A marginal cure-rate proportional hazards model for spatial survival data

Patrick Schnell[1], Dipankar Bandyopadhyay[1], Brian J. Reich[2]*, and Martha Nunn[3]
[1]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455
[2]Department of Statistics, North Carolina State University, Raleigh, NC 27695
[3]Center for Oral Health Research, Creighton University, Omaha, NE 68178

Abstract

Dental studies often produce spatially-referenced multivariate time-to-event data, such as the time until tooth loss due to periodontal disease. These data are used to identify risk factors associated with tooth loss, and to predict outcomes for an individual patient. The rate of spatial referencing can vary with various tooth locations. In addition, these event time data are heavily censored, mostly due to the fact that a certain proportion of teeth in the population are not expected to experience failure, and can be considered 'cured'. In this paper, we assume a proportional hazards (PH) model with a surviving fraction to model this clustered correlated data, and account for dependence between nearby teeth using spatial frailties which are modeled as linear combinations of positive stable random effects. This model permits predictions (conditioned on spatial frailties) that account for the survival status of nearby teeth, and simultaneously preserves the PH relationship marginally over the random effects for the susceptible teeth, allowing for interpretable estimates of the effects of risk factors on tooth loss. We explore the potential of this model via. simulation studies and application to a real dataset obtained from a private periodontal practice, and illustrate its advantages over other competing models to identify important risk factors for tooth loss and predict the remaining lifespan of a patient's teeth.

Key words: Bayesian hierarchical modeling; Cure rate; Dental data; Extreme value analysis; Frailty; Positive stable; Tooth loss.

*E-mail: brian_reich@ncsu.edu

A marginal cure-rate proportional hazards model for spatial survival data

1 Introduction

Periodontal disease (PD), which contributes to eventual tooth loss, remains a major global oral health burden. Tooth loss can inflict pain and disability, and severely compromise quality of life. The CDC estimates that one fourth of US adults aged 65 or above are toothless (CDC, National Center for Chronic Disease Prevention and Health Promotion, 2011). With Americans visiting dentists 500 million times every year and the ever-rising cost of medical and dental insurance premiums in the US, development of any future dental treatment plan should seek to utilize accurate prognosis. Therefore, developing efficient statistical models to identify risk factors for tooth loss and to predict the remaining lifespan of a tooth are needed.

Consider the motivating McGuire and Nunn (1996) dataset (henceforth, MN) recording tooth loss for 99 subjects, where subjects are followed for several years. They modelled tooth loss as a function of various covariates, such as baseline periodontal probing depth, defined as the distance (in mm) from the gingival margin to the base of the sulcus/pocket as measured by a periodontal probe. Analyzing tooth loss data presents many challenges. These data exhibit spatial dependence (Banerjee et al., 2014), because tooth survival might be associated with the disease status (and survival) of a neighboring tooth. In addition, due to diverse factors that lead to tooth loss, these data might exhibit nonstationarity, with varying patterns of spatial dependence in different regions of the mouth, as well as a complex dependence between the observations for each subject. Furthermore, tooth loss data are typically highly censored as many teeth are lost before observation begins or remain intact at the final observation. These censored survival data are also dependent within subject due to shared dietary, hygienic, and other factors. From a biological perspective, many of these teeth remain sound throughout the full lifetime of a subject and can be considered nonsusceptible to dental disease progression, or 'cured' (Sy and Taylor, 2000). Thus, the study

population constitute subjects with a mixture of both susceptible and nonsusceptible teeth.

Tooth loss data have been modeled using standard Kaplan-Meier survival analysis techniques (Chuang et al., 2001; Härkänen et al., 2002), regressions accommodating clustered interval-censored structure of the data (Wong et al., 2005), via non-spatial proportional hazards (PH) frailty models (Manda et al., 2005) and marginal models (Spiekerman and Lin, 1998; Chuang et al., 2002). Some recent advances in this direction also include the extensions of the classification and regression trees (CART) methodology to correlated survival outcomes (Fan et al., 2006, 2009).

A variety of methods have been proposed for spatially-referenced survival data. From an interpretation perspective, these methods can be broadly classified as either conditional or marginal. Conditional models introduce random effects (frailties) in either the PH (Li and Ryan, 2002; Banerjee et al., 2003; Hennerfeind et al., 2006) or proportional odds (Banerjee and Dey, 2005) framework, and then use Gaussian spatial models for the frailties to account for spatial dependence. These methods are referred to as conditional because the regression coefficients are interpreted conditional on the spatial frailties. In some cases, for example identifying population-level risk factors for tooth loss, the conditional interpretation can be awkward because the frailties are essentially nuisance parameters used to capture spatial dependence, and the effects of the regression parameters marginally over the frailties do not have a closed form. Li and Lin (2006) propose a marginal model to overcome this limitation. In this model, the data are transformed to Gaussian, and modeled directly (that is, without frailties) using a Gaussian spatial model, thus preserving the marginal interpretation of the regression coefficients. However, this model does not preserve the well-accepted PH interpretation of the regression coefficients, and does not permit estimation of spatial frailties which can be used to identify at-risk regions. Further extensions to accommodate a cured proportion along with spatial dependence in survival data appear in Banerjee and Carlin (2004) and Cooner et al. (2006).

In this paper, we model tooth loss using a mixture cure (Sy and Taylor, 2000) spatial survival model that accounts for both susceptible and nonsusceptible teeth and the dependence between the event times of nearby teeth. Here, we assume a fraction π of the entire teeth population to be

cured, and the remaining $1-\pi$ remaining susceptible (non-cured). Hence, the survival function for the whole population $S(t)$ can be written as $S(t) = \pi + (1-\pi)S_1(t)$, where $S_1(t)$ is the survival function for the non-cured teeth. Note that, under $\pi > 0$, $S(t)$ is no longer a survival function ($S(\infty) = \pi$), however, $S_1(t)$ is. We assume a PH relationship between the covariates and the survival outcome $S_1(t)$ for the non-cured teeth, and use frailties to account for spatial dependence between nearby teeth. However, rather than a Gaussian model for the random effects, we model the frailties as linear combinations of positive stable (PS) random variables (Hougaard, 1986). For non-spatial data, Liu et al. (2011) show that PS random effects preserve the PH relationship between predictors and survival marginally over the random effects. We extend this idea to the spatial setting borrowing concepts from spatial extreme value analysis (Reich and Shaby, 2012; Shaby and Reich, 2013). Note that a mixture of PH functions is no longer proportional, and $S(t)$ does not have PH structure. Throughout the paper, our focus and model development for understanding covariate-response relationship remains confined only to the non-cured teeth, and their survival function $S_1(t)$. Our spatial model not only preserves the marginal PH relationship, but also permits straight-forward predictions of the remaining lifespan of a tooth for the non-cured teeth. In addition, the spatial dependence induced by this model has nice properties, including non-stationarity and can be motivated as the limiting process under a competing risk hypothesis.

2 Description of the McGuire and Nunn data

The MN dataset consists of $n = 99$ subjects with at least 5 years of maintenance selected from a private periodontal practice of Dr. Michael K. McGuire in the Houston area. Subjects were followed for 16 years. For each tooth, the response is time after the initial visit until tooth loss due to PD. We use $p = 9$ covariates. The subject-level covariates are age at baseline, indicator of poor (fair is the baseline) hygiene, indicator of good hygiene and smoking status. The hygiene variable is an assessment of the quality of the subject's home oral hygiene, based on looking at a patient's plaque score, amount of calculus accumulation, bleeding index, and gingival inflammation. A patient with

a low plaque score, little bleeding on probing, and fairly healthy gingiva would be assessed with 'good' hygiene. In contrast, a patient with a high plaque score, a lot of bleeding on probing and puffy, inflamed gingiva would be classified with 'poor' hygiene, with the baseline category 'fair' standing somewhere in between the two extremes. The binary smoking status indicated ever a smoker, or never. The baseline tooth-level covariates are crown-to-tooth ratio indicator, probing depth, mobility, an indictor of an missing adjacent tooth on the same jaw, and an indictor of a missing directly opposing tooth on the opposite jaw. The crown-to-root ratio is defined as the ratio of the length of the part of a tooth that appears above the alveolar bone versus what lies below it (Penny and Kraal, 1979). Our dataset records the crown-to-root ratio (binary) indicator, which is 1 if it was deemed unsatisfactory by the clinician due to PD, and 0 otherwise. Mobility is defined as the amount of bone loss around the tooth, and is categorized into 4 classes starting from complete tooth stability to tooth being terminally mobile, i.e., bone loss of 1mm in any direction. In our analysis, we treat it as a continuous variable.

We analyze data for $m = 28$ teeth for each subject (third molars are excluded). The 14 teeth on the upper jaw are numbered from 2 (left) to 15 (right), and the 14 teeth on the lower jaw are numbered 18 (right) to 31 (left). Teeth are classified as molars (teeth 2, 3, 14, 15, 18, 19, 30, and 31), pre-molars (4, 5, 12, 13, 20, 21, 28, and 29), canines (6, 11, 22, and 27), and incisors (7-10 and 23-26). The spatial coordinates are defined as $\mathbf{s} = (s_1, s_2)$, where s_1 is the tooth's right/left position (so that teeth 2 and 31 have $s_1 = 1$ and teeth 15 and 18 have $s_1 = 14$) and s_2 as zero for teeth on the upper jaw and 1 for teeth on the lower jaw.

We denote the event time for the tooth at location $\mathbf{s}$ for subject i as $Y_i(\mathbf{s})$. Because the dataset is mostly subjects with well-maintained dental hygiene, the censoring rate is high. Figure 1 presents the unadjusted Kaplan-Meier survival curves for the whole data, and stratified for the four broad tooth-types: molar, premolar, canine and incisor. The survival curves level off and plateaus towards the tails for all cases, thereby suggesting the possibility of the presence of a cured proportion of teeth in the dataset. Averaging over subjects, 12.8% of teeth are missing before the first visit (missing), 82.8% of the teeth remain in place at the final visit (right-censored), and 4.4% (122 teeth) of teeth

were lost during the observation window (uncensored). The censoring indictor is $\delta_i(\mathbf{s}) = 1$ for teeth remaining at the final visit, and $\delta_i = 0$ for teeth that fall out during the monitoring period. For censored observations, we denote the final observation time as $\tilde{Y}_i(\mathbf{s})$ and thus $Y_i(\mathbf{s}) \in (\tilde{Y}_i(\mathbf{s}), \infty)$, and we assume a proportion π of these censored observations are cured.

3 Spatial survival model

We model $Y_i(\mathbf{s})$ for locations $\mathbf{s} \in \{\mathbf{s}_1, \ldots, \mathbf{s}_m\}$ and subjects $i = 1, \ldots, n$ in terms of subject- and tooth-level covariates $\mathbf{X}_i(\mathbf{s}) = (X_{i1}(\mathbf{s}), ..., X_{ip}(\mathbf{s}))$. We assume the PH framework for the susceptible/non-cured teeth in a mixture-cure model, and that event times are conditionally independent given spatial random effects $\theta_i(\mathbf{s}) > 0$. We specify the hazard function for the non-cured teeth at location $\mathbf{s}$ and time $t > 0$ as

$$\lambda_i(\mathbf{s}, t) = \theta_i(\mathbf{s}) \exp[\mathbf{X}_i(\mathbf{s})\boldsymbol{\beta}^C]\lambda_0(\mathbf{s}, t),$$

where $\lambda_0(\mathbf{s}, t) > 0$ is the baseline hazard function at location $\mathbf{s}$ (discussed further in Section 3.2). The regression coefficients, $\boldsymbol{\beta}^C = (\beta_1^C, \ldots, \beta_p^C)^T$, are interpreted as the log hazard ratio conditionally (hence the superscript 'C') on θ. The corresponding overall survival function $S(t)$ given the spatial frailty is

$$\mathrm{P}[Y_i(\mathbf{s}) > t|\theta_i(\mathbf{s})] = \pi + (1 - \pi) \exp\left\{-\theta_i(\mathbf{s}) \exp[\mathbf{X}_i(\mathbf{s})\boldsymbol{\beta}^C]H_0(\mathbf{s}, t)\right\},$$

where $H_0(\mathbf{s}, t) = \int_0^t \lambda_0(\mathbf{s}, u)du$ is the cumulative baseline hazard and π is the proportion of the teeth not diseased, i.e. the cured proportion. For estimating the remaining lifetime of a tooth, this is the survival function of interest. Marginalizing over θ gives the population survival function

$$\mathrm{P}[Y_i(\mathbf{s}) > t] = \pi + (1 - \pi) \int_0^\infty \exp\left\{-\theta \exp[\mathbf{X}_i(\mathbf{s})\boldsymbol{\beta}^C]H_0(\mathbf{s}, t)\right\} f(\theta)d\theta. \quad (1)$$

For studies comparing populations, this is the survival function of interest since it is the survival function of a typical subject with covariates $\mathbf{X}_i(\mathbf{s})$. However, (1) generally does not have a closed form or preserve the PH interpretation of the regression coefficients. In particular, the PH interpretation of the regression coefficients does not hold if the spatial random effects are Gaussian.

To preserve the PH interpretation while accounting for spatial dependence, the random effects are modeled using PS random variables. We assume the random effect is a linear combination of L PS random variables,

$$\theta_i(\mathbf{s}) = \sum_{l=1}^{L} w_l(\mathbf{s})^{1/\alpha} A_{il} \text{ with } A_{il} \overset{iid}{\sim} \text{PS}(\alpha), \tag{2}$$

where $w_l(\mathbf{s}) \geq 0$ are basis functions, $A_{il} \geq 0$ are the corresponding coefficients, and $\alpha \in (0, 1]$ controls the shape and scale of the PS distribution. Selection of L is discussed in Section 5. The kernel coefficients A_{il} follow the PS distribution with Laplace transformation $\int_0^\infty \exp(-At) f(A|\alpha) dA = \exp(-t^\alpha)$ and density $f(A|\alpha)$, which we denote $A_{il} \sim \text{PS}(\alpha)$. $f(A|\alpha)$ does not have a closed form (Stephenson, 2009), but can be written as

$$f(A|\alpha) = \int_0^1 \frac{\alpha A^{-1/(1-\alpha)}}{1-\alpha} c(B) \exp\left[-c(B) A^{-\alpha/(1-\alpha)}\right] dB, \tag{3}$$

where $c(B) = \frac{\sin(\alpha\pi B)}{\sin(\pi B)}^{1/(1-\alpha)} \frac{\sin[(1-\alpha)\pi B]}{\sin(\alpha\pi B)}$. If $\alpha = 1$, $f(A|\alpha)$ is the point mass distribution degenerate at $A = 1$, the random effects are irrelevant and there is no attenuation; as α decreases towards zero, $f(A|\alpha)$ becomes right-skewed with spatial random effect variance tending towards infinity and there is substantial attenuation. The basis functions determine the form of spatial dependence, as discussed further in Section 3.1. To fix the scale of the random effects, we restrict $\sum_{l=1}^{L} w_l(\mathbf{s}) = 1$ for all $\mathbf{s}$.

Marginally over the spatial random effects, the survival function and corresponding hazard

function become (Appendix A.1)

$$\begin{aligned} \mathrm{P}[Y_i(\mathbf{s}) > t] &= \pi + (1-\pi)\exp\left\{-\exp[\mathbf{X}_i(\mathbf{s})\boldsymbol{\beta}^M]H_0(\mathbf{s},t)^\alpha\right\} \\ \lambda_i(\mathbf{s},t) &= \exp[\mathbf{X}_i(\mathbf{s})\boldsymbol{\beta}^M][\alpha H_o(\mathbf{s},t)^{\alpha-1}\lambda_0(\mathbf{s},t)], \end{aligned} \tag{4}$$

which for the non-cured teeth correspond to a PH model with marginal (hence the superscript 'M') log hazard ratio coefficients $\boldsymbol{\beta}_M = (\beta_1^M, \ldots, \beta_p^M)^T = (\alpha\beta_1^C, \ldots, \alpha\beta_p^C)^T$ and cumulative baseline hazard $H_0(\mathbf{s},t)^\alpha$. Comparing $\boldsymbol{\beta}_C$ and $\boldsymbol{\beta}_M$, we observe that α controls the attenuation due to spatial dependence. Note that α is both the spatial dependence parameter and the attenuation parameter, thus playing a dual role of both determining the degree of spatial dependence and conditional/marginal attenuation.

3.1 Spatial dependence

In addition to the appealing marginal interpretation of the hazard ratio parameters for the susceptible tooth, this model also provides a nice interpretation of the spatial dependence as quantified by the joint survival function for two locations $\mathbf{s}_j$ and $\mathbf{s}_k$. To separate spatial variation in the marginal survival function from within-subject dependence, we present the joint survival probability assuming non-cure, as well as the same covariates and cumulative baseline hazard for all sites, i.e., $\mathbf{X}_i(\mathbf{s}) = \mathbf{X}_i$ and $H_0(\mathbf{s},t) = H_0(t)$ for all $\mathbf{s}$. Then, Appendix A.2 shows that

$$\mathrm{P}[Y_i(\mathbf{s}_j) > t, Y_i(\mathbf{s}_k) > t] = \exp\left[-\exp(\mathbf{X}_i\boldsymbol{\beta}^M)H_0(t)^\alpha \sum_{l=1}^{L}\left[w_l(\mathbf{s}_j)^{1/\alpha} + w_l(\mathbf{s}_k)^{1/\alpha}\right]^\alpha\right]. \tag{5}$$

Borrowing a concept from extreme value analysis (Coles, 2001), we summarize joint survival using the extremal coefficient (with slight abuse of notation, since the spatial extremal coefficient is usually defined in terms of the joint distribution function rather than the joint survival function) $\vartheta(\mathbf{s}_j, \mathbf{s}_k) \in [1,2]$, defined via the equation

$$\mathrm{P}[Y_i(\mathbf{s}_j) > t, Y_i(\mathbf{s}_k) > t] = \mathrm{P}[Y_i(\mathbf{s}_j) > t]^{\vartheta(\mathbf{s}_j,\mathbf{s}_k)}.$$

The extremal coefficient ranges from $\vartheta(\mathbf{s}_j, \mathbf{s}_k) = 1$ if $Y_i(\mathbf{s}_j)$ and $Y_i(\mathbf{s}_k)$ are completely dependent, to $\vartheta(\mathbf{s}_j, \mathbf{s}_k) = 2$ if $Y_i(\mathbf{s}_j)$ and $Y_i(\mathbf{s}_k)$ are independent. From (5), for the PS frailty model

$$\vartheta(\mathbf{s}_j, \mathbf{s}_k) = \sum_{l=1}^{L} \left[w_l(\mathbf{s}_j)^{1/\alpha} + w_l(\mathbf{s}_k)^{1/\alpha} \right]^{\alpha}.$$

Note that $\vartheta(\mathbf{s}_j, \mathbf{s}_k)$ holds simultaneously for all t. In our current formulation, the degree of spatial dependence remains constant over time, which is not generally the case for other frailty distributions. Certainly, the degree of spatial dependence could vary with time. In comparison to the other alternatives such as the Gaussian model (which implicitly assumes asymptotic independence) and an independence model (no spatial dependence whatsoever), our balanced model seems to be a better default choice. Further comparisons between these assumptions appear in Sections 5 and 6.

There are several potential choices for the basis functions $w_l(\mathbf{s})$. To satisfy the constraints that $w_l(\mathbf{s}) \geq 0$ and $\sum_{l=1}^{L} w_l(\mathbf{s}) = 1$ for all $\mathbf{s}$, we take $w_l(\mathbf{s}) = \frac{\exp[K_l(\mathbf{s})]}{\sum_{j=1}^{L} \exp[K_j(\mathbf{s})]}$ for $K_l(\mathbf{s}) \in \mathcal{R}$. For extreme value analysis, Reich and Shaby (2012) and Shaby and Reich (2013) use the simple model $K_l(\mathbf{s}) = -\rho(\mathbf{s} - \mathbf{v}_l)^2$, where $\mathbf{v}_1, ..., \mathbf{v}_L$ are fixed knots and $\rho > 0$ controls the range of spatial dependence. However, this model may not be rich enough to capture the complex dependence of periodontal data which display dependence not only between nearby teeth, but also between teeth in different quadrants.

To provide a flexible model for the spatial dependence function, we use Gaussian process (GP) priors for the $K_l(\mathbf{s})$ with mean $\mathrm{E}[K_l(\mathbf{s})] = \mu$, variance $\mathrm{V}[K_l(\mathbf{s})] = \sigma^2$, and anisotropic spatial correlation $\mathrm{Cor}[K_l(\mathbf{s}_j), K_l(\mathbf{s}_k)] = \exp(-d_{jk})$, where

$$d_{jk} = \left[(s_{j1} - s_{k1})^2/\rho_1^2 + (s_{j2} - s_{k2})^2/\rho_2^2 \right]^{1/2}. \tag{6}$$

We denote this model as $K_l(\mathbf{s}) \overset{iid}{\sim} \mathrm{GP}(\mu, \sigma, \rho)$. Note that this set of L functions that constitutes the basis function $w_l(\mathbf{s})$ are not identical, but exchangeable, drawn from the same prior distribution which does not depend on l, leading to different posteriors as determined by the data. The model for (2) thus resembles the spatial factor model of Lopes et al. (2008) (without the time dimension).

We must consider identifiability carefully. Since adding a constant to each K_l will not affect the function w_l, we fix $\mu = 0$. In addition, changing the labels of the basis functions, for example exchanging K_1 and K_2, does not affect the likelihood. A common approach in factor analysis is to fix several elements to zero so that the remaining elements are identified. To avoid selecting the elements to fix to zero, we elect not to constrain the K_l. Therefore, the K_l are not individually identified, which must be accounted for when monitoring MCMC convergence. Fortunately, the parameters of interest such as $\theta_i(\mathbf{s})$ and $\vartheta(\mathbf{s}_j, \mathbf{s}_k)$ are identified, and these parameters are used to assess convergence.

3.2 The baseline hazard function

The marginal and conditional survival distributions simplify in the special case of the parametric Weibull baseline hazard $\lambda_0(\mathbf{s}, t) = \kappa^C(\mathbf{s})\eta(\mathbf{s})^{-\kappa^C(\mathbf{s})}t^{\kappa^C(\mathbf{s})-1}$, with spatially-varying shape, $\kappa^C(\mathbf{s}) > 0$, and scale, $\eta(\mathbf{s}) > 0$. In this parametric model, not only is the PH assumption preserved marginally, the Weibull distribution is preserved as well. The conditional and marginal distributions for non-cured teeth are

$$\begin{aligned} Y_i(\mathbf{s})|\theta_i(\mathbf{s}) &\overset{indep}{\sim} \text{Weibull}\left[\kappa^C(\mathbf{s}), \exp\left(-\frac{\log[\theta_i(\mathbf{s})] + \mathbf{X}_i(s)\boldsymbol{\beta}^C}{\kappa^C(\mathbf{s})}\right)\eta(\mathbf{s})\right] \\ Y_i(\mathbf{s}) &\sim \text{Weibull}\left[\kappa^M(\mathbf{s}), \exp\left(-\frac{\mathbf{X}_i(s)\boldsymbol{\beta}^M}{\kappa^M(\mathbf{s})}\right)\eta(\mathbf{s})\right], \end{aligned} \tag{7}$$

where $\kappa^M(\mathbf{s}) = \alpha\kappa^C(\mathbf{s})$ is the marginal shape parameter. The survival times are independent conditionally on $\theta_i(\mathbf{s})$, and spatially dependent marginally with the Weibull marginal distributions at each $\mathbf{s}$. In our analysis in Section 5, we assume that the Weibull shape parameter is constant throughout the mouth ($\kappa^C(\mathbf{s}) \equiv \kappa^C$), and that the Weibull scale parameter varies by tooth type, with $\eta(\mathbf{s}) = \eta_1$ for molars, $\eta(\mathbf{s}) = \eta_2$ for pre-molars, and $\eta(\mathbf{s}) = \eta_3$ for canines, and $\eta(\mathbf{s}) = \eta_4$ for incisors.

The Weibull baseline hazard can also be motivated using asymptotic arguments. Appendix A.3 shows that this spatial process is minimum stable assuming a Weibull baseline hazard with

spatially-constant shape $\kappa^C(\mathbf{s}) \equiv \kappa^C$. That is, define $\tilde{Z}(\mathbf{s}) = a(N) + b(N)\min\{Z_1(\mathbf{s}), ..., Z_N(\mathbf{s})\}$ as the scaled (by some $a(N)$ and $b(N)$ so that $\tilde{Z}(\mathbf{s})$ is non-degenerate) point-wise minimum of the independent and identically distributed spatial processes $Z_l(\mathbf{s}) > 0$. Then for the large class of survival functions in the Weibull domain of attraction for $Z(\mathbf{s})$, the survival function of $\tilde{Z}(\mathbf{s})$ converges to (4) as $N \to \infty$. This motivates the use of the proposed spatial survival model in cases where the events can be thought of as the result of several competing risks, and the event time is the first of the failure times corresponding to each risk.

4 Computing details

Assuming the parametric model (7), the likelihood is simply

$$\prod_{i=1}^{N}\prod_{j=1}^{m}\left(\delta_i(\mathbf{s}_j)\pi + (1-\pi)\left\{1 - F_{ij}^C[L_i(\mathbf{s}_j)]\right\}^{\delta_i(\mathbf{s}_j)} f_{ij}^C[Y_i(\mathbf{s}_j)]^{1-\delta_i(\mathbf{s}_j)}\right), \tag{8}$$

where $F_{ij}^C(y)$ is the Weibull cumulative distribution function with parameters given in (7) for subject i at location $\mathbf{s}_j$ and $f_{ij}^C(y)$ is the corresponding density function. As stated earlier, the PS density defined in (3) does not have a closed form. Hence, for computing, we approximate this as a Gaussian quadrature over 100 equally-spaced points covering the unit interval. Further sensitivity analysis revealed that increasing the number of quadrature points only increased the computing time without any noticeable change in the posterior parameter estimates.

Due to the mixture component in the likelihood (8), none of the parameters can be factored out. Therefore, we use Metropolis-within-Gibbs algorithm (Tierney, 1994) in `R` (R Core Team, 2012) for all parameters. We use log-normal candidates for parameters with positive support, logit-normal candidates for parameters in $[0, 1]$, and Gaussian candidate distributions for parameters with support on the entire real line, with the acceptance rates tuned between 0.3 and 0.6. Details on the posterior densities of the parameters are presented in Appendix A.4. We sample two chains with widely dispersed starting values for 250,000 iterations with a spacing of 5, leaving 50,000 samples from each chain. Of these 50,000, the first 40,000 are discarded as burn-in, leaving $B = 10,000$

samples from each chain and a total of 20,000 samples for computing the posteriors. Convergence is monitored using trace plots, autocorrelation plots, and the Gelman-Rubin potential scale reduction factor $\hat{R}$ (Gelman and Rubin, 1992) combining the two chains, separately for all the model parameters. The $\hat{R}$ statistic for most parameters were between 1.0 and 1.2, with the exception of β for mobility and crown-to-root ratio having $\hat{R}$ of approximately 1.3. The consistent batch means estimates of Monte Carlo standard errors (Flegal et al., 2008) for all parameters ranged between 0.003 to 0.2.

Samples from the posterior predictive distribution for censored observations are immediately available from MCMC output. Predictions for censored $Y_i(\mathbf{s})$ are made at each iteration by sampling from the mixture of the conditional Weibull in (7) restricted to $(\tilde{Y}_i(\mathbf{s}), \infty)$ and the degenerate distribution at $+\infty$ for the cured proportion. Denoting $Y_i(\mathbf{s})^{(b)}$ as the sample for iterations $b = 1, ..., B$ in which the predictions were finite, the posterior predictive mean for diseased teeth is approximated as $\sum_{b=1}^{B} Y_i(\mathbf{s})^{(b)}/B$, and the T-year survival probability for diseased teeth is approximated as $\sum_{b=1}^{B} I(Y_i(\mathbf{s})^{(b)} > u_i + T)/B$. The T-year survival probability for all teeth is approximated as $\pi + (1-\pi)\sum_{b=1}^{B} I(Y_i(\mathbf{s})^{(b)} > u_i + T)/B$.

We conduct model comparison via the log pseudomaximum likelihood ($LPML$; Geisser and Eddy, 1979) statistic. $LPML$ is a summary of leave-one-out cross-validation using the log likelihood as the objective function. That is, $LPML = \sum_{ij} \ell_i(\mathbf{s}_j)$, where $\ell_i(\mathbf{s}_j)$ is the conditional predictive ordinate of the observation at location $\mathbf{s}_j$ for subject i, given all other observations and marginalizing over all model parameters. Despite this interpretation in terms of cross validation, $LPML$ can be computed with a single application of the MCMC algorithm to the complete dataset as shown in Gelfand and Dey (1994). Models with larger $LPML$ are preferred.

5 Analysis of the MN data

We fit cure and non-cure versions of the PS model with $L = 5, 10, 15, 20$ and 28 factors, the model with log Gaussian processes for the spatial frailties, i.e., $\log[\theta_i(\mathbf{s})] \overset{iid}{\sim} \text{GP}(0, \sigma, \rho)$, and the

independence (non-spatial) frailty model with $\log[\theta_i(\mathbf{s})] \overset{iid}{\sim} \text{PS}(\alpha)$. For reasons of identifiability, σ^2 (the variance parameter for the Gaussian process governing $K_l(.)$), ρ_1 and ρ_2 (the scale parameters in (6)) were all fixed to 1. For all models, we select non-informative priors β_j^C, $\log(\eta_j)$, $\log(\kappa^C) \overset{iid}{\sim} \text{N}(0, \text{variance} = 10^2)$, $\text{logit}(\alpha) \sim \text{N}(0,1)$, and $\pi \sim \text{Uniform}(0,1)$. Model comparison using $LPML$ statistics is presented in Table 1. We select the PS cure frailty model with $L = 20 < M$ factors as the best model (largest LPML in Table 1).

In our analysis, we assumed a parametric Weibull baseline hazard function. To evaluate the validity of this assumption, we compute the probability inverse transform statistics PIT_{ij}. Let $F_{ij}(y)$ be the cumulative distribution functions for the marginal Weibull distribution in (7) for subject i at location $\mathbf{s}_j$. The statistic is defined as $PIT_{ij} = (1-\pi)F_{ij}[Y_i(\mathbf{s}_j)]$ for uncensored observations. For censored observations, PIT_{ij} is drawn uniformly from $(1-\pi)F_{ij}[\tilde{Y}_i(\mathbf{s}_j)]$ to 1, where $\tilde{Y}_i(\mathbf{s}_j)$ is the final observation time. Assuming the model fits well, PIT_{ij} should follow the Uniform(0,1) distribution. Figure 2 shows quantile-quantile (QQ) plots of PIT_{ij} versus the Uniform(0,1) distribution. The QQ-plot is computed 10 times using different random uniform samples for censored observations. Based on these diagnostics, we do not suspect any glaring departures from the Weibull assumption, and we therefore proceed with this parametric analysis.

Table 2 summarizes the parameters (both conditional and marginal) in the final model (PS cure model with $L = 20$), and also compares the (conditional) posterior estimates from the Gaussian and the independence model. Note that these estimates can be interpreted for the susceptible (non-cured) teeth. The posterior mean (90% interval) of the spatial dependence parameter α is 0.11 (0.08, 0.14), indicating strong spatial dependence, in comparison to a moderate spatial dependence [0.29 (0.23,0.40)] for the best fitting non-cure model (with $L = 28$). Note, $\alpha = 1$ implies independence. Both the conditional and marginal hazard increases for smokers and subjects with poor hygiene across all models. The spatial predictors crown-to-root ratio, probing depth, and mobility all have strong positive effects on the hazard function. Interestingly, after accounting for the other predictors, there is no significant effect of missing opposite teeth at baseline, however a missing adjacent tooth significantly increases the hazard. Note that factoring in the cured proportion, the

conditional hazards are larger for the PS model (for most parameters), as compared to the Gaussian and independence models. The estimates of the Weibull shape parameter implies increasing failure rate with time, and the survival distribution less skewed than the exponential, across all models. The scale is the smallest for molars, followed by pre-molars, and largest for the canines and incisors also across all models. The estimates of the cure proportion π is 0.15 (0.07,0.25) for the PS model, and increases to 0.26 (0.12,0.40) for the Gaussian, and 0.53 (0.32,0.69) for the independence model.

The posterior mean of the extremal coefficient $\vartheta(\mathbf{s}_j, \mathbf{s}_k)$ in Figure 3 reveals a complex non-stationary dependency for the cure and non-cure models. From the left panel (plot from the best-fitting non-cure model), the cluster of teeth with the strongest dependence (smallest extremal coefficient) are the anterior teeth 6-11 and 22-27. These are the incisors and canines on the upper and lower jaws, respectively. Therefore, all the teeth in the front of the mouth form a cluster with strong interdependence. There is also dependence between neighboring molars and pre-molars, with less dependence across quadrants for these teeth in the back of the mouth. However, the right panel (plot from the best-fitting cure model) presents a much smoother plot, although the degree of the overall spatial dependence is stronger than the non-cure model. The dependence among these anterior teeth is now lower across all quadrants, with regions of strong dependence appearing spuriously. The strong spatial interdependence in the left panel can result from groups of proximal cured (non-susceptible) anterior teeth across the subjects. Once these are not considered in the cure model, the degree of spatial interdependence between the anterior teeth on the overall reduces substantially (right panel).

Figure 4 summarizes the posterior predictive model for a relatively healthy subject with three teeth failures during the study, and one tooth already missing at baseline. The mean residual life (MRL, defined as the expected number of additional years past the last visit until the tooth falls out) estimates for the PS model are a bit higher than the Gaussian model. Both spatial dependence and the covariates clearly affect predictions. The teeth located in the proximity of those that fell out during the monitoring period experience the smallest MRL, and teeth on the opposite side of the jaw also have low MRL. Teeth with large probing depth and non-zero mobility also have

smaller predicted MRLs. Perhaps more relevant in practice than predictive MRLs are the posterior 5- and 10-year survival probabilities. Teeth 12 and 14, the pre-molar and molar in the upper jaw, have the lowest probability of surviving 5 and 10 additional years beyond the final visit. These teeth have neighbors that have fallen out, non-zero mobility, and probing depth of at least 7mm. Their probabilities of surviving an additional five years are around 0.6, and their probabilities of surviving an additional ten years are around 0.4.

6 Simulation study

In this section, we compare the finite sample performance of the regression parameter estimates for the spatial survival models with the PS, Gaussian and independence random effects. We generate data from the cure-rate model for 50 subjects on a 5×2 rectangular grid with grid spacing 1, and cure fraction 0.15 (approximately representing the data). We use $p = 4$ covariates, with the first two $X_{ij}(\mathbf{s}) \equiv X_{ij} \sim N(0,1)$ (representing subject-level), and the next two $X_{ij}(\mathbf{s}) \overset{iid}{\sim} N(0,1)$ (representing tooth-level). The corresponding regression coefficients are $\boldsymbol{\beta}^C = (0.0, 5, 0.0, 0.5)$.

The baseline hazard is Weibull with shape and scale parameter constant across sites and subjects and fixed at $\eta = 2$ and $\kappa = 3$. Data are right censored above a threshold which is chosen as the $1-\tau$ quantile of the sample survival times, such that $100\tau\%$ of the observations are censored. Spatial random effects are simulated following either the PS or the Gaussian model. For the PS models, we choose $L = 20$ and use bandwidth $\rho = 1$. For log Gaussian random effects, we set the variance to $\sigma^2 = 0.1^2$ and the spatial range to $\rho = 1$. Designs vary based on random effect distribution (PS or Gaussian), strength of spatial dependence (α), and censoring rate τ. This leads to 4 designs representing various data generation schemes:

1. **PS - weak dependence**: PS random effects with $\alpha = 0.7$ and $\tau = 0.3$
2. **PS - strong dependence**: PS random effects with $\alpha = 0.3$ and $\tau = 0.3$
3. **Gaussian**: Gaussian random effects with $\tau = 0.3$
4. **PS - weak with higher censoring**: PS random effects with $\alpha = 0.7$ and $\tau = 0.45$

We generate 100 datasets for each simulation design and fit the PS, Gaussian and independence models with the same priors as the MN data analysis. We compare models via mean squared error (MSE), Bias, and 90% Coverage Probability (CP) for the conditional regression coefficients $\boldsymbol{\beta}^C$. Considering our parameter space $\boldsymbol{\beta}^C = (\beta_1, \beta_2, \beta_3, \beta_4)$ with β_j being an element of $\boldsymbol{\beta}^C$ and $\hat{\beta_{ij}}$ the posterior estimate of β_j from the ith dataset, we calculate the respective quantities as $\text{MSE}(\hat{\beta_j}) = \frac{1}{100}\sum_{i=1}^{100}(\hat{\beta}_{ij} - \beta_j)^2$, $\text{Bias}(\hat{\beta_j}) = \frac{1}{100}\sum_{i=1}^{100}(\hat{\beta}_{ij} - \beta_j)$, and $\text{CP}(\hat{\beta_j}) = \frac{1}{100}\sum_{i=1}^{100} I(\beta_j \in [\hat{\beta}_{ij,LCL}, \hat{\beta}_{ij,UCL}])$, where I is an indicator function for β_j lying in the interval $[\hat{\beta}_{ij,LCL}, \hat{\beta}_{ij,UCL}]$ with $\hat{\beta}_{ij,LCL}$ and $\hat{\beta}_{ij,UCL}$ the estimated lower and upper endpoints of the 90% credible intervals, respectively, of β_j for the ith dataset. Tables 3 presents these estimates for each design and covariate.

From the MSE values, it is clear that even when the truth is PS with low censoring (Designs 1 and 2), the parameters representing the null effect ($\beta_1 = \beta_3 = 0$) have higher MSE for the PS as compared to the Gaussian and independence models. However, for non-null β ($\beta_2 = \beta_4 = 0.5$), PS enjoys lower MSE mostly. When the truth is Gaussian, the differences in MSEs between the PS and Gaussian models are much higher for the non-null β as compared to the null. Under Design 4 (PS truth with moderately heavy censoring), performance of the Gaussian model is somewhat better for the null parameters and slightly better for the non-null parameters with respect to MSE. Interestingly, the independence model enjoys lower MSE compared to the PS. Under Designs 1 and 2, there is substantial bias under the Gaussian and independence models as compared to the PS for non-null β. Similarly, under Gaussian truth, the bias is considerable for the PS model with respect to non-null β. Under Design 4, bias is higher for the Gaussian model for non-null beta. Once again, under Designs 1 and 2, the Gaussian and independence models exhibit substantial undercoverage across all parameters (except the tooth-level null β_3), with the performance worsening for non-null β under Design 2. Under Gaussian truth, both the PS and independence models exhibit undercoverage for the non-null β_2 and β_4. On the overall, the non-null effects are estimated more precisely by a model representing the ground truth. Surprisingly, this is not the case for the estimation of null effects mainly with respect to MSE and bias.

As pointed out by the Associate Editor, there are some very large MSE and bias and very poor coverages for the estimation of non-null effects in misspecified models under Designs 2 and 3. In particular, all three models perform poorly in terms of MSE, with the magnitude considerably higher (43.93) for the PS model under Gaussian truth compared to the Gaussian and Independence model under PS-strong truth. However, the biases are comparable in magnitude but with opposite signs, implying overestimation by the PS model and underestimation by the Gaussian and Independence models. From these, we observe that the variances of the PS estimates, overall, might be larger under misspecification. This can be partly explained by the heavier tails in the PS density compared to the Gaussian or the Independence (Nolan, 2015) models. The underestimation in the later two is also reflected in the extremely poor CPs, whereas the CPs for the PS under Gaussian truth is not that terrible. This rekindles our enthusiasm for the PS option. We cautiously summarize that estimation of conditional effects are highly sensitive to the underlying random effects distribution. In general, conditional effects are less robust compared to marginal effects (Boehm et al., 2013), which is intuitive since more information is available for estimating marginal effects than conditional effects.

7 Conclusion

In this paper, we propose a mixture cure PH model for spatially-referenced survival data under a Bayesian paradigm. Our model induces spatial dependence via random effects, and preserves the PH interpretation of the covariates (for the susceptible units) marginally over the random effects. The prior for spatial dependence is very flexible using a spatial factor approach. The method is used to analyze PD survival data of several subjects from a private periodontal practice. Our models can estimate the remaining lifespan of the susceptible tooth, as well as the survival probabilities for each tooth. Both simulation studies and application to real data reveal that the estimation of the conditional effects is highly sensitive to the underlying random effects distribution. Hence, model fit and model comparison steps are crucial.

Due to the high censoring rate for the MN data, we elect to use the parametric Weibull model to ensure identifiability. However, semiparametric baseline hazard models are also possible. For example, Ibrahim et al. (2001) use a piece-wise constant hazard $\lambda_0(t) = \sum_{j=1}^{J} \eta_j \mathrm{I}\left(T_{j-1} < t < T_j\right)$, where $T_0, ..., T_J$ are fixed change points and η_j is the baseline hazard during the interval (T_{j-1}, T_j). The permits great flexibility for the baseline hazard while maintaining the conditional and marginal PH interpretation of the regression coefficients.

Our mixture cure proportion is assumed constant for all subjects and spatial locations. However, it can also be covariate dependent, or even spatially. To keep our model simple in the face of heavy censoring, we did not consider these currently. Other cure-rate formulations (Cooner et al., 2006) can also be considered. Furthermore, this method uses baseline periodontal measurements. An important next step is to use longitudinal markers of periodontal health such as attachment loss and probing depth to further refine predictions. This is complicated because these markers themselves evolve over time, and thus we require new methods to combine these longitudinal measurements with survival data (e.g., Tsiatis and Davidian, 2004) while accounting for spatial dependence.

8 Acknowledgments

The authors thank the Joint Editor, the Associate Editor, and two reviewers whose constructive comments and suggestions led to a considerably improved version of the manuscript. This research was supported in part by the US National Institutes of Health grants R03DE021762 and R03DE023372 (Reich and Bandyopadhyay) and R01DE019656 (Nunn).

Appendix A.1: Marginal survival function

For notational convenience we drop the indices i and $\mathbf{s}$. Denoting $\theta = \sum_{l=1}^{L} w_l^{1/\alpha} A_l$ and $\mathbf{A} = (A_1, ..., A_L)$, and recalling that $\mathrm{E}[\exp(-tA_l)] = \exp(-t^\alpha)$ and $\sum_{l=1}^{L} w_l = 1$, the survival function

$S(t) = \mathrm{P}(Y > t)$ is

$$\begin{aligned}
S(t) &= \int \left\{ \pi + (1-\pi) \exp\left[- \left(\sum_{l=1}^{L} w_l^{1/\alpha} A_l \right) \exp(\mathbf{X}\boldsymbol{\beta}^C) H_0(t) \right] \right\} f(\mathbf{A}) d\mathbf{A} \\
&= \pi + (1-\pi) \prod_{l=1}^{L} \mathrm{E}\left(\exp\left\{ -[w_l^{1/\alpha} \exp(\mathbf{X}\boldsymbol{\beta}^C) H_0(t)] A_l \right\} \right) \\
&= \pi + (1-\pi) \prod_{l=1}^{L} \exp\left\{ -[w_l^{1/\alpha} \exp(\mathbf{X}\boldsymbol{\beta}^C) H_0(t)]^{\alpha} \right\} \\
&= \pi + (1-\pi) \exp\left\{ - \sum_{l=1}^{L} [w_l \exp(\alpha \mathbf{X}\boldsymbol{\beta}^C) H_0(t)^{\alpha}] \right\} \\
&= \pi + (1-\pi) \exp\left\{ -[\sum_{l=1}^{L} w_l] \exp(\mathbf{X}\boldsymbol{\beta}^M) H_0(t)^{\alpha}] \right\} \\
&= \pi + (1-\pi) \exp\left\{ - \exp(\mathbf{X}\boldsymbol{\beta}^M) H_0(t)^{\alpha} \right\}.
\end{aligned}$$

Differentiating and computing $-S'(t)/[S(t) - \pi]$ gives the hazard function in (4).

Appendix A.2: Joint survival function

The joint survival function for non-cured teeth (5) is

$$\begin{aligned}
\mathrm{P}[Y(\mathbf{s}_j) > t, Y(\mathbf{s}_k) > t] &= \int \exp\left\{ -[\theta(\mathbf{s}_j) + \theta(\mathbf{s}_k)] \exp(\mathbf{X}\boldsymbol{\beta}^C) H_0(t) \right\} p(\mathbf{A}) d\mathbf{A} \\
&= \int \exp\left[- \left(\sum_{l=1}^{L} [w_l(\mathbf{s}_j)^{1/\alpha} + w_l(\mathbf{s}_j)^{1/\alpha}] A_l \right) \exp(\mathbf{X}\boldsymbol{\beta}^C) H_0(t) \right] p(\mathbf{A}) d\mathbf{A} \\
&= \prod_{l=1}^{L} \exp\left\{ - \left[(w_l(\mathbf{s}_j)^{1/\alpha} + w_l(\mathbf{s}_k)^{1/\alpha}) \exp(\mathbf{X}\boldsymbol{\beta}^C) H_0(t) \right]^{\alpha} \right\} \\
&= \exp\left\{ - \exp(\mathbf{X}\boldsymbol{\beta}^M) H_0(t)^{\alpha} \sum_{l=1}^{L} \left[w_l(\mathbf{s}_j)^{1/\alpha} + w_l(\mathbf{s}_k)^{1/\alpha} \right]^{\alpha} \right\}.
\end{aligned}$$

Appendix A.3: min-stability

Let $Y_1(\mathbf{s})$,...,$Y_N(\mathbf{s})$ be N independent replications of the PS spatial process for non-cured teeth that defines the process $\tilde{Y}(\mathbf{s}) = A(N) + B(N) \min\{Y_1(\mathbf{s}), ..., Y_N(\mathbf{s})\}$. To show the process is min-stable, we must show that there exist $A(N)$ and $B(N)$ such that $\tilde{Y}(\mathbf{s})$ is identical in law to $Y(\mathbf{s})$. The

joint survival function of $Y(\mathbf{s})$ is

$$\mathrm{P}[Y(\mathbf{s}_j) > t_j, j = 1, ..., m] = \int \exp\left\{-\sum_{j=1}^{m} \theta(\mathbf{s}_j) \exp(\mathbf{X}_j \boldsymbol{\beta}^C) H_0(\mathbf{s}_j, t_j)\right\} p(\mathbf{A}) d\mathbf{A}. \quad (9)$$

Setting $A(N) = 0$, $B(N) = N^{\kappa}$, and observing that for the Weibull cumulative hazard $NH_0(\mathbf{s}_j, t_j N^{-\kappa}) = N\left[\frac{t_j N^{-\kappa}}{\eta(\mathbf{s}_j)}\right]^{\kappa} = \left[\frac{t_j}{\eta(\mathbf{s}_j)}\right]^{\kappa} = H_0(\mathbf{s}_j, t)$, the joint survival function of $\tilde{Y}(\mathbf{s})$ can be written as

$$\begin{aligned}
\mathrm{P}[\tilde{Y}(\mathbf{s}_j) > t_j, j = 1, ..., m] &= \mathrm{P}[Y_i(\mathbf{s}_j) > t_j N^{-\kappa}, j = 1, ..., m, i = 1, ..., N] \\
&= \mathrm{P}[Y(\mathbf{s}_j) > t_j N^{-\kappa}, j = 1, ..., m]^N \\
&= \int \exp\left\{-N \sum_{j=1}^{m} \theta(\mathbf{s}_j) \exp(\mathbf{X}_j \boldsymbol{\beta}^C) H_0(\mathbf{s}_j, t_j N^{-\kappa})\right\} p(\mathbf{A}) d\mathbf{A} \\
&= \int \exp\left\{-\sum_{j=1}^{m} \theta(\mathbf{s}_j) \exp(\mathbf{X}_j \boldsymbol{\beta}^C) H_0(\mathbf{s}_j, t_j)\right\} p(\mathbf{A}) d\mathbf{A},
\end{aligned}$$

which equals (9).

Appendix A.4: Outline of the conditional posterior distributions

The full likelihood (expanding 8) is given by

$$\begin{aligned}
&\mathcal{L}(\pi, \kappa^C, \eta^C, \boldsymbol{\beta}^C, \alpha, \Theta | X, Y, L, D) \\
&= \prod_{i=1}^{N} \prod_{j=1}^{m} \left(\delta_i(\mathbf{s}_j)\pi + (1-\pi) \left\{ \exp\left[-\left(\frac{L_i(\mathbf{s}_j)}{\eta(\mathbf{s}_j)} \right)^{\kappa^C} \theta_i(\mathbf{s}_j) \exp\left(\mathbf{X}_i(\mathbf{s}_j) \boldsymbol{\beta}^C \right) \right] \right\}^{\delta_i(\mathbf{s}_j)} \right. \\
&\left. \times \left\{ \kappa^C \eta(\mathbf{s}_j)^{-\kappa^C} \theta_i(\mathbf{s}_j) \exp\left[\mathbf{X}_i(\mathbf{s}_j) \boldsymbol{\beta}^C \right] Y_i(\mathbf{s}_j)^{\kappa^C - 1} \times \exp\left[-\left(\frac{L_i(\mathbf{s}_j)}{\eta(\mathbf{s}_j)} \right)^{\kappa^C} \theta_i(\mathbf{s}_j) \exp\left(\mathbf{X}_i(\mathbf{s}_j) \boldsymbol{\beta}^C \right) \right] \right\}^{1-\delta_i(\mathbf{s}_j)} \right),
\end{aligned}$$

where Θ is the matrix of spatial frailties $\theta_i(\mathbf{s}_j)$, Y is the matrix of observed time of tooth loss $Y_i(\mathbf{s}_j)$, X represent the design matrix of covariates, L is the matrix of censoring time $L_i(\mathbf{s}_j)$ (lower bound for tooth loss), and D is the matrix of censoring indicators $\delta_i(\mathbf{s}_j)$ for location $\mathbf{s}_j$ of individual i. Due to the mixture likelihood, none of the parameters can be factored out. Our sampling strategy substitutes the Gibbs steps with convenient Metropolis steps via. the Metropolis-within-Gibbs

(Tierney, 1994) sampling. The conditional posteriors for the parameters are presented below.

i) We use a Uniform$(0,1)$ prior for π, so

$$p(\pi|\text{rest}) \propto \mathcal{L}(\pi, \kappa^C, \eta^C, \boldsymbol{\beta}^C, \alpha, \Theta|X, Y, L, D) I(0 < \pi < 1)$$

ii) We use N$(0, \text{variance}=10^2)$ priors for β_k^C, $\log(\eta(\mathbf{s}_j))$, and $\log(\kappa^C)$, yielding

$$p(\beta_k^C|\text{rest}) \propto \mathcal{L}(\pi, \kappa^C, \eta^C, \boldsymbol{\beta}^C, \alpha, \Theta|X, Y, L, D) \times \exp\left[-\left(\beta_k^C\right)^2 / \left(2 \times 10^2\right)\right]$$

$$p(\eta(\mathbf{s}_j)|\text{rest}) \propto \mathcal{L}(\pi, \kappa^C, \eta^C, \boldsymbol{\beta}^C, \alpha, \Theta|X, Y, L, D) \times (\eta(\mathbf{s}_j))^{-1} \exp\left[-(\log \eta(\mathbf{s}_j))^2/(2 \times 10^2)\right]$$

$$p(\kappa^C|\text{rest}) \propto \mathcal{L}(\pi, \kappa^C, \eta^C, \boldsymbol{\beta}^C, \alpha, \Theta|X, Y, L, D) \times \left(\kappa^C\right)^{-1} \exp\left[-(\log \kappa^C)^2/(2 \times 10^2)\right]$$

iii) We use a N$(0,1)$ prior for logit(α), yielding

$$\begin{aligned} p(\alpha|\text{rest}) &\propto \mathcal{L}(\pi, \kappa^C, \eta^C, \boldsymbol{\beta}^C, \alpha, \Theta|X, Y, L, D) \\ &\times \prod_{i=1}^{N} \prod_{l=1}^{L} \left(\int_0^1 \left\{ \frac{\alpha A_{il}^{-1/(1-\alpha)}}{1-\alpha} \left(\frac{\sin(\alpha\pi B)}{\sin(\pi B)} \right)^{1/(1-\alpha)} \frac{\sin[(1-\alpha)]\pi B}{\sin(\alpha\pi B)} \right.\right. \\ &\times \left.\left. \exp\left[- \left(\frac{\sin(\alpha\pi B)}{\sin(\pi B)} \right)^{1/(1-\alpha)} \frac{\sin[(1-\alpha)]\pi B}{\sin(\alpha\pi B)} A_{il}^{-\alpha/(1-\alpha)} \right] \right\} dB \right) \\ &\times \frac{1}{\alpha(1-\alpha)} \exp\left[-\frac{1}{2} \left(\log \frac{\alpha}{1-\alpha} \right)^2 \right]. \end{aligned}$$

iv) The spatial random effects term $\theta_i(\mathbf{s}_j)$ is constructed as

$$\theta_i(\mathbf{s}_j) = \sum_{l=1}^{L} \left(\frac{\exp[K_l(\mathbf{s}_j)]}{\sum_{j=1}^{m} \exp[K_l(\mathbf{s}_j)]} \right)^{1/\alpha} A_{il}.$$

Hence, we outline the conditionals of its components A_{il} and $\boldsymbol{K}_l$ as follows:

$$
\begin{aligned}
p(A_{il}|\text{rest}) &\propto \mathcal{L}(\pi, \kappa^C, \eta^C, \boldsymbol{\beta}^C, \alpha, \Theta|X,\, Y, L, D) \\
&\times \int_0^1 \left\{ \frac{\alpha A^{-1/(1-\alpha)}}{1-\alpha} \left(\frac{\sin(\alpha\pi B)}{\sin(\pi B)} \right)^{1/(1-\alpha)} \frac{\sin[(1-\alpha)]\pi B}{\sin(\alpha\pi B)} \right. \\
&\times \left. \exp\left[- \left(\frac{\sin(\alpha\pi B)}{\sin(\pi B)} \right)^{1/(1-\alpha)} \frac{\sin[(1-\alpha)]\pi B}{\sin(\alpha\pi B)} A^{-\alpha/(1-\alpha)} \right] \right\} dB.
\end{aligned}
$$

and

$$
p(\boldsymbol{K}_l|\text{rest}) \propto \mathcal{L}(\pi, \kappa^C, \eta^C, \boldsymbol{\beta}^C, \alpha, \Theta|X,\, Y, L, D) \times \left[-\frac{1}{2} \boldsymbol{K}_l^T E^{-1} \boldsymbol{K}_l \right]
$$

where $E(\mathbf{s}_j, \mathbf{s}_k) = \exp\left[\sqrt{(s_{j1} - s_{k1})^2 + (s_{j2} - s_{k2})^2} \right]$.

References

Banerjee, S. and Carlin, B. P. (2004) Parametric spatial cure rate models for interval-censored time-to-relapse data. *Biometrics*, **60**, 268–275.

Banerjee, S. and Dey, D. K. (2005) Semiparametric proportional odds models for spatially correlated survival data. *Lifetime Data Analysis*, **11**, 175–191.

Banerjee, S., Gelfand, A. E. and Carlin, B. P. (2014) *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, 2nd edn.

Banerjee, S., Wall, M. and Carlin, B. (2003) Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*, **4**, 123–142.

Boehm, L., Reich, B. J. and Bandyopadhyay, D. (2013) Bridging conditional and marginal inference for spatially referenced binary data. *Biometrics*, **69**, 545–554.

CDC, National Center for Chronic Disease Prevention and Health Promotion (2011) *Oral Health: Preventing Cavities, Gum disease, Tooth loss, and Oral Cancers at a Glance 2011*. Atlanta, GA: CDC.

Chuang, S., Tian, L., Wei, L. and Dodson, T. (2001) Kaplan-meier analysis of dental implant survival: a strategy for estimating survival with clustered observations. *Journal of Dental Research*, **80**, 2016–2020.

— (2002) Predicting dental implant survival by use of the marginal approach of the semi-parametric survival methods for clustered observations. *Journal of Dental Research*, **81**, 851–855.

Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag London Ltd.

Cooner, F., Banerjee, S. and McBean, A. M. (2006) Modelling geographically referenced survival data with a cure fraction. *Statistical methods in medical research*, **15**, 307–324.

Fan, J., Nunn, M. and Su, X. (2009) Multivariate exponential survival trees and their application to tooth prognosis. *Computational Statistics and Data Analysis*, **53**, 1110–1121.

Fan, J., Xu, X., Levine, R., Nunn, M. and LeBlanc, M. (2006) Trees for correlated survival data by goodness of split with applications to tooth prognosis. *Journal of the American Statistical Association*, **101**, 959–967.

Flegal, J. M., Haran, M. and Jones, G. L. (2008) Markov chain monte carlo: Can we trust the third significant figure? *Statistical Science*, **23**, 250–260.

Geisser, S. and Eddy, W. (1979) A predictive approach to model selection. *Journal of the American Statistical Association*, **72**, 153–160.

Gelfand, A. E. and Dey, D. K. (1994) Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B*, **56**, 501–514.

Gelman, A. and Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, 457–472.

Härkänen, T., Larmas, M., Virtanen, J. and Arjas, E. (2002) Applying modern survival analysis methods to longitudinal dental caries studies. *Journal of Dental Research*, **81**, 144–148.

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

Hougaard, P. (1986) A class of multivanate failure time distributions. *Biometrika*, **73**, 671–678.

Ibrahim, J., Chen, M. and Sinha, D. (2001) *Bayesian survival analysis*. New York: Springer.

Li, Y. and Lin, X. (2006) Semiparametric normal transformation models for spatially correlated survival data. *Journal of the American Statistical Association*, **101**, 591–603.

Li, Y. and Ryan, L. (2002) Modeling spatial survival data using semiparametric frailty models. *Biometrics*, **58**, 287–297.

Liu, D., Kalbleeisch, J. and Schaubel, D. (2011) A positive stable frailty model for clustered failure time data with covariate dependent frailty. *Biometrics*, **67**, 8–17.

Lopes, H. F., Salazar, E. and Gamerman, D. (2008) Spatial dynamic factor analysis. *Bayesian Analysis*, **3**, 759–792.

Manda, S. O., Gilthorpe, M. S., Tu, Y.-K., Blance, A. and Mayhew, M. T. (2005) A bayesian analysis of amalgam restorations in the royal air force using the counting process approach with nested frailty effects. *Statistical Methods in Medical Research*, **14**, 567–578.

McGuire, M. K. and Nunn, M. E. (1996) Prognosis versus actual outcome II: The effectiveness of commonly taught clinical parameters in developing an accurate prognosis. *Journal of Periodontology*, **67**, 658–665.

Nolan, J. P. (2015) *Stable Distributions - Models for Heavy Tailed Data*. Boston: Birkhauser. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

Penny, R. E. and Kraal, J. H. (1979) Crown-to-root ratio: Its significance in restorative dentistry. *The Journal of Prosthetic Dentistry*, **42**, 34–38.

R Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model from extreme precipitation. *The Annals of Applied Statistics*, **6**, 1430–1451.

Shaby, B. A. and Reich, B. J. (2013) Bayesian spatial extreme value analysis to assess the changing risk of concurrent extremely high temperatures across large portions of European cropland. *Environmetrics*. In press.

Spiekerman, C. F. and Lin, D. (1998) Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, **93**, 1164–1175.

Stephenson, A. G. (2009) High-dimensional parametric modelling of multivariate extreme events. *Australian & New Zealand Journal of Statistics*, **51**, 77–88.

Sy, J. P. and Taylor, J. M. (2000) Estimation in a cox proportional hazards cure model. *Biometrics*, **56**, 227–236.

Tierney, L. (1994) Markov chains for exploring posterior distributions. *The Annals of Statistics*, **22**, 1701–1728.

Tsiatis, A. A. and Davidian, M. (2004) Joint modeling of longitudinal and time-to-event data: An overview. *Statistical Science*, **14**, 809–834.

Wong, M., Lam, K. and Lo, E. (2005) Bayesian analysis of clustered interval-censored data. *Journal of Dental Research*, **84**, 817–821.

Table 1: LPML values for various choices of PS models (varying with L), the Gaussian and the Independence model with and without cure proportions

Model	$L = 5$	$L = 10$	$L = 15$	$L = 20$	$L = 28$	Gaussian	Independence
cure	-536	-530	-485	-383	-437	-594	-648
non-cure	-519	-511	-499	-506	-497	-524	-654

Table 2: Posterior mean and 90% credible intervals of log hazard ratio parameters (β), the Weibull shape (κ) and scale (η), PS spatial dependence parameter (α), and cured proportion (π). 'PS' stands for positive stable and 'Indep' stands for independence model. 'Cond' and 'Marg' represents respectively, the conditional and marginal estimates.

Parameter	PS-Cond.	PS-Marg.	Gaussian-Cond.	Indep-Cond.
smoking status	7.02 (2.42,11.52)	0.77 (0.27, 1.27)	1.06 (0.69,1.43)	0.81 (0.23,1.41)
age	-0.02 (-0.12, 0.78)	-0.00 (-0.01, 0.01)	-0.01 (-0.03,0.00)	0.00 (-0.04,0.03)
poor hygiene	8.60 (2.79,14.26)	0.95 (0.31, 1.57)	1.16 (0.76,1.56)	1.43 (0.82,2.06)
good hygiene	-2.00 (-7.28, 3.10)	-0.22 (-0.80, 0.34)	-0.43 (-0.92,0.04)	-0.61 (-1.26,0.02)
crown-to-root ratio	10.36 (5.14,15.72)	1.14 (0.57, 1.73)	1.47 (1.09,1.87)	1.87 (1.34,2.4)
probing depth	1.43 (0.63, 2.18)	0.16 (0.07, 0.24)	0.23 (0.13,0.33)	0.45 (0.28,0.62)
mobility	6.40 (4.38, 8.83)	0.70 (0.48, 0.97)	0.77 (0.56,0.97)	0.94 (0.67, 1.21)
missing adjacent tooth	5.93 (3.09, 8.72)	0.65 (0.34, 0.96)	0.55 (0.15,0.96)	1.06 (0.46,1.68)
missing opposite tooth	-1.72 (-3.86, 1.87)	-0.13 (-0.42, 0.21)	-0.15 (-0.64,0.32)	0.40 (-0.28,1.04)
Shape	15.34 (11.18, 19.45)	1.69 (1.23, 2.14)	1.87 (1.68,2.07)	2.60 (2.16,3.13)
		PS	Gaussian	Indep
Scale, molar	–	135 (90, 202)	119 (57,213)	1135 (339,2536)
Scale, pre-molar	–	209 (138, 314)	180 (90,316)	1375 (422,3018)
Scale, canine	–	313 (198, 439)	267 (127,472)	1644 (492,3665)
Scale, incisor	–	269 (186, 363)	246 (119,435)	1599 (481,3516)
Positive stable	–	0.11 (0.08, 0.14)	–	0.3 (0.3,0.3)
Cured proportion	–	0.15 (0.07, 0.25)	0.26 (0.12,0.40)	0.53 (0.32,0.69)

Table 3: Mean squared error, bias and 90% coverage probabilities for estimating conditional regression coefficients from the simulation study.

(a) Mean squared error (multiplied by 100)

		Parameters			
Design	Model	$\beta_1 = 0$	$\beta_2 = 0.5$	$\beta_3 = 0$	$\beta_4 = 0.5$
1. PS - weak	PS	2.00	2.50	0.78	1.32
	Gaussian	1.09	3.55	0.47	2.97
	Independent	1.88	2.83	0.76	1.77
2. PS - strong	PS	11.92	12.19	0.87	1.69
	Gaussian	1.18	15.76	0.35	12.52
	Independent	2.21	11.40	0.56	11.29
3. Gaussian	PS	1.89	38.10	1.29	43.93
	Gaussian	0.59	0.64	0.41	0.47
	Independent	0.98	1.80	0.72	1.81
4. PS - higher censoring	PS	2.63	3.73	1.10	2.50
	Gaussian	0.96	3.52	0.66	2.31
	Independent	1.70	3.08	1.05	1.59

(b) Bias (multiplied by 100)

		Parameters			
Design	Model	$\beta_1 = 0$	$\beta_2 = 0.5$	$\beta_3 = 0$	$\beta_4 = 0.5$
1. PS - weak	PS	1.48	5.77	0.08	4.49
	Gaussian	1.33	-15.89	0.40	-15.86
	Independent	1.87	-9.89	0.91	-9.32
2. PS - strong	PS	-1.26	13.15	0.19	7.59
	Gaussian	-0.71	-32.9	0.56	-34.69
	Independent	-0.06	-30.17	1.11	-32.36
3. Gaussian	PS	1.09	28.77	-0.15	29.63
	Gaussian	0.53	-1.18	-0.11	-1.49
	Independent	0.01	8.07	-0.44	8.39
4. PS - higher censoring	PS	0.46	8.59	0.28	9.63
	Gaussian	0.02	-13.95	0.91	-12.28
	Independent	0.28	-8.12	0.62	-6.35

(b) Coverage probability of 90% intervals (multiplied by 100)

		Parameters			
Design	Model	$\beta_1 = 0$	$\beta_2 = 0.5$	$\beta_3 = 0$	$\beta_4 = 0.5$
1. PS - weak	PS	88	90	90	87
	Gaussian	69	33	89	19
	Independent	62	51	79	61
2. PS - strong	PS	90	86	92	85
	Gaussian	68	4	92	0
	Independent	58	11	91	1
3. Gaussian	PS	87	63	88	60
	Gaussian	85	85	91	86
	Independent	74	62	81	64
4. PS - higher censoring	PS	91	90	91	81
	Gaussian	84	47	81	49
	Independent	67	60	74	70

Figure 1: Kaplan-Meier survival plots for the full MN dataset, and also stratified by the 4 tooth types: molar, premolar, canine and incisor.

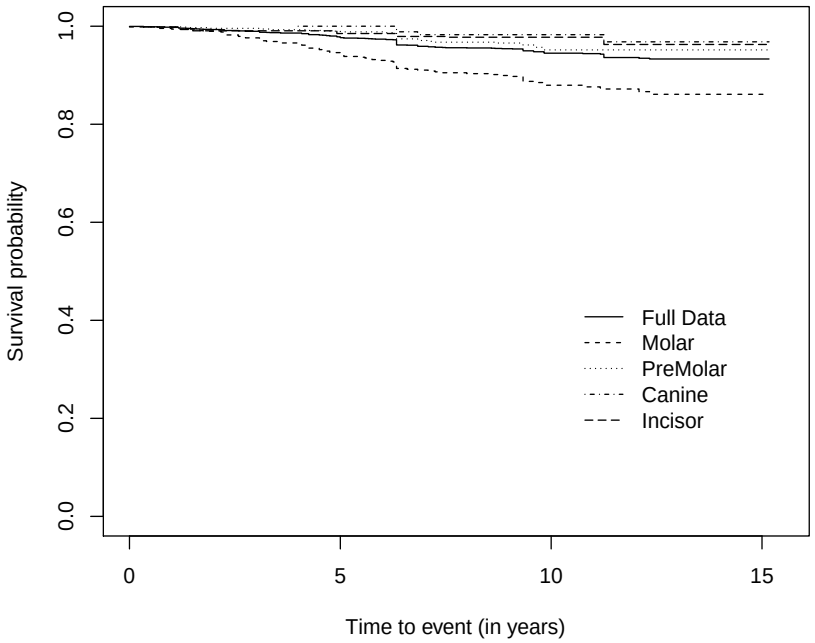

Figure 2: PIT_{ij} diagnostic for the final PS frailties model. The red line is the 45-degree line, and the dashed lines plot the observed versus expected quantiles for 10 random draws of probabilities for censored observations.

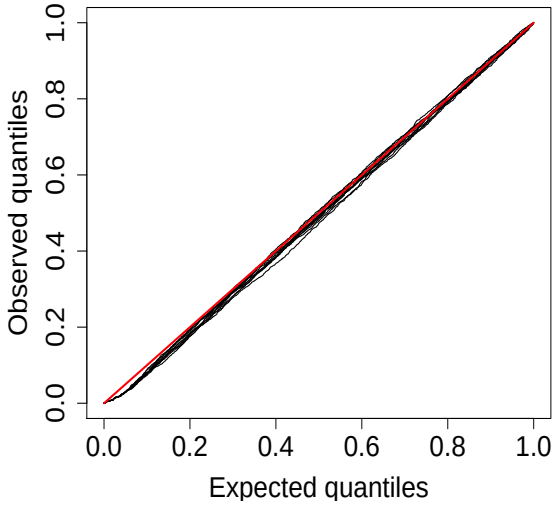

Figure 3: Posterior mean of the extremal coefficient $\vartheta(\mathbf{s}_j, \mathbf{s}_k)$ for all teeth pairs. The left panel is from the best-fitting PS non-cure model while the right panel is from the best fitting PS cure model. The posterior standard deviations are all less than 0.1.

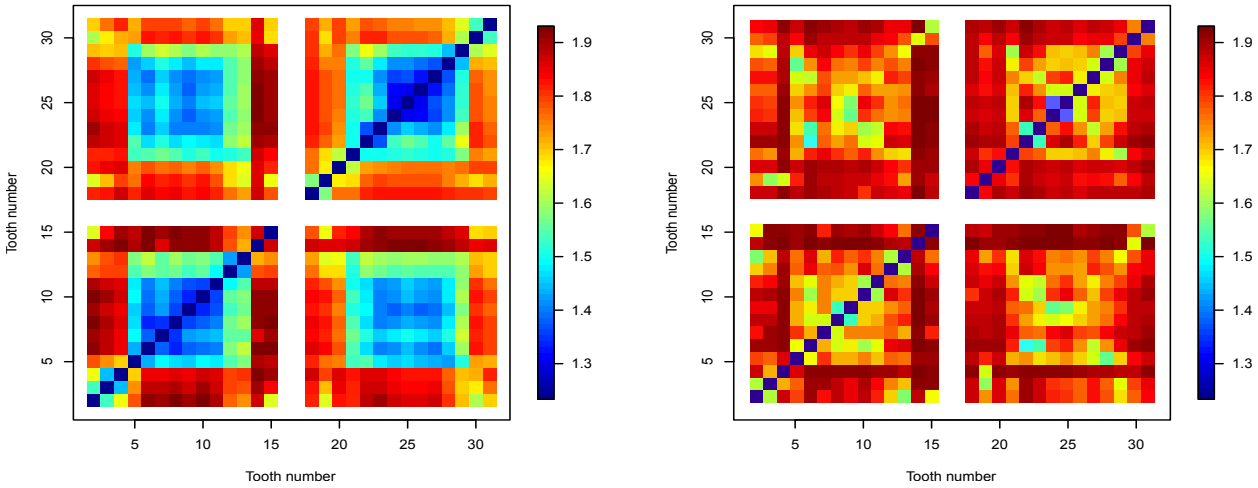

Figure 4: Covariates (probing depth and mobility), posterior mean residual life, and posterior survival probabilities (5- and 10-yrs) from the PS model for a subject with poor oral health. Vertical dashed lines represent teeth missing at baseline, and vertical solid lines represent teeth that fell out during the monitoring period. Means are thresholded at 100 years.