

Cleveland State University
EngagedScholarship@CSU



Mathematics Faculty Publications

Mathematics Department

2-1-2016

A Transformation Class for Spatio-temporal Survival Data with a Cure Fraction

Sandra M. Hurtado Rua
Cleveland State University, s.hurtadorua@csuohio.edu

Dipak K. Dey
University of Connecticut

Follow this and additional works at: https://engagedscholarship.csuohio.edu/scimath_facpub

 Part of the [Mathematics Commons](#), and the [Survival Analysis Commons](#)

How does access to this work benefit you? Let us know!

Repository Citation

Hurtado Rua, Sandra M. and Dey, Dipak K., "A Transformation Class for Spatio-temporal Survival Data with a Cure Fraction" (2016). *Mathematics Faculty Publications*. 268.
https://engagedscholarship.csuohio.edu/scimath_facpub/268

This Article is brought to you for free and open access by the Mathematics Department at EngagedScholarship@CSU. It has been accepted for inclusion in Mathematics Faculty Publications by an authorized administrator of EngagedScholarship@CSU. For more information, please contact library.es@csuohio.edu.

A transformation class for spatio-temporal survival data with a cure fraction

Sandra M Hurtado Rúa and Dipak K Dey

Abstract

We propose a hierarchical Bayesian methodology to model spatially or spatio-temporal clustered survival data with possibility of cure. A flexible continuous transformation class of survival curves indexed by a single parameter is used. This transformation model is a larger class of models containing two special cases of the well-known existing models: the proportional hazard and the proportional odds models. The survival curve is modeled as a function of a baseline cumulative distribution function, cure rates, and spatio-temporal frailties. The cure rates are modeled through a covariate link specification and the spatial frailties are specified using a conditionally autoregressive model with time-varying parameters resulting in a spatio-temporal formulation. The likelihood function is formulated assuming that the single parameter controlling the transformation is unknown and full conditional distributions are derived. A model with a non-parametric baseline cumulative distribution function is implemented and a Markov chain Monte Carlo algorithm is specified to obtain the usual posterior estimates, smoothed by regional level maps of spatio-temporal frailties and cure rates. Finally, we apply our methodology to melanoma cancer survival times for patients diagnosed in the state of New Jersey between 2000 and 2007, and with follow-up time until 2007.

1 Introduction

Survival curves are modeled with a pre-specified functional form which is built on a set of assumptions that are sometimes difficult to verify a priori. One of the most popular models in survival analysis is the Cox proportional hazard (PH) model,^{1,2} which assumes PHs across two sets of covariates. When the proportionality assumption does not hold, a generalized PH framework or a proportional odds (PO) model³ is usually considered. The generalized PH model

assumes that the proportional effect on the hazard function may be time varying, while the PO model adopts a time constant odds ratio between two sets of covariates.

Research addressing situations when the assumptions for the PH and PO models are violated exists in the literature.^{5,4} For example, a linear transformation is proposed by Cheng et al.⁵ Let T be the time to event, X the covariate vector, and β the corresponding regression parameter, then the linear transformation is given by equation (1), where $h(\cdot)$ is a strictly increasing function and $S(t|X)$ the survival function of t given the covariate X . If $\psi(y) = \log[-\log(y)]$, the linear transformation reduces to the PH model, and when $\psi(y) = -\text{logit}(y)$, equation (1) corresponds to the PO model

$$\psi\{S(t|X)\} = h(t) + \beta^T X \quad (1)$$

Model 1 assumes that all the subjects at risk will eventually experience the event of interest after a sufficiently long follow-up time. However, survival curves that flatten out at a value greater than zero are commonly observed in many areas including cancer survival times and epidemiological studies. A well-pronounced plateau is considered as evidence of the existence of a proportion of subjects for whom the event of interest will never occur. When analyzing time to event data, this proportion of patients is called cure rate. The most popular type of survival model incorporating this assumption is the mixture model.⁶ The mixture cure rate survival model formulates a population survival curve based on a latent random variable Y that controls the cure process, $Y=1$ indicates cure with probability θ , otherwise $Y=0$. Given $Y=0$, T 's are independently distributed $f(t|X)$, with corresponding survival function $S(t|X)$. Then, marginalizing over Y , the population survival mixture model, $S_p(t|X)$, is given by equation (2).

$$S_p(t|X) = \theta + (1 - \theta)S(t|X) \quad (2)$$

where θ is the cure fraction and $S(t|X)$ is a proper survival function for the susceptible individuals in the sense that $\lim_{t \rightarrow \infty} S(t|X) = 0$. Cure rate modeling has been studied in both frequentist⁷⁻⁹ and Bayesian¹⁰⁻¹² setups, and it has been extended in ways that facilitate biomedical interpretation of the cure process.^{11,12}

Model 1 also assumes that the survival times are independent, but in many applications the independence assumption may not be satisfied. Dependent survival times arise from clustering effects. For example, in cancer research, observations within the same county and year of diagnosis may be correlated. A frailty or random effect model accounts for correlation. It formulates the nature of the underlying dependence through a covariance specification and then the survival times are assumed to be independent conditionally on the unobserved frailty. Clayton¹³ uses a gamma frailty survival model without covariates, although this frailty model incorporates heterogeneity among clusters, it does not consider spatial or temporal correlations. Although spatial frailty models have been mainly used in other contexts, such as disease mapping,¹⁴⁻¹⁶ Banerjee and Dey¹⁷ model survival data accounting for spatial correlation and have shown that misleading results could be obtained using traditional non-spatial models when data exhibit a spatial structure. Conditionally autoregressive (CAR) models are the most popular areal models in the spatial literature.¹⁸⁻²⁰ Given a map, they model a correlation structure for the locations based on neighboring information.

It has been shown in the literature the need to model spatio-temporal clustered lifetime data with a flexible methodology, beyond the traditional cure rate PH model, yet there is not methodology that comprehensibility integrates all these assumptions. In this article, we propose a model for spatio-temporally arranged survival data that integrates cure rates and a flexible transformation.

We do not assume any proportionality of the survival curve, instead we use a single parameter based transformation on $S_p(t|\cdot)$. The benchmark PH and PO models are special cases of our transformation model. Here, $S_p(t|\cdot)$ is modeled as a function of a baseline cumulative distribution function (cdf), cure rates, and spatio-temporal frailties. The likelihood function is formulated assuming that the single parameter controlling the transformation is unknown (it needs to be estimated) and full conditional distributions are derived. A Markov chain Monte Carlo (MCMC) procedure is proposed for Bayesian estimation of the single parameter that controls the transformation, cure rates and spatio-temporal frailty parameters.

The format of this article is as follows. Section 2 introduces our transformation class of spatio-temporal cure rate survival models and its properties. Section 3 describes the likelihood formulation, posterior properties and Bayesian modeling, including model diagnosis and the description of the MCMC technique used. Section 4 is dedicated to the Bayesian analysis of a melanoma cancer data set from the Surveillance, Epidemiology and End Results (SEER) database. Finally, conclusions and discussion points are presented in Section 5.

2 The model

2.1 Motivation and general formulation

Assume that the cure process for a particular subject is determined by N , the number of latent competing risks¹² (i.e. carcinogenic cells left active after a treatment phase) and Y_1, Y_2, \dots, Y_N , the latent competing risks (i.e. the incubation times for each carcinogenic cell). Each N is assumed to have a probability mass function $P_\theta(N=n)$ and Y_1, Y_2, \dots, Y_N are assumed to be independent and identically distributed (iid) with common survival function $S(y) = 1 - F_0(y)$. The time to event, T , (i.e. time to death due to colon cancer) is defined by $T = \min\{Y_i, 0 \leq i \leq N\}$, where $P(Y_0 = \infty) = 1$. Assuming that N is independent of $Y_i \forall_i$, the survival function for T is given by $S_p(t) = \sum_{n=0}^{\infty} P(Y_1 > t, \dots, Y_n > t | N = n) P_\theta(N = n)$, which reduces to

$$S_p(t) = \sum_{n=0}^{\infty} [1 - F_0(t)]^n P_\theta(N = n) = M_\theta(\ln[1 - F_0(t)]) \quad (3)$$

where $M_\theta(\cdot)$ is the moment generating function of $N|\theta$. For example, if $N|\theta \sim \text{Poisson}(\theta)$, then $S_p(t) = \exp[-\theta F_0(t)]$ which is the model proposed by Chen et al.¹² Clearly, $S_p(t) = \exp[-\theta F_0(t)]$ yields to a PH structure and the cure fraction is given by $\exp(\theta)$. There is also a mathematical relationship between models 3 and 2. Assuming that $N=0$ with probability θ and $N=1$ with probability $1 - \theta$, model 3 reduces to the mixture cure rate model⁶ given by equation (2).

A more general transformation for $M_\theta(\ln[1 - F_0(t)])$ in equation (3) can be defined as

$$S_p(t) = \psi[F_0(t)\theta] \quad (4)$$

where $\psi(\cdot)$ is a known monotone decreasing function on $[0, 1]$, $F_0(t)$ is a proper CDF such that $F_0(0) = 0$ and $F_0(\infty) = 1$, and θ is the parameter that models the cure fraction. Because $S_p(\infty) = \psi(\theta) > 0$, equation (4) is not a proper survival function and $\psi(\theta)$ is the cure rate. Since $\psi(\cdot)$ is a monotone decreasing function on $[0, 1]$, as $\theta \rightarrow \infty$, the cure fraction $\psi(\theta) \rightarrow 0$ while as $\theta \rightarrow 0$, the cure fraction $\psi(\theta) \rightarrow 1$.

The corresponding population density and hazard functions are, respectively, given by $f_p(t) = -\psi'[F_0(t)\theta]f(t)\theta$ and $h_p(t) = -\frac{\psi'[F_0(t)\theta]f(t)\theta}{\psi[F_0(t)\theta]}$, where $\psi'(\cdot) = \frac{d\psi(\cdot)}{d\theta}$ and $f(t) = \frac{dF_0(t)}{dt}$. Similarly, $f_p(t)$

and $h_p()$ are not a proper probability density and a proper hazard functions, respectively, because $S_p(\infty) \neq 0$; however, $h_p()$ is integrable ($\int_0^\infty h_p() < \infty$). On the other hand, $f()$ and $h()$ are a proper probability density and a proper hazard functions, respectively.

The transformation class of models given by equation (4) is a broader family of cure rate models, for example, consider a Box–Cox type of transformation¹⁰ indexed by a unique parameter $\alpha \in [0, \infty)$ like the one shown in equation (5). Note that $\lim_{\alpha \rightarrow 0} (1 + \alpha x)^{-\frac{1}{\alpha}} = \exp(-x)$, additionally when $\alpha = 1$, equation (5) has a PO structure model, while $\alpha = 0$ is a PH model. A transformation class of models given by equations (4) and (5) is a rich family of models beyond the traditional PO and PH models

$$\psi(x) = \begin{cases} (1 + \alpha x)^{-\frac{1}{\alpha}}, & \alpha > 0 \\ \exp(-x), & \alpha = 0 \end{cases} \quad (5)$$

The model defined by equations (4) and (5) is identifiable under the conditions listed in Theorem 2.1. Consider the class of models of the form: $\mathcal{H} = \{S_p(t|z) = \psi[F(t)\theta(z)], t < C; F(t) \in \mathcal{F}; 0 < \theta(z) < \infty \forall z \in \mathcal{Z}\}$, where $\psi(\cdot)$ is given by equation (5), C is the censoring time, \mathcal{Z} the design space, and $\mathcal{F} = \{F(t) : F(t) \text{ a proper cumulative distribution on } [0, +\infty)\}$. Theorem 2.1 lists some conditions under which the proposed model is identifiable. The proof of Theorem 2.1 is given in Appendix 1.

Theorem 2.1 *Assume that $\theta(z) = \exp(\delta + \beta z)$, then*

- (1) *If α is assumed to be known and $F(t) \equiv F(t|\gamma)$, where γ is a parameter or vector of parameters, the model $S_p(t|z) = \psi[F(t)\theta(z)], t < C$ is identifiable for all $\theta(z)$ (unspecified or specified by any link function).*
- (2) *If α is assumed to be known and $F(t)$ is unspecified (i.e. modeled non-parametrically), the model $S_p(t|z) = \psi[F(t)\theta(z)], t < C$ is not identifiable unless $\delta = 0$.*
- (3) *If $\alpha > 0$ is unknown and $\delta = 0$, then the model $S_p(t|z) = \psi[F(t)\theta(z)], t < C$ is identifiable for all $F(t)$ (unspecified or parametrically specified).*

2.2 Model extension for spatio-temporal clustered survival data

We extend the model given by equations (4) and (5) to a spatio-temporal clustered survival setting. Assume that there are n_{jk} subjects located at the j th region who were diagnosed in the k th year where $j = 1, \dots, J$ and $k = 1, \dots, K$. The observed response is $t_{ijk} = \min\{T_{ijk}, C_{ijk}\}$, where T_{ijk} is the survival time for the i th subject located at the j th region who was diagnosed in the k th year for $i = 1, \dots, n_{jk}$ and C_{ijk} the censoring variable. Suppose we observe for the ijk individual a $L \times 1$ vector of covariates Z_{ijk} and that the baseline CDF $F_{0j}(\cdot) \equiv F_{0j}$ for $j = 1, \dots, J$ (i.e. the baseline CDF's are spatially dependent).

Using the general class of transformation discussed in Section 2.1, the population survival curve with spatio-temporal frailties and cure rates is given by equation (6)

$$S_p(t|Z_{ijk}, W_{jk}) = \psi[F_{0j}(t)\theta(\gamma, Z_{ijk}) \exp(w_{jk})] \quad (6)$$

where $\psi()$ is given by equation (5), w_{jk} is the spatio-temporal frailty terms, and $\theta(\gamma, Z_{ijk})$ is an increasing positive link function that models the cure fraction. Since $S_p(\infty) = \psi[\theta(\gamma, Z_{ijk}) \exp(w_{jk})]$ and $\theta(\gamma, Z)$ is an increasing positive link function, we can interpret the role of the

regression coefficients in the cure fraction. A large negative regression coefficient leads to a small value of θ which implies a larger cure fraction when the covariates are positive.

When $\alpha = 0$, the population survival function reduces to equation (7) which is an extension of the cure rate PH model¹² to a spatio-temporal survival model with cure fraction. The cure fraction from equation (7) is given by $S_p(\infty) = \exp[\theta(\gamma, Z_{ijk}) \exp(w_{jk})] = [\exp(\theta(\gamma, Z_{ijk}))]^{\exp(w_{jk})}$

$$S_p(t|X_{ijk}, Z_{ijk}, W_{jk}) = \exp[-F_{0j}(t)\theta(\gamma, Z_{ijk}) \exp(w_{jk})] \quad (7)$$

If $\alpha = 1$, equation (6) reduces to equation (8) which is a spatio-temporal PO survival model with cure fraction where $S_p(\infty) = [1 + \theta(\gamma, Z_{ijk}) \exp(w_{jk})]^{-1}$

$$-\log \left[\frac{S_p(t|Z_{ijk}, W_{jk})}{1 - S_p(t|Z_{ijk}, W_{jk})} \right] = \log F_{0j}(t) + \log \theta(\gamma, Z_{ijk}) + w_{jk} \quad (8)$$

The parameters that control the cure process, $\theta(\cdot)$, and the spatio-temporal frailties, w_{jk} , are modeled through a hierarchical specification. The cure rates are controlled by the positive increasing link function, $\theta(\gamma, Z_{ijk})$. Although many link functions can be used, we model $\theta(\gamma, Z_{ijk})$ as

$$\theta(\gamma, Z_{ijk}) = \exp[\mathbf{z}'_{ijk}\boldsymbol{\gamma}] \quad (9)$$

where $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_L)$ (model without intercept). Some authors⁹ have worked with the exponential link function given by equation (9). One advantage of the exponential link is that it can be interpreted in the context of a canonical link for a Poisson regression model.¹² We prove that a model with exponential link function is identifiable under different setups (Theorems 2.1 and 3.1).

To capture the spatio-temporal relationships, we introduced the random effects w_{jk} in equation 6. Let us define $W_k = (w_{1k}, \dots, w_{Jk})'$ as the frailty for the k th year. In the absence of spatial correlation, the elements of W_k are modeled as iid random variables for all k . However, here, we allow $\{w_{1k}, \dots, w_{Jk}\}$ to be spatially correlated across the regions with a suitable prior specification for each k . Under the assumption of spatial correlation among regions, the distribution for the frailty for the k th year needs to reflect this structure. A popular choice is the CAR distribution.²⁰ We assume that W_k have a CAR distribution with parameters (ξ, σ_k^2) . To define this, consider that the $J \times 1$ vector of spatial frailties for the k th year is assumed to have a multivariate normal distribution. Let $M = \{m_{ij}\}$ be the adjacency matrix of the graph of the geographical region, i.e. let $m_{ij} = 1$, if the j th and i th counties are neighbors and 0 otherwise. Define $m_{j+} = \sum_b m_{jb}$, the number of neighbors for the j th county and $D_M = \text{Diag}\{m_{j+}\}$. Assume that $0 < \xi < 1$ is a smoothing parameter, which can be viewed as a measure of spatial association in the sense that $\text{Cov}(w_{jk}, w_{j'k} | \xi, \sigma^2) = \frac{\xi \sigma_k^2 m_{ij}}{m_{j+} m_{j'+} + \xi m_{ij}}$. Now, define $\Sigma(\xi, \sigma_k^2) = \sigma_k^2 [D_M - \xi M]^{-1}$, then

$$W_k | \xi, \sigma_k^2 \sim N_J(\mathbf{0}, \Sigma(\xi, \sigma_k^2)) \quad (10)$$

Furthermore, the vectors of spatial frailties W_k , for $k = 1, \dots, K$, are assumed to be correlated through a prior specification for σ_k^2 .

In this section, we have derived a transformation class of cure rate models, listed its properties and extended to a spatio-temporal clustered survival setting. We also presented the conditions for the model to be identifiable. In the following section, we propose a Bayesian estimation methodology and study the properties of the posterior distribution.

3 Bayesian estimation

3.1 Likelihood, posterior distribution, and properties

Bayesian inference provides a viable alternative to jointly estimate the transformation parameter α , the spatio-temporal frailties w_{jk} , and the cure rate related parameters γ . We fit the model in a Bayesian setting for right censored data as follows: Let $\delta_{ijk} = \mathbb{I}(t_{ijk} < C_{ijk})$, where $\mathbb{I}(\cdot)$ is the indicator function. The observed data are given by $\mathfrak{D} = \{t_{ijk}, \delta_{ijk}, \mathbf{z}_{ijk}, M\}$. The ijk contribution to the conditional likelihood function given W_K is of the form $L_{ijk}(\psi, F_{0j}, \gamma | W_k, \mathfrak{D}) =$

$$\left\{ \frac{\psi' \left[F_{0j}(t_{ijk}) \exp(\mathbf{z}'_{ijk} \gamma + w_{jk}) \right] f_{0j}(t) \exp(\mathbf{z}'_{ijk} \gamma + w_{jk})}{\psi \left[F_{0j}(t) \exp(\mathbf{z}'_{ijk} \gamma + w_{jk}) \right]} \right\}^{\delta_{ijk}} \times \psi \left[F_{0j}(t) \exp(\mathbf{z}'_{ijk} \gamma + w_{jk}) \right] \quad (11)$$

Since $\frac{\psi'(x)}{\psi(x)} = -(1 + \alpha x)^{-1} \mathbb{I}\{\alpha > 0\}$, equation (11) reduces to equation (12) for $\alpha \in (0, +\infty)$. The likelihood construction for the case when $\alpha = 0$ is given in Appendix 2.

$$L_{ijk}(\alpha, \gamma, F_{0j} | \mathbf{W}, \mathfrak{D}) = \left\{ 1 + \alpha F_{0j}(t) \exp(\mathbf{z}'_{ijk} \gamma + w_{jk}) \right\}^{-\frac{1}{\alpha}} \times \left\{ \frac{f_{0j}(t) \exp(\mathbf{z}'_{ijk} \gamma + w_{jk})}{1 + \alpha F_{0j}(t) \exp(\mathbf{z}'_{ijk} \gamma + w_{jk})} \right\}^{\delta_{ijk}} \quad (12)$$

We assume that the variables T_{ijk} and C_{ijk} are conditionally independent given \mathbf{z}_{ijk} . We also assume that the observations in the same spatio-temporal cluster jk are dependent but exchangeable. Let $\mathbf{F}_0 = (F_{01}, \dots, F_{0J})$, $\gamma = (\gamma_1, \dots, \gamma_L)$, and assume that $0 < \xi < 1$, $\sigma_k > 0 \forall_k$, all the priors are independent and that the prior of α is proper. Then, the joint posterior distribution of $(\alpha, \gamma, \mathbf{F}_0)$ based on \mathfrak{D} is given by

$$\pi(\alpha, \gamma, \mathbf{F}_0 | \mathfrak{D}) \propto \prod_{k=1}^K \prod_{j=1}^J \prod_{i=1}^{n_{jk}} L_{ijk}(\alpha, \gamma, \mathbf{F}_0 | \mathbf{W}_k, \mathfrak{D}) [\mathbf{W}_k | \xi, \sigma_k^2] \pi(\alpha) \pi(\gamma) \pi(\mathbf{F}_0) \quad (13)$$

where $\pi(\alpha)$, $\pi(\gamma)$, and $\pi(\mathbf{F}_0)$ are the priors of α , γ , and \mathbf{F}_0 , respectively, and $[\mathbf{W}_k | \xi, \sigma_k^2]$ given by equation (10). Prior to drawn inferences, we investigate the posterior properties of equation (13). The posterior propriety under an improper prior for γ established in Theorem 3.1 implies that under certain conditions for \mathbf{F}_0 the model is identifiable and the estimation of the regression coefficients γ that control the cure rates can contain little subjective information. The proof of Theorem 3.1 is given in Appendix 3.

Theorem 3.1: *Suppose that Z^* is a $n \times L$ matrix with rows $\delta_{ijk} z'_{ijk}$, $n = \sum_{ijk} n_{ijk}$, $d = \sum_{ijk} \delta_{ijk}$, and $\pi(\gamma) \propto 1$. Assume that (a) Z^* is full rank, (b) $\pi(\alpha)$ is proper, (c) $0 < \xi < 1$ and $\sigma_k > 0 \forall_k$, and (d) the ratio $\frac{f_{0j}(t)}{F_{0j}(t)}$ is bounded for all $j = 1, \dots, J$ and $0 < t < C$, then the posterior (13) is proper.*

Condition (a) of Theorem 3.1 is frequently required in many regression setups and commonly satisfied by many data sets. Condition (b) indicates that we need some prior information about the parameter α which can be easily elicited. Condition (c) ensures a CAR model for the spatio-temporal frailties W . Finally, condition (d) is a more technical condition but not as restrictive as it seems. Some parametric specifications for \mathbf{F}_0 satisfy condition (d), for example, if the baseline CDF is based on the exponential (λ) model, $F_{0j}(t) \equiv F_{0j}(t | \lambda)$, $\infty < \lambda < \infty$. It can be shown that there exists a constant $R > 0$ such that $\frac{f_{0j}(t)}{F_{0j}(t)} = \frac{e^\lambda \exp\{-e^\lambda t\}}{1 - \exp\{-e^\lambda t\}} < R \forall_t > \epsilon > 0$. Additionally, for a Weibull distribution with parameters (ζ, λ) , it can be shown that $\frac{f_{0j}(t)}{F_{0j}(t)} = \frac{\zeta t^{\zeta-1} e^\lambda \exp\{-e^\lambda t^\zeta\}}{1 - \exp\{-e^\lambda t^\zeta\}} \leq R \max\{t^{-1}\} \zeta$, which under a

proper hyperprior for ζ will lead to a proper posterior under the conditions of Theorem 3.1. When F_0 is estimated non-parametrically, $\frac{f_{0j}(t)}{F_{0j}(t)} < R$ for some $R > 0$, for all $j = 1, \dots, J$ and $0 < t < C$ guarantees a proper posterior even when $\pi(\gamma) \propto 1$, but note that the conditions are sufficient but not necessary.

3.2 Model implementation: an example

To implement model (11), we need to specify \mathbf{F}_0 and priors and hyperpriors. In this particular section, we describe a model implementation with a non-parametric model for \mathbf{F}_0 , and weakly informative but proper priors and hyperpriors for all the parameters.

We first turn our attention to the modeling of $F_{0j}(t)$. We model $F_{0j}(t)$ using a dense class of monotone transformations arising from mixture of beta distribution functions.²¹ Since any continuous density in $[0, 1]$ can be approximated as a discrete mixture of beta densities,²² we consider the function $J_{0j}(t) = \frac{a_0 F_{0j}(t)}{a_0 F_{0j}(t) + b_0}$ which is increasing and maps $[0, 1]$ into $[0, 1]$, where $a_0 > 0$ and $b_0 > 0$. $J_{0j}(t)$ can be modeled as a mixture of $Beta(r_e, s_e)$ CDFs as follows $J_{0j}(t) = \sum_{e=1}^E \rho_{ej} IB(\tilde{J}_{0j}(t); r_e, s_e)$, where $\sum_{e=1}^E \rho_{ej} = 1$; \forall_j , $IB(\cdot; r_e, s_e)$ denotes the incomplete beta function, and $E, a_0, b_0, \tilde{J}_0(t), r_e$, and s_e are fixed. $F_{0j}(t)$ and $f_{0j}(t)$ are then defined by

$$F_{0j}(t) = \frac{b_0 \sum_{e=1}^E \rho_{ej} IB(\tilde{J}_{0j}(t); r_e, s_e)}{a_0 \left[1 - \sum_{e=1}^E \rho_{ej} IB(\tilde{J}_{0j}(t); r_e, s_e) \right]} \quad (14)$$

$$f_{0j}(t) = \frac{b_0 \frac{\partial}{\partial t} \tilde{J}_{0j}(t) \sum_{e=1}^E \rho_{ej} Beta(\tilde{J}_{0j}(t); r_e, s_e)}{a_0 \left[1 - \sum_{e=1}^E \rho_{ej} IB(\tilde{J}_0(t); r_e, s_e) \right]^2} \quad (15)$$

The values of $\{(r_e, s_e)\}$ are chosen such that the $Beta(r_e, s_e)$ CDF's have equally spaced means, and are centered around $\tilde{J}_0(t)$. $\tilde{J}_0(t)$ is chosen such that $\tilde{J}_{0j}(t) = \frac{a_0 F_{0j}(t)}{a_0 F_{0j}(t) + b_0}$ and a suitable hyper-prior is assigned to $\tilde{F}_{0j}(t)$. E, a_0 , and b_0 are usually fixed.

The second modeling issue is the selection of the priors. We assume that γ , ρ , and α are independent and their components are also independent a priori. Specifically, we take $\gamma \sim \mathcal{N}_L(0, \sigma_\gamma^2 I_L)$, a normal distribution with mean zero and covariance matrix $\sigma_\gamma^2 I_L = \text{Diag}(\sigma_\gamma^2)$, where I_L is the identity $L \times L$ matrix. $\rho_j \sim \mathcal{D}(\frac{1}{E} \mathbf{1}_E) \forall_j$, a Dirichlet prior where $\mathbf{1}_E$ is the unit vector of size E . To facilitate the interpretability of the parameters that control the cure rate, we restrict the parameter α to $[0, 1]$. α is then assumed to have a prior beta distribution, $\alpha \sim \mathcal{B}(a_\alpha, b_\alpha)$. The prior specification for each vector of frailties W_k is given by equation (10) and $W_k | \sigma_k^2, \xi$, is assumed to be independent a priori \forall_k . The hyperpriors for $\sigma^2 = (\sigma_1^2, \dots, \sigma_K^2)$, and ξ are assumed to independent a priori, and their components are also independent, such that $\sigma_k^2 \sim \mathcal{IG}(a_\sigma, b_\sigma)$, \forall_k and $\xi \sim \mathcal{B}(a_\xi, b_\xi)$.

To implement a Gibbs sampler with Metropolis–Hastings steps, we need to obtain all the conditional densities for the parameters α , γ , ρ , \mathbf{W} , σ , and ξ . For $j = 1, \dots, J$, $k = 1, \dots, K$, the full conditional distribution functions are given as follows

$$[\alpha | \gamma, \rho, \mathbf{W}, \mathcal{D}] \propto L(\alpha, \gamma, \rho | \mathbf{W}, \mathcal{D}) \alpha^{a_\alpha} (1 - \alpha)^{b_\alpha}$$

$$[\gamma | \alpha, \rho, \mathbf{W}, \mathcal{D}] \propto L(\alpha, \gamma, \rho | \mathbf{W}, \mathcal{D}) \exp \left[-\gamma' \Sigma_\gamma^{-1} \gamma \right]$$

$$\begin{aligned}
[\rho_j | \alpha, \gamma, \rho_{(-j)}, \mathbf{W}, \mathfrak{D}] &\propto L(\alpha, \gamma, \rho | \mathbf{W}, \mathfrak{D}) \prod_{e=1}^M \rho_{je}^{\frac{1}{e}} \\
[W_k | \alpha, \gamma, \rho, W_{(-k)}, \sigma^2, \xi, \mathfrak{D}] &\propto L(\alpha, \gamma, \rho | \mathbf{W}, \mathfrak{D}) \exp \left[-\frac{1}{2} \sum_{k=1}^K \frac{1}{\sigma_k^2} W'_k [D_M - \xi M] W_k \right] \\
[\xi | \mathbf{W}, \sigma^2, \mathfrak{D}] &\propto \\
\left[\prod_{k=1}^K (\sigma_k^2)^{-J/2} \right] \det [D_M - \xi M]^{K/2} \exp \left[-\frac{1}{2} \sum_{k=1}^K \frac{1}{\sigma_k^2} W'_k [D_M - \xi M] W_k \right] &\xi^{a\xi} (1 - \xi)^{b\xi} \\
[\sigma_k^2 | \xi, \mathbf{W}_k, \mathfrak{D}] &\propto [\sigma_k^2]^{-(a_\sigma + \frac{J}{2} + 1)} \exp \left[-\frac{1}{\sigma_k^2} \left(\frac{1}{2} W'_k [D_M - \xi M] W_k + b_\sigma \right) \right]
\end{aligned}$$

where $\rho_{(-j)}$ represents all the vectors $\rho_{(i)}$, $i \neq j$, and $W_{(-k)}$ all the vectors $W_{(i)}$, $i \neq k$.

The proposed model (6) induces a flexible family of competing models for several modeling choices of $F_{0j}(t)$, α , and ξ ; so, a model selection approach needs to be implemented. Model selection is a procedure to compare a class of competing models, and select the one that best fits the data. Our model selection is based on the conditional predictive ordinate (CPO) and deviance information criterion (DIC) measures.

The CPO²³ is extensively used in the literature. Large values of CPO_{ijk} are indicators of a better fit. The sum of the logarithms of the pseudomarginal-likelihood (LPML) measures the goodness of fit and is given by $LPML = \sum_{i,j,k} \log CPO_{ijk}$, where large values indicate better fit. LPML is a well-defined statistics as long as the posterior density is proper, so it is possible to find it even when the prior distribution is improper. It is easy to compute from the MCMC samples and it is computationally stable. The DIC²⁴ is a penalized likelihood criteria, such as the Akaike information criterion. The DIC is based on the posterior distribution of the deviance statistics and it is defined as $DIC = \overline{Dev} + p_{Dev}$, where \overline{Dev} is the posterior mean of the deviance and p_{Dev} a penalty term for model complexity and depends on the total number of model parameters.

4 Application

We illustrate our methodology with a melanoma cancer data set from the National Cancer Institute SEER database.²⁵ The cohort extracted is a population of 10,337 subjects who were diagnosed with melanoma cancer in one of the counties of the state of New Jersey between 2000 and 2007 and had been undergoing treatment since diagnosis. In the state of New Jersey, hospitals, physicians, and laboratories located in all counties are required to report primary cancer cases along with clinical, temporal, and geographical variables. All the cases are followed annually and vital status is recorded. For deceased cases, the underlying cause of death is also included. One of the goals of New Jersey registry is to describe cancer patterns in the state to adopt programs and services that increase survival and identify risk factors. The lifetimes here are naturally clustered by region (counties) and year of diagnosis and inferences about survival curves and cure rates can be used by health care providers, public health officials, and administrators to adopt new programs, educational campaigns, and services.

All the subjects were diagnosed with only melanoma cancer and were followed up since diagnosis until December 2007. Subjects who have died due to melanoma cancer were considered failed and the rest (those who die due to other causes, drop out, or survive until the end of the study) were

considered censored. By the end of 2007, 1480 patients have died of melanoma cancer (14%) while the remaining were censored.

The variables considered in this analysis are lifetime in months since diagnosis, year of diagnosis, and county of residence while undergoing treatment (Table 1). Subjects who moved during the treatment phase were removed from the data set. The patient level covariates included were: gender, race (white, black, or other), age, stage of the disease (local, regional, or distant), radiation therapy (whether or not the patient received radiation therapy), surgery (whether or not the patient received surgery), marital status (married or divorced, separated, widowed, single), and median family income on the county of residence (in thousands). A clinical interaction of interest is the interaction between radiation and stage since radiation is a treatment used mainly for melanoma patients whose stage at diagnosis is not local. It is also known that surgery is the first course of action for melanoma treatment, so it may be of interest to analyze if radiation and surgery interact.

The marginal estimated Kaplan–Meier curves for all the counties in the state of New Jersey show a plateau in the survival curve (see, for example, Figure 1), thus a cure rate model appears to be suitable for these data. Kaplan–Meier curves also show regional differences on the survival curve which can be explained using spatially correlated frailties. Survival curves over a long period of time are usually not time invariant so temporal components should be included. Additionally, the Kaplan–Meier curves evidence that non-proportionality of hazard occurs with (possibly strong) between-county survival difference including crossing of survival curves over time. Proportionality of the hazard functions is a strong assumption for these data.

The characteristics of the above-mentioned data violate the assumptions of many classical survival models. Classical survival models often consider that all the subjects are at risk of death, non-spatio-temporal clusters, and either proportionality of the hazard or the survival odds when covariates are present.

This section has two goals. First, we model our data using the transformation class spatio-temporal cure rate model in equation (6) with non-parametric baseline given by equation (14) (full model). We carry out a Bayesian analysis with covariates using the proposed priors and obtain the usual posterior estimates, smoothed by county level maps of spatio-temporal frailties and cure rate parameters. Second, we compare the full model with the standard PH ($\alpha = 1$) and PO ($\alpha = 0$) cure rate models with iid frailties ($\xi = 0$) and spatio-temporal frailties.

The Gibbs sampler with Metropolis–Hastings steps described in Section 3.2 was implemented. For some parameters, transformations were necessary and the hyperparameters for the prior distributions were chosen to obtain flat distributions. We coded the algorithm in R.²⁶

The modeling constants E , $\{(r_e, s_e)\}_{e=1}^E$, $\bar{J}_{0f}(t)$, a_0 , and b_0 in equation (14) were chosen as follows. Based on Gelfand and Mallick,²¹ we fixed $E = 5$ since keeping E relatively small produces inferences

Table 1. Summary of the melanoma cancer data.

Survival time		Status		Year of diagnosis		
(months)		(count)		year:(count)		
Median	34	Censored	(8857)	00: (955)	01: (1159)	02:(1305) 03:(1170)
IQR	46	Death	(1480)	04: (1387)	05: (1393)	06:(1436) 07:(1532)

IQR: interquartile range.

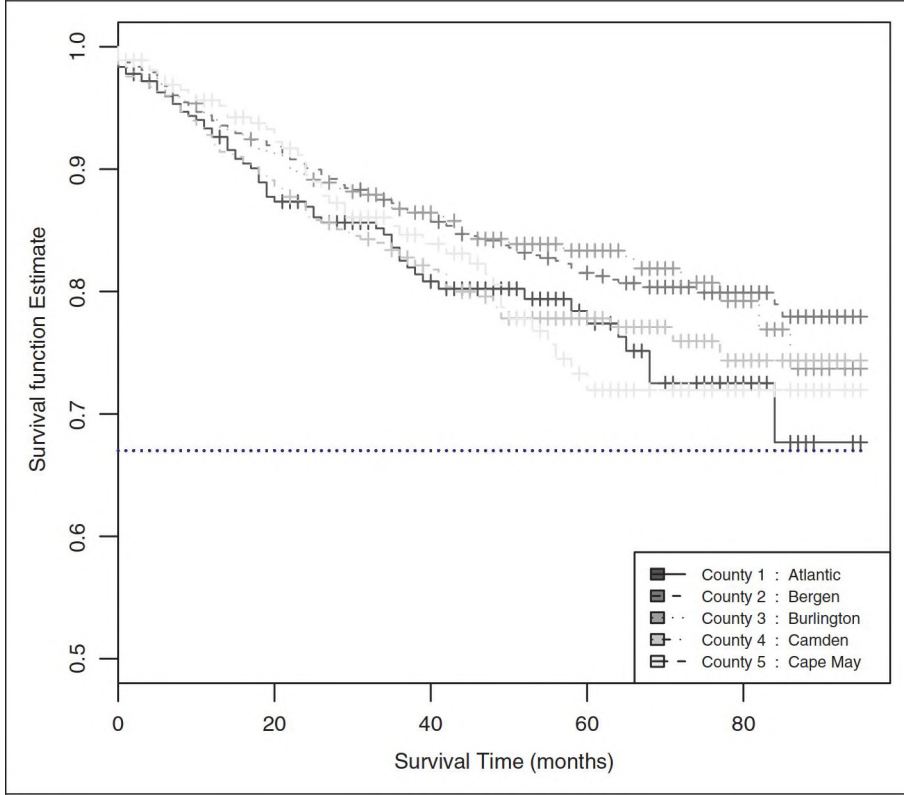


Figure 1. Marginal Kaplan–Meier plot for some counties in the state of New Jersey.

that are almost indistinguishable from large $E. \{(r_e, s_e)\}_{e=1}^E$ where chosen to produce evenly spaced Beta CDFs by setting $\{r_e\} = (1, 2, 3, 4, 5)$ and $\{s_e\} = (5, 4, 3, 2, 1)$. $\tilde{F}_{0j}(t)$, a_0 , and b_0 were selected such that $\tilde{F}_{0j}(t)$ substantially maps into $[0, 1]$. $\tilde{F}_{0j}(t)$ is a central distribution around which F_{0j} is distributed. The exponential distribution provides a flexible option in controlling the vagueness of the priors, so we set $\tilde{F}_{0j}(t) = 1 - \exp(-\lambda_0 t)$; V_j , a CDF from an exponential distribution with mean $\frac{1}{\lambda_0}$, where λ_0 was fixed at 0.025 (corresponding to a large variance) which keeps our prior confidence about $F_{0j}(t)$ vague. We found that choosing $a_0 = \frac{100}{\lambda_0}$ and $b_0 = 400$ led to values of $\tilde{J}_0(t) = \frac{a_0 \tilde{F}_{0j}(t)}{a_0 \tilde{F}_{0j}(t) + b_0}$ that substantially cover the interval $[0, 1]$.

Our algorithm was ran three initially overdispersed parallel MCMC chains for 20,000 iterations each and 25% of the runs were used as burn in period. Convergence was assessed using autocorrelation plots, density plots, trace plots, acceptance rates, Heidelberger–Welch stationarity tests, and Gelman and Rubin’s convergence diagnostic.^{27,28} The acceptance rate oscillates between 0.25 and 0.64. All the plots are satisfactory with the exception of some autocorrelation plots.

Table 2 displays the posterior medians and 95% highest posterior density (HPD) intervals for the transformation parameter α , the γ parameters involve in the cure rate specification ($\psi[\theta(\gamma, Z_{ijk})]$) and the hyperparameters used in the modeling of the spatio-temporal frailties w_{jk} . The estimates for the interaction between radiation and surgery are not displayed since they are not statistically significant; so, the model was refitted without this interaction. The posterior median for α is

Table 2. Transformation, cure and spatio-temporal parameter estimates and 95% HPD intervals.

Parameter	Median	HPD Intervals
Full Model		
$\theta(\gamma, Z_{ijk}) = \exp[\mathbf{z}_{ijk}^t \gamma]$, $W_{jk} \sim CAR(\xi, \sigma_k^2)$, α random		
Transformation parameter		
α	0.1565	(0.149, 0.198)
Cure rates parameters		
γ_{WHITE}	3.3187	(3.152, 3.474)
γ_{FEMALE}	-0.2856	(-0.398, -0.179)
γ_{AGE}	0.8685	(0.803, 0.933)
$\gamma_{\text{LOCALIZED TUMOR}}$	-0.7387	(-2.361, -0.421)
$\gamma_{\text{RADIATION}}$	-0.6022	(-1.025, 0.173)
$\gamma_{\text{RAD*LOC TUMOR}}$	-0.0898	(-1.059, 0.614)
$\gamma_{\text{NO RAD*LOC TUMOR}}$	-0.8501	(-1.236, -0.548)
$\gamma_{\text{RAD*REG/DIST TUMOR}}$	-0.6402	(-0.851, -0.273)
γ_{SURGERY}	-0.7765	(-0.908, -0.641)
γ_{MARRIED}	0.0532	(-0.049, 0.186)
γ_{INCOME}	-0.1058	(-0.162, -0.051)
Spatio-temporal hyperparameters		
ξ	0.7182	(0.591, 0.815)
σ_2^2	10.78	(6.227, 21.566)
$\sigma_{2,1}^2$	10.74	(6.179, 21.319)
$\sigma_{2,2}^2$	10.73	(6.201, 21.415)
$\sigma_{2,3}^2$	10.82	(6.230, 21.504)
$\sigma_{2,4}^2$	10.77	(6.189, 21.221)
$\sigma_{2,5}^2$	10.76	(6.142, 21.308)
$\sigma_{2,6}^2$	10.82	(6.187, 21.231)
σ_8^2	10.79	(6.163, 21.108)

HPD: highest posterior density.

0.1565 with a 95% HPD of (0.149, 0.198) which indicates that the fitted model is neither a PO or PH model.

According with the HPD intervals for the γ parameters, all the variables involved in the cure rate specification are significant with the exception of the variables marital status and radiation; however, the interaction effect of radiation and stage of the tumor (local versus regional or distant) is statistically significant. The posterior median for the coefficient γ_{AGE} is 0.8685 which implies that the cure rate decreases with age at diagnosis. For example, the odds of cure given one is diagnosed at age 25 over at age 45 are 1.37, assuming all other variables have zero effect. Similarly, any other overall cure odds for different age points can be computed using the formula $\psi[\theta(\gamma, Z_{ijk})]$. The cure rate increases as the median family income on the county of residence (in thousands) increases. Note that all the variables were standardized.

The posterior median for the variables gender and surgery are negative, indicating a higher cure rate effect for the female group and the patients that have a surgical procedure, respectively, while all the other variables are constant. The posterior median for the variable race is positive, indicating a higher cure rate effect for non-white group, while all the other variables are constant.

Table 3. Estimates of the interaction parameters and cure rates stratified by stage at diagnosis and radiation therapy.

Stage	Radiation	Overall γ effect		Cure rates		
		Estimate	95% HPD intervals	Mean	Median	IQR
Localized tumor	Yes	-1.1297	(-1.860, -0.038)	0.691	1.000	(0.137, 1.000)
	Not	-0.9617	(-2.064, -0.526)	0.415	0.164	(0.000, 1.000)
Regional/distant tumor	Yes	-0.4211	(-2.249, -0.049)	0.356	0.008	(0.000, 1.000)
	Not	-	-	0.014	0.000	(0.000, 0.000)

HPD: highest posterior density and IQR: interquartile range.

The interaction between radiation and stage is significant. Table 3 presents the estimates of the overall interaction effects and summary statistics for the estimated cure rates stratified by stage at diagnosis and radiation therapy. The sign of the estimates is negative, indicating higher cure rates with respect to the baseline group (patients diagnosed with regional or distant tumor with no radiation treatment), while all the other variables are constant. The hyperparameters for the spatio-temporal components are all significant according to Table 2. Especially, ξ different from zero shows a spatial effect on the survival curve. The HPD intervals for the variables σ_k^2 overlap which implies that the temporal effect may be not strong.

Figure 2 shows a combination of a box plot and a kernel density plot for the estimated cure rates, $\hat{\psi}[\theta(\hat{\gamma}, Z_{ijk})]$. The plot starts with a box plot. It then adds a rotated kernel density plot to each side of the box plot. The plots show the estimated cure rates classified by gender, race, surgery versus no surgery, and interaction between stage at diagnosis and radiation, respectively. All plots indicate low rates with high density around zero for the categories male, white, and no surgery, respectively. Patients in the group regional–distant stage tumor and no radiation have lower cure rate than those in any other classification of the radiation and stage interaction. The kernel is bimodal for the non-white category, indicating low cure rates for some non-white subjects and a higher proportion of non-white subjects with higher cure rates.

Figures 3 and 4 map the posterior medians of the frailties w_{jk} for the years 2000 and 2007. The pattern of the frailties varies from year to year. The counties cluster into three groups: low, median, and high frailties. The small frailties (darker colors) indicate higher cure rates. For the year 2000, five of the counties in the southwest of the state have the lowest median frailties (darker colors), while for the year 2007, counties in the southeast have the lowest frailties (darker colors).

The scale of maps 3 and 4 are different. The scales suggest an overall decreasing pattern in the frailties over time. The pattern of the frailties by year is visualized better in Figure 5. It shows the box plots of the posterior medians of the w_{jk} frailties over time. Figure 5 suggests a constant trend for the first three years, followed by a decreasing trend for the last three years. The overall total decrease in median frailties from year 2000 to 2007 is 0.60 units over the observation period. We can also observe that the variability of the posterior medians changes over time.

We now compare several other models that belong to our family of proposed models using two criteria: LPML and DIC. Table 4 lists the DIC and LPML for models with α fixed at two levels, and several frailty structures. Note that the frailties are assumed to be iid if $\xi=0$, a model with CAR(ξ, σ^2) structure is a model with spatial association, while a model with CAR(ξ, σ_k^2) is a model with spatio-temporal association. The cure rate PH CAR(ξ, σ^2) model has the larger LPML statistics suggesting that the best model is the one with PH structure and spatial frailties. However, the best model according to the DIC criteria is the cure rate transformation model with α random

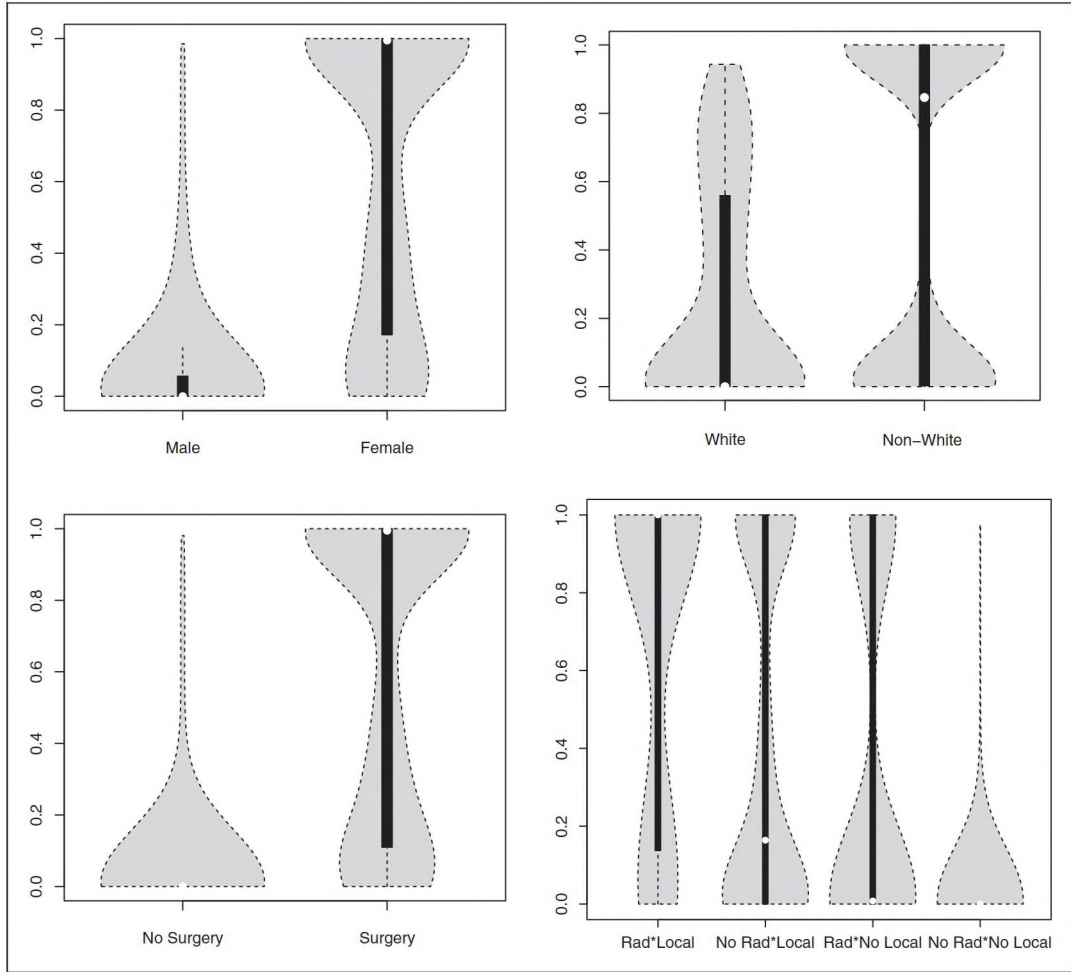


Figure 2. Box-kernel density plot for $(\psi[\theta(\gamma, Z_{ijk})])$.

and $CAR(\xi, \sigma_k^2)$ structure for the frailties suggesting a model with spatio-temporal frailties to be the best. For any transformation structure, cure rate models with spatial frailties and spatio-temporal frailties outperformed models with iid frailties.

The medians for the parameters involved in the cure rate specification (γ) and the hyperparameters that control the spatio-temporal frailties (ξ, σ_k^2) for models with PH ($\alpha = 0$) and PO ($\alpha = 1$) structure are listed in Table 5 along with their 95% HPD intervals. The conclusions are consistent with the ones reached from Table 2. All the variables involved in the cure rate specification are significant with the exception of marital status and education level. Additionally, the hyperparameters for the spatio-temporal components are all significant, in particular ξ is different from zero (spatial frailty effect).

We have proposed a hierarchical Bayesian methodology to model spatially arranged survival data with possibility of cure and temporal effects. We propose a flexible transformation class of survival curves by assuming a continuous transformation of the population survival function indexed by a

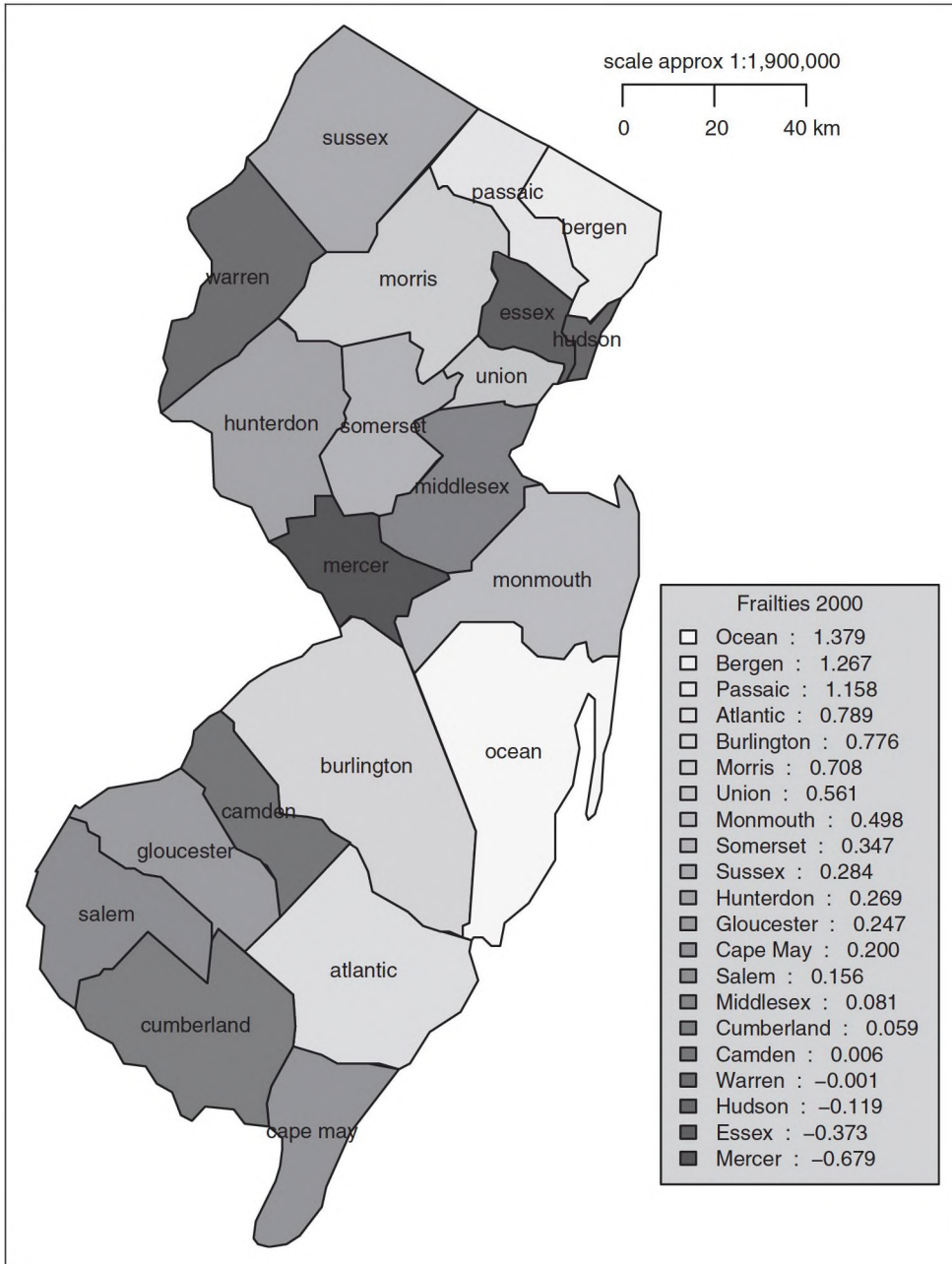


Figure 3. Fitted spatio-temporal frailties, New Jersey counties for the year 2000.

single parameter assumed to lie in $[0, +\infty)$. Our transformation model is a larger class of models containing two special cases of the well-known existing models: PH and PO models.

The survival curve is modeled as a function of a baseline curve, cure rates, and spatio-temporal frailties. Further, our model does not assume any particular shape for the baseline survival function,

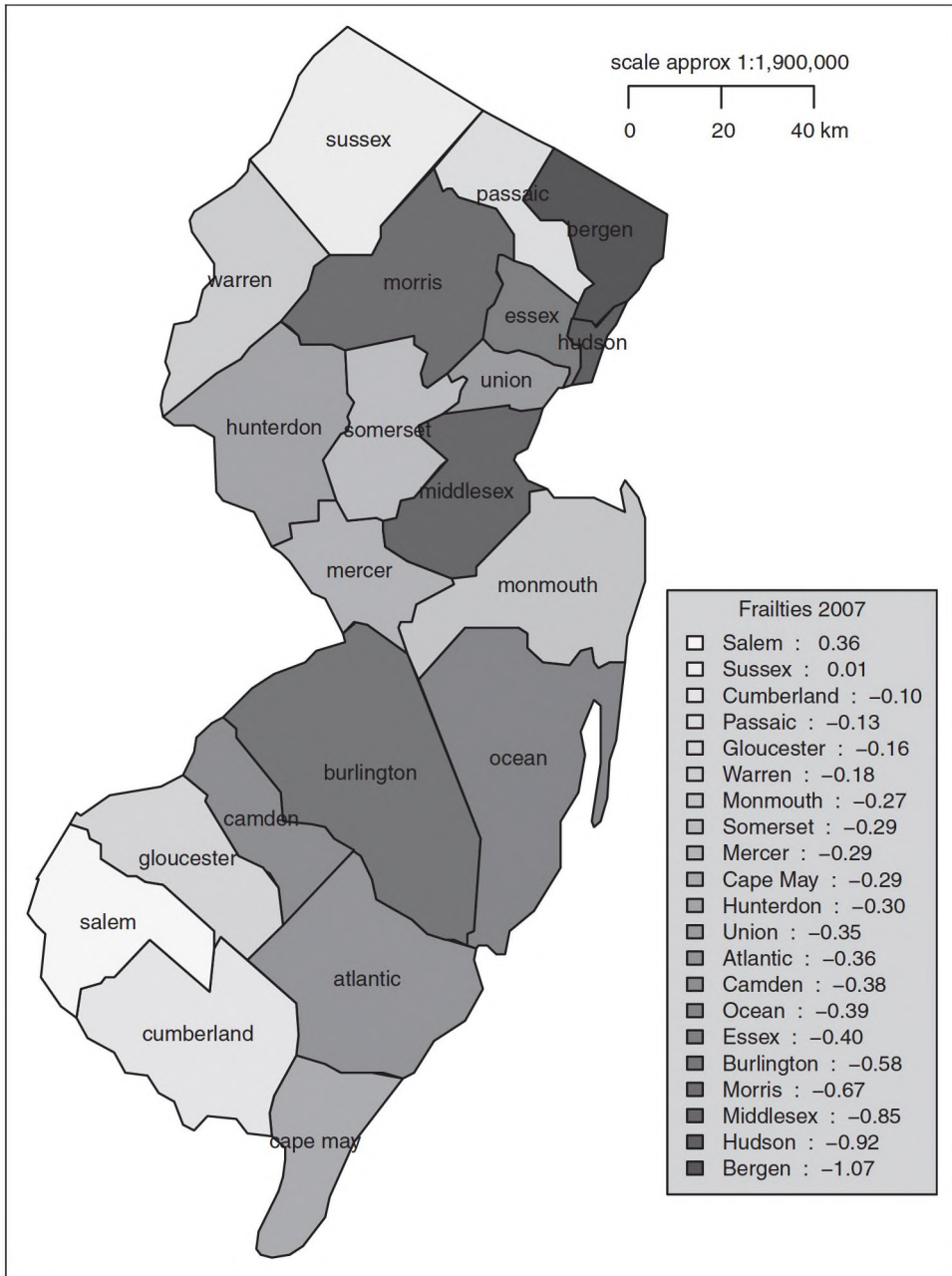


Figure 4. Fitted spatio-temporal frailties, New Jersey counties for the year 2007.

instead a non-parametric baseline function is incorporated. We integrate cure rates and spatio-temporal frailties into the survival curve specification. The cure rates are modeled through a covariate specification. The spatial frailties are specified using a CAR model with time-varying parameters resulting in a spatio-temporal formulation. The likelihood function is formulated

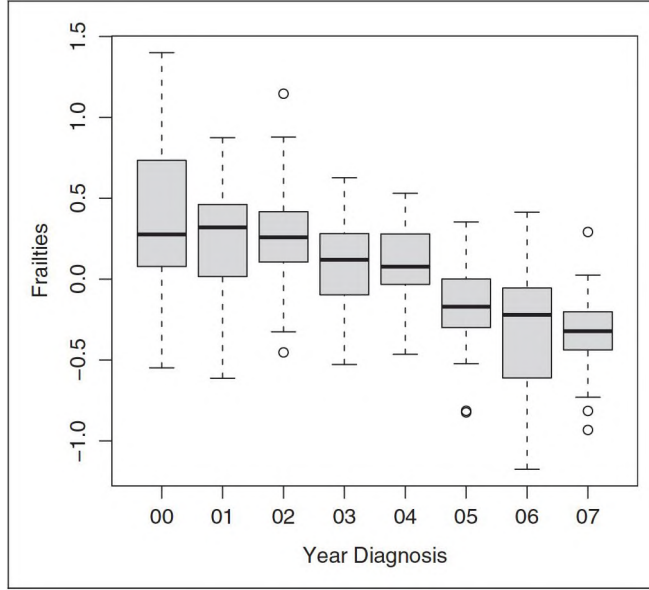


Figure 5. Box plot of the posterior medians for the spatial frailties W_{jk} for each year k .

Table 4. LPML and DIC for various cure rate competing models.

Model Number	Transformation α	Frailties W_{jk}	LPML	DIC
1	Random	iid	-12,760.01	32,797.78
2		$CAR(\xi, \sigma^2)$	-12,743.62	32,891.27
3		$CAR(\xi, \sigma_k^2)$	-12,711.46	32,526.63
4	PH ($\alpha = 0$)	iid	-12,755.22	32,949.74
5		$CAR(\xi, \sigma^2)$	-12,698.31	32,725.81
6	PO ($\alpha = 1$)	$CAR(\xi, \sigma_k^2)$	-12,734.22	32,850.00
7		iid	-12,854.29	33,189.76
8		$CAR(\xi, \sigma^2)$	-12,852.93	33,192.48
9		$CAR(\xi, \sigma_k^2)$	-12,883.72	33,379.13

LPML: logarithms of the pseudomarginal-likelihood; DIC: deviance information criterion; PH: proportional hazard; and PO: proportional odds.

assuming that the single parameter controlling the transformation is unknown (it needs to be estimated) and full conditional distributions are derived based on non-informative priors. Our melanoma survival example suggests that cure rate models with spatial or spatio-temporal frailties outperform cure rates models with iid frailties.

5 Discussion

We have formulated a class of models for spatially arranged survival data by simultaneously incorporating a cure fraction and adjusting for spatio-temporal correlation. The model formulation does not assume any proportionality of the survival curve, instead a single parameter

Table 5. Parameter estimates and 95% HPD intervals for PH and PO models with $\theta(\gamma, Z_{ijk}) = \exp[\gamma]$ and $W_{jk} \sim \text{CAR}(\xi, \sigma_k^2)$.

Parameter	PH model ($\alpha = 0$)		PO model ($\alpha = 1$)	
	Median	HPD intervals	Median	HPD intervals
Cure rates parameters				
γ_{WHITE}	3.2824	(3.115, 3.430)	3.6063	(3.378, 3.777)
γ_{FEMALE}	-0.2810	(-0.399, -0.169)	-0.2981	(-0.435, -0.159)
γ_{AGE}	0.8388	(0.779, 0.903)	0.9545	(0.880, 1.029)
$\gamma_{\text{LOCALIZED TUMOR}}$	-0.3383	(-0.832, -0.225)	-1.0287	(-1.829, -0.285)
$\gamma_{\text{RADIATION}}$	-0.2142	(-0.722, -0.128)	-0.0794	(-0.534, 0.976)
$\gamma_{\text{RAD*LOC TUMOR}}$	-0.4507	(-0.653, 1.059)	-0.0826	(-0.198, 1.219)
$\gamma_{\text{NO RAD*LOC TUMOR}}$	-0.9540	(-1.506, -0.488)	-0.4105	(-1.120, -0.007)
$\gamma_{\text{RAD*REG/DIST TUMOR}}$	-0.8982	(-1.399, 0.399)	-0.5588	(-0.238, 0.239)
γ_{SURGERY}	-0.7256	(-0.846, -0.604)	-1.0121	(-1.180, -0.859)
γ_{MARRIED}	0.0412	(-0.059, 0.172)	0.0519	(-0.077, 0.199)
γ_{INCOME}	-0.1000	(-0.149, -0.050)	-0.1343	(-0.194, -0.072)
Spatio-temporal hyperparameters				
ξ	0.7191	(0.581, 0.810)	0.7195	(0.598, 0.807)
σ_1^2	10.81	(6.156, 21.551)	10.73	(6.138, 21.233)
σ_2^2	10.75	(6.154, 21.415)	10.72	(6.161, 21.584)
σ_3^2	10.77	(6.165, 21.102)	10.79	(6.148, 21.308)
σ_4^2	10.81	(6.167, 21.116)	10.75	(6.184, 21.459)
σ_5^2	10.80	(6.192, 21.156)	10.74	(6.167, 21.337)
σ_6^2	10.76	(6.137, 21.395)	10.76	(6.145, 21.044)
σ_7^2	10.74	(6.121, 21.241)	10.70	(6.183, 21.204)
σ_8^2	10.78	(6.209, 21.149)	10.78	(6.131, 20.949)

based transformation on $S_p(t|\cdot)$ was proposed. The benchmark PH and PO models are special cases of our transformation model. The model does not assume any particular shape for the baseline survival function, instead a non-parametric baseline function is incorporated. The cure rates are modeled through a covariate specification which allows for a biomedical interpretation and is suitable for Bayesian estimation. The spatial frailties are specified using a CAR model with time-varying parameters resulting in a spatio-temporal formulation. The likelihood function is formulated assuming that the single parameter controlling the transformation is unknown (it needs to be estimated) and full conditional distributions are derived based on quickly informative priors.

We also considered several values of $\alpha \in [0, 1]$ to compare our model with the two benchmark models, PH and PO. Different values of α led to a various modeling structures. The interpretation of γ and $\{w_{jk}\}$ is the same for different transformations, but the effect on the cure rates, $\psi[\theta(\gamma, Z_{ijk})]$, and spatio-temporal relative risks, $\psi[\theta(\gamma, Z_{ijk})]$, is different, causing difficulties when comparison with PH or PO models is desired.

Our method is implemented via MCMC and it is naturally suitable for model comparison using the LPML and DIC criteria. An MCMC procedure was proposed for Bayesian estimation of the single parameter that controls the transformation, cure rates, and spatio-temporal frailties parameters. Another MCMC procedure was proposed when the parameter that control the transformation is assumed to be known. The computational efforts in terms of time and hardware are moderate,

especially for large samples like the one used here. We found some problems with some of the autocorrelation plots by, in general, the convergence diagnostics were satisfactory. Our results for the melanoma cancer data set support that a CAR prior and cure rates parameters can be useful in the modeling of the survival curves.

The model formulation given by equation (6) supports many other specifications for $\theta(\gamma, Z_{ijk})$ that lead to identifiable models as long as $F(t)$ is parametrically specified (Theorems 2.1 and 3.1). Model (6) also supports other areal and geostatistical model specifications for the spatio-temporal frailties and it can be easily generalized to incorporate other random effects in addition to the random intercepts. Although our application does not include interaction effects in the cure rate specification, the model is suitable to include any kind of variables in the cure rate link function, $\theta(\gamma, Z_{ijk})$.

Funding

Sandra M Hurtado Rúa acknowledges partial support from the Clinical Translational Science Center grant (UL1-RR024996).

Acknowledgements

We acknowledge the associate editor and reviewers for their constructive suggestions which improved the presentation of the manuscript.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Cox DR. Regression models and life tables (with discussion). *J R Stat Soc* 1972; **34**: 187–220.
2. Sy JP and Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics* 2000; **56**: 227–236.
3. Bennett S. Analysis of survival data by the proportional odds model. *Stat Med* 1983; **2**: 835–845.
4. Cortese G, Scheike T and Martinussen T. Flexible survival regression modelling. *Stat Meth Med Res* 2010; **19**: 5–28.
5. Cheng SC, Wei LJ and Ying Z. Analysis of transformation models with censored data. *Biometrika* 1995; **82**: 835–845.
6. Berkson J and Gage RP. Survival curve for cancer patients following treatment. *J Am Stat Assoc* 1952; **47**: 501–515.
7. Rodrigues J, Balakrishnan N, Cordeiro G, et al. A unified view on lifetime distributions arising from selection mechanisms. *Comput Stat Data An* 2011; **55**: 3311–3319.
8. Yu B, Tiwari R and Feuer E. Estimating the personal cure rate of cancer patients using population-based grouped cancer survival data. *Stat Meth Med Res* 2010; **20**: 261–274.
9. Tsoodikov A, Ibrahim J and Yakovlev A. Estimating cure rates from survival data: An alternative to two-component mixture models. *J Am Stat Assoc* 2003; **98**: 1063–1078.
10. Yin G. Bayesian transformation cure frailty models with multivariate failure time data. *Stat Med* 2008; **27**: 5929–5940.
11. Cooner F, Banerjee S, Carlin BP, et al. Flexible cure rate modelling under latent activation schemes. *J Am Stat Assoc* 2007; **102**: 560–572.
12. Chen MH, Ibrahim JG and Sinha D. A new bayesian model for survival data with a surviving fraction. *J Am Stat Assoc* 1999; **94**: 909–919.
13. Clayton DG. A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 1978; **65**: 141–151.
14. Turechek WW and Madden LV. A generalized linear modeling approach for characterizing disease incidence in spatial hierarchy. *Phytopathology* 2002; **93**: 458–466.
15. Biggeri A, Marchi M, Lagazio C, et al. Nonparametric maximum likelihood estimators for disease mapping. *Stat Med* 2000; **19**: 2539–2554.
16. Waller LA, Carlin BP, Xia H, et al. Hierarchical spatio-temporal mapping of disease rates. *J Am Stat Assoc* 1997; **92**: 607–617.
17. Banerjee S and Dey D. Semiparametric proportional odds model for spatially correlated survival data. *Lifetime Data Anal* 2005; **11**: 175–191.
18. Gelfand AE and Vounatsou P. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* 2003; **4**: 11–25.
19. Cressie NAC. *Statistics for spatial data*. New York: Wiley, 1993.

20. Besag J. Spatial interaction and the statistical analysis of lattice systems (with discussion). *J R Stat Soc Ser B* 1974; **36**: 192–236.
21. Gelfand AE and Mallick BK. Bayesian analysis of proportional hazard functions built from monotone functions. *Biometrics* 1995; **51**: 843–852.
22. Diaconis P and Ylvisaker D. Qualifying prior opinion. In: Bernard JM, DeGroot MH, Andley DVI, et al. (eds) *Bayesian statistics 2*. North Holland: Amsterdam, 1985, pp.133–156.
23. Gelfand AE, Dey DK and Chang H. Model determination using predictive distributions with implementation via sampling-based methods. *Bayesian Stat* 1992; **4**: 147–167.
24. Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B* 2002; **64**(4): 583–639.
25. Surveillance, Epidemiology, and End Results (SEER) Program, (www.seer.cancer.gov) (2010) SEER*Stat Database: Incidence - SEER 17 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2009 Sub (1973-2007 varying) - Linked To County Attributes - Total U.S., 1969-2007 Counties. *National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch*.
26. R Development Core Team. (2010) *R: A language and environment for statistical computing*. Vienna, Austria, R Foundation for Statistical Computing. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
27. Gelman A and Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci* 1992; **7**: 457–511.
28. Brooks SP and Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 1997; **7**: 434–455.

Appendix I

Proof of Theorem 2.1

Proof of (a) Without loss of generality assume that z is one dimensional and $\mathcal{Z} = (z_0, z_1)$; $-\infty < z_0 < z_1 < \infty$. In each case, we need to show that $S_p(t|z) \equiv S_p^*(t|z)$ if and only if $F(t) = F^*(t)$ for almost all $t < C$ and $\theta(z) = \theta^*(z)$ for all $z \in \mathcal{Z}$. The ‘if’ part is clearly true, so we prove here the ‘only if’ part.

Assume that α is known and $F(t) \equiv F(t|\gamma)$, where γ is a parameter or vector of parameters. To show the ‘only if’ part, assume $S_p(t|z) \equiv S_p^*(t|z)$. For both cases, $\alpha = 0$ or $\alpha > 0$, $S_p(t|z) \equiv S_p^*(t|z)$ implies $\frac{\theta(z)}{\theta^*(z)} = \frac{F^*(t|\gamma^*)}{F(t|\gamma)} = a$, where a is a positive constant (functionally independent of z and t) for all $t < C$ and $z \in \mathcal{Z}$. If $F^*(t|\gamma^*) = aF(t|\gamma)$, then $F^*(t|\gamma^*)$ does not belong to the same parametric family to which $F(t|\gamma)$ belongs if $a \neq 1$, unless $\gamma^* = \gamma$. Then, $a = 1$ and so $F(t) = F^*(t)$ for all $t < C$ and $\theta^*(z) = \theta(z)$ for all $z \in \mathcal{Z}$. Then, the model is identifiable regardless of whether $\theta(z)$ is unspecified or specified.

Proof of (b) Assume that α is fixed and $\delta = 0$. The ‘if’ part is clearly true, so we prove here the ‘only if’ part.

For both cases, $\alpha = 0$ or $\alpha > 0$, $S_p(t|z, w) \equiv S_p^*(t|z, w)$ implies $\frac{\theta(z)}{\theta^*(z)} = \frac{F^*(t)}{F(t)} = a$, where a is a positive constant (functionally independent of z and t) for all $t < C$ and $z \in \mathcal{Z}$. It follows that $\theta(z) = a\theta^*(z)$ for all $z \in \mathcal{Z}$, which implies that $\theta(z) = a\theta^*(z) = \exp(\beta^*z + \log(a))$ for all $z \in \mathcal{Z}$. It follows that $\theta(z)$ cannot be written as $\exp(\beta z)$ unless $a = 1$. Then, the model is identifiable when $\delta = 0$. Now suppose $\delta \neq 0$ and $\exists a \neq 1$ such that $\theta(z) = a\theta^*(z)$ for all $z \in \mathcal{Z}$ and $F^*(t) = aF(t)$ for all $t < C$. Then, $S_p^*(t|z, w) \in \mathcal{H}$ and $S_p^*(t|z, w) = \psi[F^*(t)^{\frac{1}{a}} \exp(\delta + \beta z) \exp(w)] = S_p(t|z, w)$. This shows that the model is not identifiable for $\delta \neq 0$.

Proof of (c) Assume that $\alpha > 0$ is unknown and that $\delta = 0$. The ‘if’ part is clearly true, so we prove here the ‘only if’ part.

Let $S_p(t|z) \equiv S_p^*(t|z)$. It follows that $[1 + \alpha\theta(z)F(t)]^{\frac{\alpha}{a}} = [1 + \alpha^*\theta^*(z)F^*(t)]$, for all $t < C$, $z \in \mathcal{Z}$. Let $g(t, z) = \theta(z)F(t)$, $g^*(t, z) = \theta^*(z)F^*(t)$ and consider the Taylor series expansion around $(1, 0)$ of the following functions $G(t, z) = [1 + \alpha g(t, z)]^{\frac{\alpha}{a}}$ and $G^*(t, z) = [1 + \alpha^* g^*(t, z)]$. $G(t, z) = G^*(t, z)$ for all $t < C$, $z \in \mathcal{Z}$ if and only if all the terms of the Taylor series expansion of $G(t, z)$ and $G^*(t, z)$ are equal. In particular, the second term of the Taylor series expansions of $G(t, z)$ and $G^*(t, z)$ around $(1, 0)$ are equal if and only if $(1 + \alpha)^{\frac{\alpha}{a}-1} = 1$. It follows that $\alpha^* = \alpha$ which implies that $\frac{\theta(z)}{\theta^*(z)} = \frac{F^*(t)}{F(t)} = a$, where a is a positive constant (functionally independent of z and t) for all $t < C$ and $z \in \mathcal{Z}$. It follows that $\theta(z) = a\theta^*(z) = \exp(\beta^*z + \log(a))$ for all $z \in \mathcal{Z}$. $\theta(z)$ cannot be written as $\exp(\beta z)$ unless $a = 1$. Then, the model is identifiable when $\delta = 0$.

Appendix 2

Likelihood for $\alpha = 0$

Since $\frac{\psi'(x)}{\psi(x)} = -(1 + \alpha x)^{-1} \mathbb{I}\{\alpha > 0\} - 1 \mathbb{I}\{\alpha = 0\}$, the contribution ijk to the conditional likelihood reduces to equation (16) when $\alpha > 0$ and to equation (17) when $\alpha = 0$

$$L_{ijk}(\gamma, \boldsymbol{\rho} | \mathbf{W}, \mathfrak{D}) = \left\{ 1 + \alpha F_{0j}(t | \boldsymbol{\rho}_j) \exp(\mathbf{z}'_{ijk} \boldsymbol{\gamma} + w_{jk}) \right\}^{-\frac{1}{\alpha}} \times \left\{ \frac{f_{0j}(t | \boldsymbol{\rho}_j) \exp(\mathbf{z}'_{ijk} \boldsymbol{\gamma} + w_{jk})}{1 + \alpha F_{0j}(t | \boldsymbol{\rho}_j) \exp(\mathbf{z}'_{ijk} \boldsymbol{\gamma} + w_{jk})} \right\}^{\delta_{ijk}} \quad (16)$$

$$L_{ijk}(\gamma, \boldsymbol{\rho} | \mathbf{W}, \mathfrak{D}) = \exp \left\{ -F_{0j}(t | \boldsymbol{\rho}_j) \exp(\mathbf{z}'_{ijk} \boldsymbol{\gamma} + w_{jk}) \right\} \times \left\{ f_{0j}(t | \boldsymbol{\rho}_j) \exp(\mathbf{z}'_{ijk} \boldsymbol{\gamma} + w_{jk}) \right\}^{\delta_{ijk}} \quad (17)$$

Appendix 3

Proof of Theorem 3.1

We prove the theorem for $\alpha > 0$. A similar proof holds for $\alpha = 0$. First, we proof that $\exists R(\alpha) > 0$ such that

$$L_{ijk}(\alpha, \gamma, F_{0j} | \mathbf{W}, \mathfrak{D}) \leq R(\alpha); \forall t, z \quad (18)$$

When $\delta_{ijk} = 0$, $L_{ijk}(\alpha, \gamma, F_{0j} | \mathbf{W}, \mathfrak{D}) = [1 + \alpha F_{0j}(t) \exp(\mathbf{z}'_{ijk} \boldsymbol{\gamma} + w_{jk})]^{-\frac{1}{\alpha}}$ which one can show is less than or equal to 1 using basic calculus, so equation (18) holds. For $\delta_{ijk} = 1$, $L_{ijk}(\alpha, \gamma, F_{0j} | \mathbf{W}, \mathfrak{D})$ can be written as $F_{0j}(t) \exp(\mathbf{z}'_{ijk} \boldsymbol{\gamma} + w_{jk}) [1 + \alpha F_{0j}(t) \exp(\mathbf{z}'_{ijk} \boldsymbol{\gamma} + w_{jk})]^{-\frac{1}{\alpha} - 1} \frac{f_{0j}(t)}{F_{0j}(t)}$. Let $h(x) = x(1 + \alpha x)^{-(1 + 1/\alpha)}$, $x > 0$. It can be shown that $h(x) < (1 + \alpha)^{-(1 + 1/\alpha)} < (1 + \alpha)^{-1} < (1 + \alpha)$. Since $\frac{f_{0j}(t)}{F_{0j}(t)} < B$ for some $B > 0$, equation (18) holds when $\delta_{ijk} = 1$ and $R(\alpha) = B(1 + \alpha)$.

Since Z^* is full rank, there exists L independent row vectors $z'_{ijk_1}, \dots, z'_{ijk_L}$ such that $\delta_{ijk_l} = 1$, $\forall l = 1, \dots, L$. Using equation (18) in the posterior $\pi(\alpha, \gamma, \mathbf{F}_0 | \mathfrak{D})$ (equation (13)), we get

$$\begin{aligned} \pi(\alpha, \gamma, \mathbf{F}_0 | \mathfrak{D}) &\leq \int_0^1 \int_{(0,1]^J} \int_{\mathfrak{R}^K} \int_{\mathfrak{R}^L} R(\alpha)^{d-L} \prod_{l=1}^L f_{0j}(t_{ijk_l}) \exp(\mathbf{z}'_{ijk_l} \boldsymbol{\gamma} + w_{jk_l}) \left[1 + \alpha F_{0j}(t_{ijk_l}) \exp(\mathbf{z}'_{ijk_l} \boldsymbol{\gamma} + w_{jk_l}) \right]^{-\frac{1}{\alpha} - 1} \\ &\quad \times [\mathbf{W}_k] \pi(\mathbf{F}_0) \pi(\alpha) d\gamma d\mathbf{W} d\mathbf{F}_0 d\alpha, \end{aligned} \quad (19)$$

where \mathfrak{R}^a denotes the a -dimensional Euclidean space and $(0, 1]^J$ the unit hypercube in \mathfrak{R}^J . Let $v_l = z_{ijk_l}$, $\boldsymbol{\gamma}$ be the 1-1 linear transformation from $\boldsymbol{\gamma}$ to $\mathbf{v} = (v_1, \dots, v_L)'$. Then equation (19) is proportional to

$$\begin{aligned} &\int_0^1 \int_{(0,1]^J} \int_{\mathfrak{R}^K} \int_{\mathfrak{R}^L} R(\alpha)^{d-L} \prod_{l=1}^L f_{0j}(t_{ijk_l}) \exp^{v_l} \exp^{w_{jk_l}} \left[1 + \alpha F_{0j}(t_{ijk_l}) \exp^{v_l} \exp^{w_{jk_l}} \right]^{-\frac{1}{\alpha} - 1} \\ &\quad \times [\mathbf{W}_k] \pi(\mathbf{F}_0) \pi(\alpha) d\mathbf{v} d\mathbf{W} d\mathbf{F}_0 d\alpha \end{aligned} \quad (20)$$

Integration out \mathbf{v} , equation (20) reduces to

$$\int_0^1 \int_{(0,1]^J} \int_{\mathfrak{R}^K} R(\alpha)^{d-L} \prod_{l=1}^L \left[\frac{f_{0j}(t_{ijk_l})}{F_{0j}(t_{ijk_l})} \right] [\mathbf{W}_k] \pi(\mathbf{F}_0) \pi(\alpha) d\mathbf{W} d\mathbf{F}_0 d\alpha \quad (21)$$

By assumption (d) of Theorem 3.1, $\frac{f_{0j}(t_{ijk_l})}{F_{0j}(t_{ijk_l})} < B$ for some $B > 0$, and so equation (21) is less than or equal that

$$\int_0^1 \int_{(0,1]^J} \int_{\mathfrak{R}^K} R(\alpha)^{d-L} B^L [\mathbf{W}] \pi(\mathbf{F}_0) \pi(\alpha) d\mathbf{W} d\mathbf{F}_0 d\alpha \leq \infty \quad (22)$$