 7. \tag{2}$$

As a result, 14 tooth specific baseline hazards h_j for failure will have to be considered in the model. Thus the conditional probability density specified by the intensity function (1) is

$$p\{b_{ij} | a_{ij}, h_j, Z_i\} = Z_i \cdot h_j(b_{ij} - a_{ij}) \exp\left\{ -Z_i \int_{a_{ij}}^{b_{ij}} h_j(s - a_{ij})\, ds \right\}. \tag{3}$$

Although the main interest here is the (joint) estimation of the baseline hazards h_j and the frailty coefficients Z_i, the inferential situation is complicated further by the fact that the exact values of the variables a_{ij} are also unknown. Therefore they need to be estimated as well, and for this we need a model for the tooth eruption times.

To do so, we introduce another set of subject specific variables $\eta_i (\leqslant a_{ij})$, calling them *birth times of detention*. The motivation is analogous to that justifying the introduction of a subject specific frailty parameter: the positive correlation found between eruption times a_{ij} within a subject is explained by the fact that they share a common time η_i starting the tooth eruption process. On the other hand, different subjects i will generally have different η_i values. And, of course, different teeth j, unless they are in a right–left symmetric position to each other, may take very different times after η_i until they erupt. We now postulate an intensity model (Model e) for the eruption of tooth j of subject i as follows:

$$\lambda_{ij}^{(e)}(t) := f_j(t - \eta_i) \cdot 1\{\eta_i < t \leqslant a_{ij}\}. \tag{4}$$

We again assume a right–left symmetry: $f_{(1,\nu)} \equiv f_{(2,\nu)}$ and $f_{(3,\nu)} \equiv f_{(4,\nu)}$ for all $\nu = 1, 2, \ldots, 7$, and therefore need 14 tooth specific non-parametric baseline hazards f_j to describe the eruption process. The probability densities of the eruption, are thus

$$p\{a_{ij} | f_j, \eta_i\} = f_j(a_{ij} - \eta_i) \exp\left\{ -\int_{\eta_i}^{a_i} f_j(s - \eta_i)\, ds \right\}, \tag{5}$$

The joint density of the eruption time a_{ij} and the failure time b_{ij} of a tooth (i, j) is then the product of (3) and (5):

$$p\{a_{ij}, b_{ij} | f_j, h_j, Z_i, \eta_i) = p\{b_{ij} | a_{ij}, h_j, Z_i\} \times p\{a_{ij} | f_j, \eta_i\},$$

The times of the dental examinations are assumed to follow a fixed protocol, and therefore be independent of the dental development or of the level of caries activity.

Due to interval censoring, a measurement model must be specified. The posterior density of a_{ij}, given the intensity function $\lambda_{ij}^{(e)}$, the birth time of dentition η_i, the index $k(i, j)$ of the examination (i.e. the observation) in which the eruption was recorded, and the examination protocol $\{u_{ik}\}$, is proportional to

$$\begin{aligned} p\{a_{ij}|f_j, \eta_i, k(i, j), (u_{ik})\} &\propto p\{a_{ij}, k(i, j)|f_j, \eta_i, (u_{ik})\} \\ &= p\{a_{ij}|f_j, \eta_i, (u_{ik})\}p\{k(i, j)|a_{ij}, (u_{ik}), f_j, \eta_i\} \\ &= p\{a_{ij}|f_j, \eta_i\} \cdot 1\{u_{i,k(i,j)-1} < a_{ij} \leqslant u_{i,k(i,j)}\}. \end{aligned}$$

Similarly, the posterior density of b_{ij}, given the intensity function $\lambda_{ij}^{(c)}$, the eruption time a_{ij}, the index $l(i, j)$ of the examination (i.e. the observation) in which the failure was recorded, and the examination protocol (u_{ik}), is proportional to

$$\begin{aligned} p\{b_{ij}|h_j, Z_i, a_{ij}, l(i, j), (u_{ik})\} &\propto p\{b_{ij}, l(i, j)|h_j, Z_i, a_{ij}, (u_{ik})\} \\ &= p\{b_{ij}|h_j, Z_i, a_{ij}\} \cdot 1\{u_{i,l(i,j)-1} < b_{ij} \leqslant u_{i,l(i,j)}\}. \end{aligned}$$

3.2. *Prior distributions*

As an approximation facilitating the numerical integration of the posterior distribution, we assume that the nonparametric random intensity functions h_j and f_j are piecewise constant (see Arjas & Gasbarra, 1994). Using the generic notation g for such functions supported by the interval $[0, T_{\max}]$, we can write

$$g(t) := \sum_{n=0}^{n(T)} a_n 1_{(T_n, T_{n+1}] \cap [0, T_{\max}]}(t),$$

where $n(T)$ is the number of jump points on $(0, T_{\max})$. The prior distribution of the jump points $0 =: T_0 < T_1 < \cdots < T_{n(T)+1} := T_{\max}$ is assumed to be a Poisson process on $(0, T_{\max})$ with (hyper) parameter μ.

3.2.1. *Baseline intensity functions of failures*

The tooth lifetimes are known with an approximate precision of one year: the only situation in which more precise information is available is when both the eruption and the failure of a tooth take place during the same examination interval. Corresponding to this lack of precision in the data, we impose an additional smoothness condition on (the prior of) the functions h_j: The distance between jump points T_n and T_{n+1} must be greater than 0.5 years for all $n \geqslant 0$.

The prior distribution of the levels $a_0, a_1, \ldots, a_{n(T)}$ is specified inductively as follows: we suppose that the initial level a_0 is gamma distributed with parameters α_0 (shape) and β_0 (scale), and that, given $a_0, a_1, \ldots, a_{n-1}$, the next level a_n is gamma distributed with parameters α and α/a_{n-1}. Note that the function h_j is "trend free" with respect to the prior, in the sense that the prior means satisfy $\mathbb{E}[a_n|a_0, a_1, \ldots, a_{n-1}] = a_{n-1}$ for all $n > 0$.

The hyperparameters α and μ control the *a priori* fluctuation of the functions h_j, and α_0 and β_0 the expectation and the variance of the initial level $h_j(0+)$. The behaviour of the functions $h_j = h_{(\kappa,\nu)}$ can be expected to depend quite strongly on index ν: molar teeth are known to carry a much higher risk for caries that the other teeth, so we choose the prior expectation of $h_{(\kappa,\nu)}(0+)$ to be higher for molar teeth than for the other teeth. Here we let $\alpha_0/\beta_0 = 0.3$ for $\nu = 6, 7$, and $\alpha_0/\beta_0 = 0.1$ for $\nu = 1, 2, \ldots, 5$.

3.2.2. Baseline intensity functions of eruptions

The eruption times in the data are concentrated on a relatively short age interval, and until the beginning of that interval, the intensity remains low, after which it rises steeply. As a consequence, unless sufficient attention is paid to parameter identifiability, the dentition times η_i will be identified from the data only very poorly: when considering functions f_j, replacing simultaneously T_n by $T_n + \Delta$ $(n \geqslant 1)$ and η_i by $\eta_i - \Delta$ $(\forall i)$ for any (fixed) Δ, the conditional probability distribution (5) of a_{ij} would remain unchanged. This concern is reflected in the specification of the prior distribution for the functions f: The lowest first incisors and molars ($j' = (\kappa, \nu)$ for $\kappa = 3, 4$ and $\nu = 1, 6$) are the first erupting permanent teeth. For improved calibration of the estimates, we tie the corresponding eruption times $a_{ij'}$ to the corresponding η_i by setting $f_{j'}(0+) = a_0 = 2$. For the other teeth we suppose that the initial level $f_j(0+) = a_0 = 0$, so that these functions remain zero until a first positive value is attained at a random time T_1 (depending on the function in question).

The functions f_j are then assumed (*a priori*) to be increasing: let $d_0 \equiv a_0$, $d_n := a_n - a_{n-1}$ $(n \geqslant 1)$, and suppose that the increments $d_i \sim \text{gamma}(\cdot; \alpha, \beta)$ for all $i = 1, 2, \ldots$. In the present study, the hyperparameters of the tooth specific functions f and h were chosen as in Table 2. The increments d_i are then *a priori* gamma(1, 1) distributed and thus of mean size 1.

3.2.3. Other parameters

The individual frailties Z_i in the intensity model (1) for failure are *a priori* assumed to be conditionally independent and identically distributed, given a hyperparameter ϕ:

$$p\{Z_i|\phi\} = \text{gamma}(\phi, \phi), \quad i = 1, 2, \ldots, 240.$$

The (conditional) *a priori* expectation and variance of Z_i are therefore $\mathbb{E}[Z_i|\phi] = 1$ and $\text{var}(Z_i|\phi) = \phi^{-1}$. An additional level of hierarchy is introduced to control the variance: it is assumed that parameter ϕ is a random variable with prior distribution gamma$(\cdot; 2, 2)$. In this way it becomes possible to learn from the data about the intersubject heterogeneity in the population.

We choose the prior densities of the parameters η_i to be normal $N(\xi, \tau^{-2})$. We assume that, *a priori*, the parameters $\xi \sim N(5, 1)$ and $\tau^2 \sim \gamma(2, 2)$. The assumption $\mathbb{E}[\xi] = 5$ reflects our concept that the first permanent teeth erupt around the age of 5 years, but considering ξ and τ^2 as random variables, we can learn about the distribution of the birth and dentition times η_i in the population.

3.3. The joint distribution of the parameters

A graphical representation (a directed acyclic graph) of this Bayesian hierarchical model is shown in Fig. 2. Following the convention of Spiegelhalter *et al*. (1996), a box represents variables with known values (observations or hyperparameters) and an oval unknown parameters or random variables. A solid arrow represents stochastic dependence and a dashed arrow deterministic dependence. The joint posterior distribution of the unobservables (including parameters) is proportional to the expression

Table 2. *The hyperparameters of the simple functions* ($\kappa = 1, 2, 3, 4$).

Function	Tooth ν	α	β	μ	Function	Tooth ν	α_0	β_0	α	μ
$f_{(\kappa,\nu)}$	$1, 2, \ldots, 7$	1	1	0.5	$h_{(\kappa,\nu)}$	$1, \ldots, 5$	0.1	1	1	0.1
						6,7	0.3	1	1	0.1

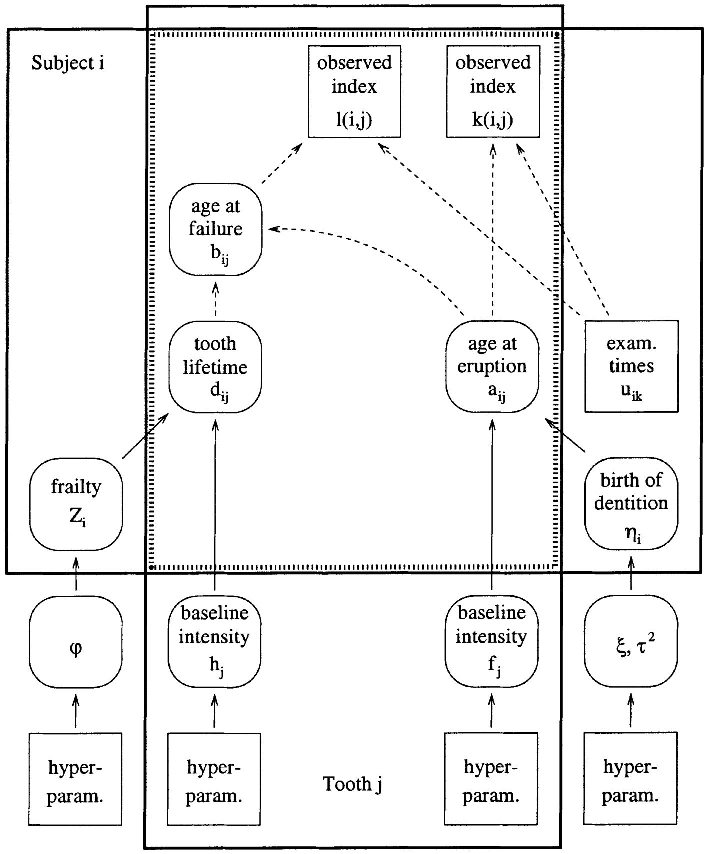

Fig. 2. A hierarchial model of the observed data.

$$
\begin{aligned}
&p\{(a_{ij}), (f_j), (\eta_i), (b_{ij}), (h_j), (Z_i), \phi, \xi, \tau^2 | (k(i, j)), (l(i, j)), (u_{ik})\}\\
&\quad \propto \prod_{i,j} [p, \{a_{ij}|f_j, \eta_i, k(i, j), (u_{ik})\} p\{b_{ij}|h_j, Z_i, a_{ij}, l(i, j), (u_{ik})\}]\\
&\qquad \times \prod_i [p\{Z_i|\phi\} p\{\eta_i|\xi, \tau^2\}] \times \prod_j [p\{f_j\} p\{h_j\}] \times p\{\phi\} p\{\xi\} p\{\tau^2\}.
\end{aligned}
$$

4. Results

We calculated the posterior expectations numerically by using versions of the Metropolis–Hastings–Green algorithm (see Green 1995) and the adaptive Metropolis (AM) algorithm (see Haario *et al.*, 1998). The latter was used in group updating of $(\eta_i, (a_{ij})_{j \in J})$ (for each i), where $J := \{(3, 1), (3, 6), (4, 1), (4, 6)\}$ refers to the first erupting permanent teeth. Single component updating, together with the imposed restriction $\eta_i \leqslant a_{ij}$ for all i, j, would have resulted in a slow convergence of the posterior of parameters η_i and ξ, but application of the AM algorithm improved the situation markedly.

We ran 6000 iterations of burn-in, and, in addition, 18 000 iterations with a 1:3 thinning to obtain a sample from the posterior distribution. Such a run took about 42 hours on a 400 MHz Pentium II PC. It seems that fewer iterations would have been sufficient for reasonably good convergence, but the speed of convergence was quite sensitive to the choice of the controlling hyperparameters.

We calculated the posterior expectations of the baseline intensity functions f and h, together with their symmetric 90% pointwise credible intervals. Each curve was cut off at a point in time after which the credible intervals became too wide (because of the small risk set size) to have any practical value. The estimates of eruption baseline intensity functions f of upper jaw are presented in Fig. 3, and those of failure baseline intensity functions h in Fig. 4.

The estimates of functions f_j reflect the nature of the eruption process: Supposing that the birth of dentition time η_i of subject i is given, the time interval during which tooth (i, j) is likely to erupt is rather short. Corresponding to this, the pointwise posterior mean estimate of f_j remains at zero (except for the lower first incisors and molars for which we had set $f_j(0+) = 2$), and then starts to rise quite steeply. The variances of the eruption times during the first phase (first and second incisors I1 and I2, and first molars M1) are smaller than the variances during the second phase (canines C, first and second premolars B1 and B2, and second molars M2). (See Table 1 for terminology and notations.) This was already noted by Virtanen *et al.* (1994), and here it is reflected by the fact that the estimates of the functions f_j of the first phase teeth rise even more steeply than those of the second phase teeth.

We also drew logarithmic predictive hazard rates (Fig. 5) thus eliminating the need to use "credible bands". Recall (e.g. Arjas & Gasbarra, 1996) that if $F_{i^*j}(t)$ is the predictive cumulative distribution function (CDF) of the lifetime, from eruption to failure, to tooth j of a generic individual i^* (not in the data), and $f_{i^*j}(t)$ is the corresponding density function, then the predictive hazard rate for failure of tooth j is given by $r_{i^*j}(t) := f_{i^*j}(t)/(1 - F_{i^*j}(t))$.

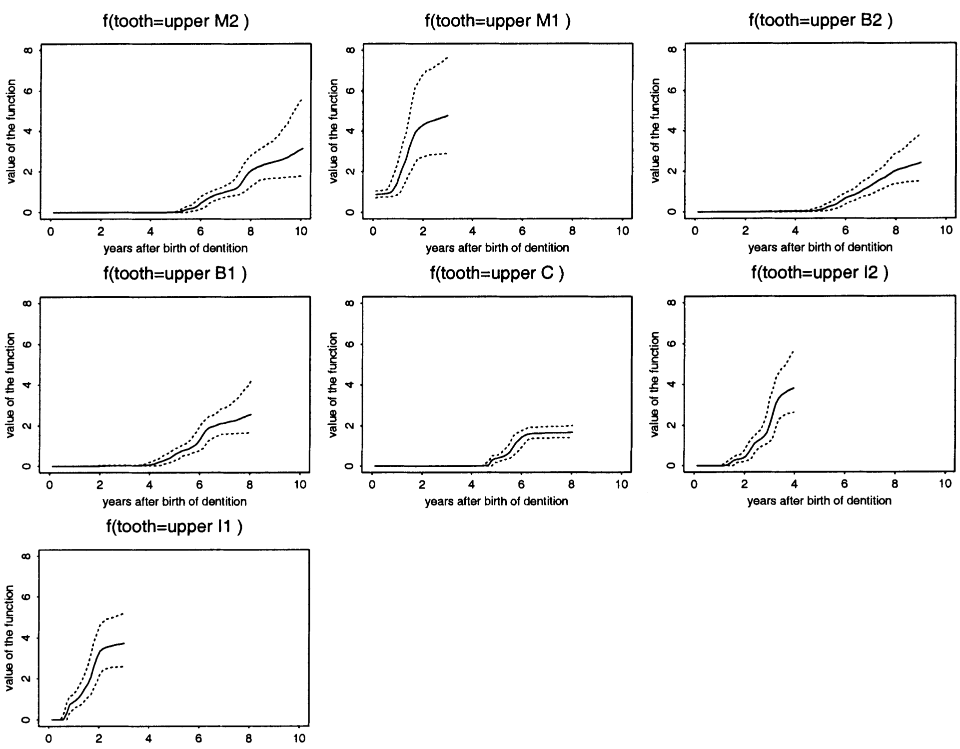

Fig. 3. Baseline eruption intensity functions f_j of the upper jaw.

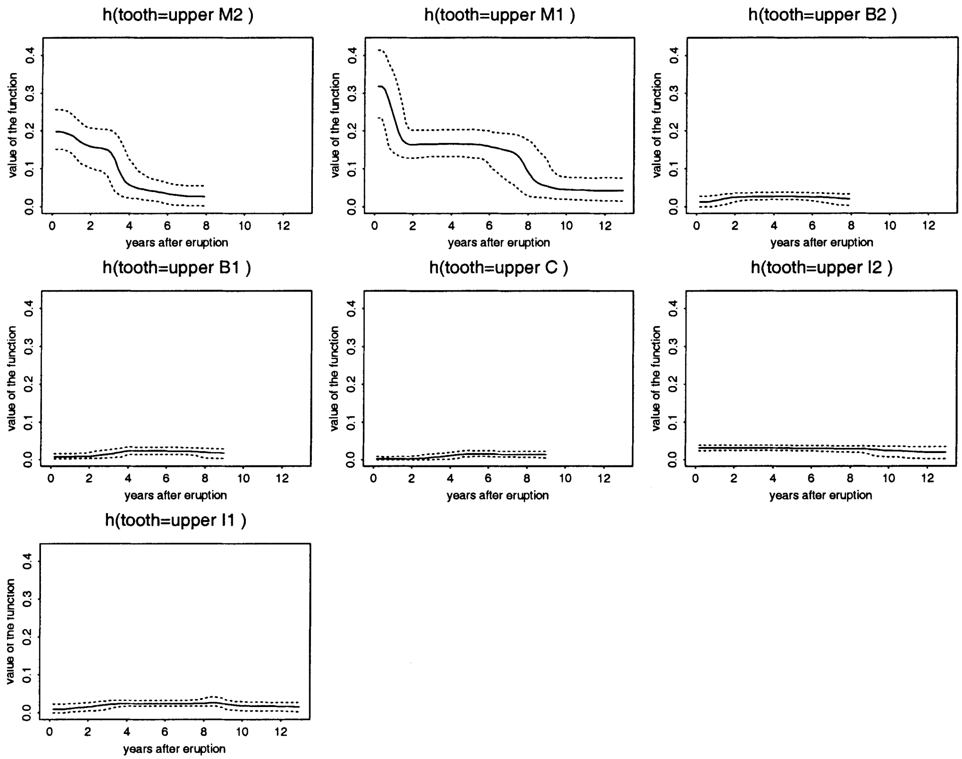

Fig. 4. Baseline failure intensity functions h_j of the upper jaw.

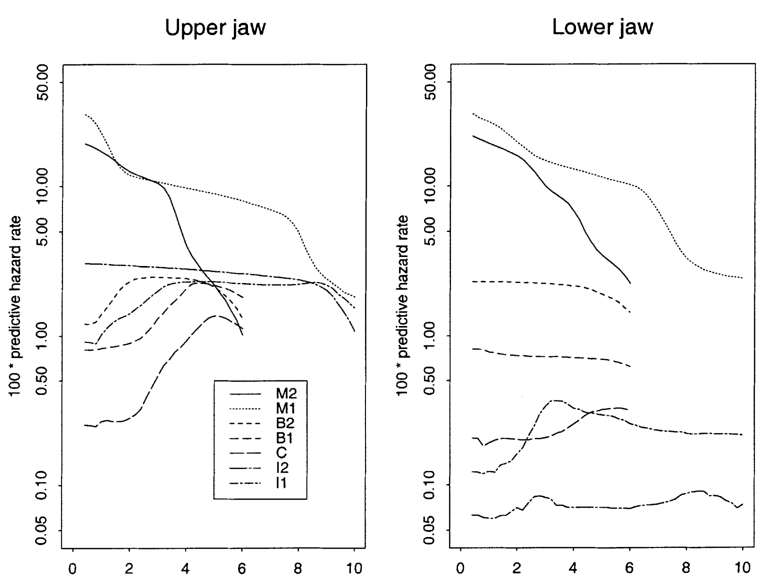

Fig. 5. Predictive failure hazard rates (logarithmic scale).

Intuitively, $r_{i^*j}(t)\Delta t$ (for small Δt) will represent the probability, given the data, and given that tooth (i^*, j) has not failed within t years after eruption, that it will do so during the time interval $(t, t+\Delta t]$. In practice, the determination of the predictive hazard rates requires numerical integration with respect to the joint posterior of the model.

The characteristic differences between the cariotic process in different teeth become even clearer from Fig. 5 than from Fig. 4, where the failure hazard rates of incisors and canines, particularly in the lower jaw, were too low to show any time trend. The predictive failure hazard rates of molars have their highest values right after the time of eruption, after which they decrease monotonically with the dental age of the tooth.

Predictive distributions can also be used as an instrument for model checking. Although we did not make a systematic evaluation of the predictive performance of our model, the following experiment was tried: considering again a hypothetical subject (index by i^*) drawn from the same population as the 240 subjects in the data, we produced the predictive distribution of the eruption and failure times. For comparison, we followed Virtanen *et al.* (1996) and produced empirical CDF's by approximating the eruption and failure times by the midpoints $(u_{i,k(i,j)-1} + u_{i,k(i,j)})/2$ and $(u_{i,l(i,j)-1} + u_{i,l(i,j)})/2$, respectively, of the corresponding examination intervals.

The CDFs of teeth (1,7) are shown in Fig. 6. The predictive CDFs match the empirical CDFs quite well, although the predictive CDF of eruption time seems to start rising later than the empirical CDF and thus the corresponding predictive CDF appears to produce predictions for the later eruption times which are a few months too large. The empirical CDFs were, however, constructed by the rough mid-point approximation, so it is not clear whether the differences are due to an excessive stiffness of our prior or not.

In the present study the estimation of the birth of dentition times η_i was hampered by the fact that some of the first eruptions a_{ij} had taken place already before the first examination time u_{i1}. The posterior expectations $\hat{\eta}_i := \mathbb{E}[\eta_i|\text{data}]$ resemble the prior assumptions: Their mean is $\hat{\eta} := \sum_i \hat{\eta}_i/240 = 5.5 = \mathbb{E}[\zeta|\text{data}]$, and the variance $\sum_i(\hat{\eta}_i - \hat{\eta})^2/240 = 0.6$ is in accordance with the precision parameter $\mathbb{E}[\tau^2|\text{data}] = 1.6$.

We assumed *a priori* that frailties should be 1 on average, which is in accordance with the

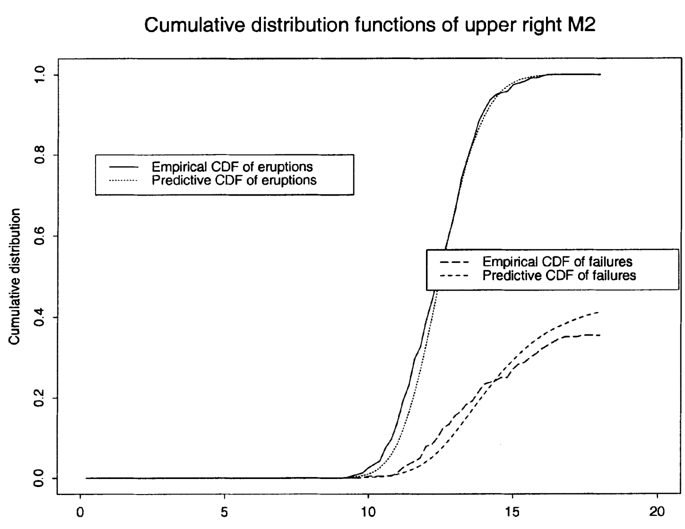

Fig. 6. Empirical and estimated CDFs of eruptions and failures of tooth (1, 7).

posterior expectations $\hat{Z}_i := \mathbb{E}[Z_i|\text{data}]$ of the frailties ($\sum_i \hat{Z}_i/240 = 0.9$). The median of $(\hat{Z}_i)$ is 0.7, and 65% of $\hat{Z}_i$s are below 1. The variance $\sum_i(\hat{Z}_i - \hat{Z})^2/240 = 0.62$ (where $\hat{Z} := \sum_i \hat{Z}_i/240$) is a little smaller than indicated by the posterior expectation of ϕ (i.e. $\mathbb{E}[\phi^{-1}|\text{data}] = 0.88$). The posterior expectation of ϕ is 1.14.

We used the CODA software package (Best *et al.*, 1995) to assess the convergence of parameters ξ, τ^2, and ϕ. The autocorrelations were close to zero with lags greater than 10. The Gelman and Rubin shrink factors (calculated with three other parallel chains) for these parameters were 1.05, 1.01 and 1.00 for the 97.5% quantile which is evidence of good mixing. The dependence factors assuming values of the Raftery and Lewis diagnostic indicated some serial correlation in the chains (values 3.22, 4.15, and 2.53, respectively), but the 2.5% quantile was estimated with a reasonable accuracy (± 0.01 with probability 0.95). The Geweke diagnostic showed no evidence against the convergence of the posterior estimates (Z-score got values 1.71, 1.13 and 0.84, respectively). The Heidelberger and Welch method indicated that the chains were stationary (by accepting the null-hypothesis of Cramer–von Mises test), and the estimates met the accuracy criterion $\epsilon = 0.01$ (of the halfwidth test).

5. Discussion

The characteristic features of the preceding analysis of dental caries attack times are: (i) handling the problem of interval censored eruption and attack times by augmenting the unobserved exact times to the data; (ii) controlling the within subject caries activity/susceptibility by means of corresponding subject-specific frailty parameters; (iii) describing the cariotic process, from eruption to failure, by means of nonparametrically defined tooth-specific baseline hazard rate functions; and finally (iv) applying Bayesian inferential methods and MCMC sampling in the practical execution of the estimation.

One possible improvement of the model could be in (ii): the assumption of time-independent frailty may be unrealistic because the environmental conditions, e.g. quality of oral hygiene, or level of sugar consumption, could well change over time in a way that can be different from one individual to another. We did not try to elaborate on this point here, however.

Our model has random elements on several levels of hierarchy, which has the effect that some of its parameters may not be identifiable from the data very well: Not even the survival data points were observed exactly. For this reason, we chose the levels of h_j to be at least 0.5 years long. The choice of the hyperparameters of the functions f_j is more free, but for very small values of α the gamma prior prefers small increments which can lead to an excessive number of jump points in the approximating step functions. On the other hand, if $\alpha > 1$, adding new levels (and jump points) can be difficult because the prior density of small jumps is low, which can result in a "staircase"-like behaviour in the posterior expectations of functions f_j.

The model could have been simplified by imputing some "sensible looking" values, taken directly from the earliest eruption times, for the birth times of dentition η_i. However, this is *ad hoc*, and particularly so if some early eruption times are not recorded in the data.

A final point of discussion is to what extent more traditional inferential methods could have been applied here, and whether they could be expected to lead comparable results. An application of the EM algorithm in handling the interval censored data, adding a penalty term to the M-step to control the smoothness of the non-parametrically defined baseline hazard rates, would be an obvious candidate. It is doubtful, however, that the data analysis or the computations would be any simpler. Classical non-parametric methods for analysing survival data, such as the Kaplan–Meier and Nelson–Aalen estimators, would be troubled by the interval censoring, and also difficult to extend so that they could cope with the complicated dependence structure present in the multivariate survival data. A parametric model would be able to handle interval censored data, but the estimates might not be realistic: it is difficult to kind a parametric family of functions which would allow for the kind of versatility in the behaviour of hazard rate estimates as is displayed in Fig. 4.

Acknowledgements

We would like to thank Professor Markku Larmas for helpful comments.

References

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993) *Statistical models based on counting processes*. Springer Verlag New York. New York.

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

Arjas, E. & Gasbarra, D. (1996) Bayesian inference of survival probabilities, under stochastic ordering constraints. *J. Amer. Statist. Assoc.* **91**, 1101–1109.

Best, N. G., Cowles, M. K. & Vines, S. K. (1995) *CODA manual version 0.30*. MRC Biostatistics Unit, Cambridge, UK.

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

Clayton, D. (1991) A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* **47**, 467–485.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996) *Markov chain Monte Carlo in practice*. Chapman & Hall, London.

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

Haario, H., Saksman, E. & Tamminen, J. (1998) An adaptive Metropolis algorithm. Reports of the Department of Mathematics, University of Helsinki, Preprint 176.

Spiegelhalter, D. J., Thomas, A., Best, N. G. & Gilks, W. R. (1996). *BUGS: Bayesian inference using Gibbs sampling, version 0.5, (version ii)*. MRC Biostatistics Unit, Cambridge.

Virtanen, J. I., Bloigu, R. S. & Larmas, M. A. (1994) Timing of eruption of permanent teeth: standard Finnish patient documents. *Community Dent. Oral Epidemiol*. **22**, 286–288.

Virtanen, J. I., Bloigu, R. S. & Larmas, M. A. (1996) Effect of early or late eruption of permanent teeth on caries susceptibility. *J. Dent*. **24**, 245–250.

Received August 1998, in final form October 1999

Tommi Härkänen, Rolf Nevanlinna Institute, P.O.B. 4, FIN-00014 University of Helsinki, Finland.