Lifetime Data Analysis manuscript No.
(will be inserted by the editor)

Generalized Accelerated Failure Time Spatial Frailty Model for Arbitrarily Censored Data

Haiming Zhou · Timothy Hanson · Jiajia Zhang

the date of receipt and acceptance should be inserted later

Note: The final publication is available at Springer via http://dx.doi.org/10.1007/s10985-016-9361-4

H. Zhou
Division of Statistics, Northern Illinois University, DeKalb, IL 60115, USA
E-mail: zhouh@niu.edu

T. Hanson
Department of Statistics, University of South Carolina, Columbia, SC 29208, USA

J. Zhang
Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208, USA

Abstract Flexible incorporation of both geographical patterning and risk effects in cancer survival models is becoming increasingly important, due in part to the recent availability of large cancer registries. Most spatial survival models stochastically order survival curves from different subpopulations. However, it is common for survival curves from two subpopulations to cross in epidemiological cancer studies and thus interpretable standard survival models can not be used without some modification. Common fixes are the inclusion of time-varying regression effects in the proportional hazards model or fully nonparametric modeling, either of which destroys any easy interpretability from the fitted model. To address this issue, we develop a generalized accelerated failure time model which allows stratification on continuous or categorical covariates, as well as providing per-variable tests for whether stratification is necessary via novel approximate Bayes factors. The model is interpretable in terms of how median survival changes and is able to capture crossing survival curves in the presence of spatial correlation. A detailed Markov chain Monte Carlo algorithm is presented for posterior inference and a freely available function `frailtyGAFT` is provided to fit the model in the R package `spBayesSurv`. We apply our approach to a subset of the prostate cancer data gathered for Louisiana by the Surveillance, Epidemiology, and End Results program of the National Cancer Institute.

Keywords Interval-censored data · Heteroscedastic survival · Linear dependent tailfree process · Spatial data · Stratified AFT model

1 Introduction

Spatially correlated survival data are commonly observed in biomedical and epidemiological studies. For example, in cancer data arising from the Surveillance Epidemiology and End Results (SEER) program of the National Cancer Institute, survival times of patients from the same or adjacent counties are often expected to be more alike than those from distant counties due to region-specific similarities in environments and treatment resources. Simultaneously modeling the risk factors and geographical pattern that explain the differences in survival probabilities is of particular importance for establishing policies to improve national health systems. Choosing an appropriate survival model that takes into account the spatial dependence across counties is a must for valid statistical inferences.

The use of parametric and semiparametric hierarchical frailty survival models has become quite popular for analyzing spatially correlated survival data. First approaches augmented the linear predictor in the proportional hazards (PH) model with random effects, i.e. frailties, following some version of a Gaussian process for point-referenced data, or a Gaussian Markov random field for data observed on a lattice such as county-level data. Examples include Henderson et al. (2002), Li and Ryan (2002), Banerjee et al. (2003), Banerjee and Carlin (2003), and Hennerfeind et al. (2006). Proportional odds (PO) spatial frailty models were considered by, e.g. Banerjee and Dey (2005). Zhang and

Lawson (2011) and Wang et al. (2012) developed spatial frailty accelerated failure time (AFT) models. Zhao et al. (2009) considered all three commonly used survival models (AFT, PH and PO) with Gaussian spatial frailties, assuming the same nonparametric prior on baseline survival $S_0(\cdot)$. All of these spatial survival models assume a static, homogeneous baseline distribution $S_0(\cdot)$ coupled with a parametric part of the model (e.g. PO, AFT, PH).

Stratification is commonly used in semiparametric survival modeling when the shape of the baseline hazard changes dramatically with a discrete covariate, such as clinic identification. Stratified analyses allow a completely different baseline hazard for each strata. The partial likelihood of Cox (1975) is easily modified to accommodate stratified variables and this can be implemented, e.g. via the `STRATA` subcommand in SAS procedure `PHREG`. Unfortunately, tests concerning the stratification variable(s) cannot be carried out and continuous variables cannot be stratified on. Recently, Zhao and Hanson (2011) generalized the stratified PH model to allow spatially smooth baseline hazards over an areal map. Similarly, Hanson et al. (2012) considered longitudinally smoothed baseline hazards in a stratified PH model. In both cases, spatial or longitudinal smoothing improved predictive performance over completely independent baseline hazards or a common baseline hazard.

Semiparametric AFT models, the natural competitor to PH, have been widely studied, including Christensen and Johnson (1988), Kuo and Mallick (1997), Walker and Mallick (1999), Kottas and Gelfand (2001), Hanson and Johnson (2002, 2004) and Hanson (2006), typically using Dirichlet process mixtures or Polya trees for the homogeneous baseline distribution. However, very limited work has been completed to date on stratified AFT models. Marginal models (Chiou et al., 2015) only allow for discrete stratification and average over the strata, giving a population-averaged interpretation for the acceleration factors. Significance of semiparametric risk factors can be watered down and often the conditional interpretation is more relevant as, e.g. in the case of stratifying on gender, the gender is known and should not be averaged over. Conditionally stratified AFT has been considered, but termed "AFT with heteroscedastic error." In most of these models the error term in the log-linear model specifying the survival times is simply a single index model of the covariates times a residual, e.g. Pang et al. (2015). In such cases covariates affect the location and scale of the log-survival times, but the overall shape of the density remains static across covariates. To relax the assumption of common baseline distributional shape, Hanson and Jara (2013) discussed an AFT model with general heteroscedastic error terms. Their model can be viewed as a particular kind of censored quantile regression (Koenker, 2008), but with heteroscedastic error that changes with covariate levels. However, the finite sample performance of their model has not been well studied, and how to incorporate spatial dependence into the model remains a question.

In this paper we propose a generalized accelerated failure time spatial frailty model which allows stratification on continuous or categorical covariates, as well as per-variable tests for whether stratification is necessary via novel approximate Bayes factors. This model extends the linear dependent

tailfree process of Jara and Hanson (2011) to the interval-censored data setting, and also incorporates exchangeable or spatially varying areal-level frailties. In contrast to the aforementioned index models (Pang et al., 2015), the entire shape of the residual density changes smoothly with strata covariates. Furthermore, the residual density is median-zero, providing greatly enhanced interpretation of non-strata variables in terms of how median survival changes. The proposed model includes the traditional parametric and semiparametric AFT spatial frailty models as special cases; therefore, it provides a means of testing the fit of simpler, commonly-used models. A highly novel aspect of this work is the development of fast, approximate Bayes factors based only on a fit of the full model for testing the adequacy of many important, commonly-fit reduced models. Methods of obtaining posterior inference are carefully detailed and a freely-available function `frailtyGAFT` is provided in the R package `spBayesSurv` (`https://cran.r-project.org/web/packages/spBayesSurv`) that provides relevant output including per-variable tests for stratification covariates. All Markov chain Monte Carlo (MCMC) relies on a thoroughly-tested self-contained algorithm; no MCMC tuning is required. In addition, functions for obtaining survival curves, etc. are available. The model accommodates general interval-censored data, including standard right-censored data as well as special cases such as Case I (current status data) and Case II interval-censoring.

The rest of the paper is organized as follows. Section 2 describes the proposed model together with the MCMC implementation of posterior inference and Bayesian hypothesis tests. Section 3 presents simulation studies to evaluate the performance of the proposed model. Section 4 provides a detailed analysis of the SEER prostate cancer data. The paper is concluded by a discussion in Section 5.

2 Generalized accelerated failure time spatial frailty model

2.1 Standard survival modeling

Let t_{ij} be a random event time associated with the jth subject in the ith county and $\mathbf{z}_{ij} = (z_{1ij}, \ldots, z_{pij})'$ be a related p-dimensional vector of covariates, $j = 1, \ldots, n_i, i = 1, \ldots, m$. Let δ_{ij} be a censoring indicator equaling 1 if t_{ij} is an observed event time and equaling 0 if the event time is censored to lie in the interval $(l_{ij}, u_{ij}]$ with $l_{ij} < u_{ij}$, where $l_{ij} = 0$ ($u_{ij} = \infty$) corresponds to left (right) censoring. The event times from the same county of residence are expected to be correlated due to sharing common unobserved characteristics, such as region-specific similarities in environments and treatment resources. To incorporate spatial dependence, a traditional way is to introduce a random effect (frailty) into the linear predictor of semiparametric survival models. In this paper we consider the AFT spatial frailty model, specified as

$$S_{\mathbf{z}_{ij}}(t) = S_0\left(e^{-\mathbf{z}'_{ij}\boldsymbol{\beta}-v_i}t\right), \tag{2.1}$$

where v_i is an unobserved frailty associated with county i, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a vector of regression coefficients, and $S_0(t)$ is the baseline survival function corresponding to $\mathbf{z}_{ij} = \mathbf{0}$ and $v_i = 0$. In practice, $\mathbf{z}_{ij}$s are usually normalized so that $S_0(t)$ can serve as a reference. Often $S_0(t)$ is assumed to be a static parametric or nonparametric survival function, free of covariates. However this assumption implies that the resulting survival curves are not allowed to cross for different covariates, which can be unrealistic in practical applications (De Iorio et al., 2009). In what follows, we present a generalized AFT spatial frailty model, where $S_0(t)$ is allowed to flexibly vary with covariates under some identifiability constraints, yielding a particular kind of stratified AFT model that allows stratification on continuous or categorical covariates. In comparison to standard AFT spatial survival models, the proposed model has increased flexibility while retaining interpretability of model parameters.

2.2 Spatial frailty modeling

For modeling the spatial frailties, we consider a version of conditionally autoregressive prior (Besag, 1974). Given frailties $v_1, \ldots, v_m$ associated with counties $1, \ldots, m$, we define an $m \times m$ symmetric *proximity matrix* W with the ijth entry w_{ij} representing some type of connection between counties i and j, where w_{ii} is customarily set to 0. Typically $w_{ij} = 1$ if i and j share a common boundary and zero otherwise; this is the measure used in this paper. Note that w_{ij} could instead reflect some fashion of meaningful "distance" between counties, e.g. Mahalanobis distance of median household income or a ruralness measure between counties i and j. The frailties are assumed to follow the conditional independence assumption

$$v_i | \{v_j\}_{j \neq i}, \tau \sim N\left(\sum_{j=1}^{m} w_{ij} v_j / w_{i+},\ \tau^2 / w_{i+} \right),\ i = 1, \ldots, m, \tag{2.2}$$

where $w_{i+} = \sum_{j=1}^{m} w_{ij}$ and τ is a scale parameter. A little algebra yields the joint density of frailties as

$$p(\mathbf{v}) = p(v_1, \ldots, v_m) \propto \left(\frac{1}{\tau^2} \right)^{\frac{m-1}{2}} \exp\left\{ -\frac{1}{2\tau^2} \mathbf{v}'(D_w - W)\mathbf{v} \right\}, \tag{2.3}$$

where $\mathbf{v} = (v_1, \ldots, v_m)'$ and D_w is an $m \times m$ diagonal matrix with $(D_w)_{ii} = w_{i+}$. Note this joint density is improper since we can add any constant to all of the v_is and (2.3) is unaffected; that is, the v_is are not centered. We consider the constraint $\sum_{i=1}^{m} v_i = 0$, which provides the needed centering to avoid identifiability issues. Under this constraint, the conditionally autoregressive prior becomes

$$v_i | \{v_j\}_{j \neq i} \sim N\left(\sum_{j \neq i}^{m-1} w^*_{ij} v_j / w^*_{i+},\ \tau^2 / w^*_{i+} \right),\ i = 1, \ldots, m-1, \tag{2.4}$$

where $w^*_{i+} = w_{i+} + w_{m+} + 2w_{mi}$ and $w^*_{ij} = w_{ij} - w_{m+} - w_{mj} - w_{mi}$. We prefer the incoporation of the sum-to-zero constraint directly into the prior and model to avoid the *ad hoc* adjustments typically made within the MCMC scheme itself.

2.3 Mixture of linear dependent tailfree processes prior for $S_0(t)$

We allow the baseline survival function $S_0(t)$ to depend on certain covariates, say a q-dimensional vector $\mathbf{x}_{ij}$ which is often a subset of $\mathbf{z}_{ij}$, yielding the generalized AFT (GAFT) spatial frailty model

$$S_{\mathbf{z}_{ij}}(t) = S_{0,\mathbf{x}_{ij}}\left(e^{-\mathbf{z}'_{ij}\boldsymbol{\beta}-v_i}t\right), \tag{2.5}$$

where v_i has the CAR prior (2.3) with $\sum_{i=1}^m v_i = 0$. For ease of handling identifiability issues, we rewrite the model (2.5) as:

$$y_{ij} = \log(t_{ij}) = \tilde{\mathbf{z}}'_{ij}\tilde{\boldsymbol{\beta}} + v_i + \epsilon_{ij}, \tag{2.6}$$

where $\tilde{\mathbf{z}}_{ij} = (1, \mathbf{z}'_{ij})'$ includes an intercept, $\tilde{\boldsymbol{\beta}} = (\beta_0, \boldsymbol{\beta}')'$ is a vector of corresponding coefficients, ϵ_{ij} is a heteroscedastic error term independent of v_i, and $P(e^{\beta_0+\epsilon_{ij}} > t|\mathbf{x}_{ij}) = S_{0,\mathbf{x}_{ij}}(t)$. Here we assume

$$\epsilon_{ij}|G_{\mathbf{x}_{ij}} \overset{ind.}{\sim} G_{\mathbf{x}_{ij}},$$

where $G_{\mathbf{x}}$ is a probability measure defined on $\mathbb{R}$ for every $\mathbf{x} \in \mathcal{X}$; this defines a model for the entire collection of probability measures $\mathcal{G}_{\mathcal{X}} = \{G_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$ so that each element is allowed to smoothly change with the covariates $\mathbf{x}$.

We consider a mixture of linear dependent tailfree processes (LDTFP) prior (Jara and Hanson, 2011) for $\mathcal{G}_{\mathcal{X}}$. In particular, we focus on an LDTFP centered at a normal distribution Φ_σ with mean 0 and variance σ^2, that is, $E(G_{\mathbf{x}}) = N(0, \sigma^2)$ for every $\mathbf{x} \in \mathcal{X}$. The LDTFP successively partitions the real line into finer and finer partitions; each refinement of a partition produces the next level of sets in the process. At level l, the process partitions $\mathbb{R}$ into 2^l intervals $B_{l,k} = \left[\Phi_\sigma^{-1}((k-1)2^{-l}), \Phi_\sigma^{-1}(k2^{-l})\right]$, $k = 1, \ldots, 2^l$, with $B_{l,2^l}$ being right-open, so that an L-level set of nested partitions is defined as $\Pi^{L,\sigma} = \{B_{l,k} : k = 1, \ldots, 2^l, l = 1, \ldots, L\}$. Note that $B_{l,k} = B_{l+1,2k-1} \cup B_{l+1,2k}$. Given that an observation is in set k at level l, say $B_{l,k}$, it could then be in either $B_{l+1,2k-1}$ or $B_{l+1,2k}$ at level $l+1$ with conditional probability $Y_{l+1,2k-1}$ or $Y_{l+1,2k}$ respectively. Clearly they must sum to one for every $\mathbf{x} \in \mathcal{X}$, and so we consider logistic regression for each of these probabilities, allowing the entire shape of the density to change with covariates. Specifically, we assume $Y_{l+1,2k-1}(\mathbf{x}) = h(\tilde{\mathbf{x}}'\boldsymbol{\gamma}_{l,k})$ and $Y_{l+1,2k}(\mathbf{x}) = 1 - h(\tilde{\mathbf{x}}'\boldsymbol{\gamma}_{l,k})$, where $\tilde{\mathbf{x}} = (1, \mathbf{x}')'$ includes an intercept, $\boldsymbol{\gamma}_{l,k} = (\gamma_{l,k,0}, \ldots, \gamma_{l,k,q})'$ is a vector of coefficients, and $h(\cdot) = \frac{\exp\{\cdot\}}{1+\exp\{\cdot\}}$. Finally, there are $2^L - 1$ regression coefficient vectors $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_{l,k}\}$, e.g. for $L = 3$, $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_{0,1}, \boldsymbol{\gamma}_{1,1}, \boldsymbol{\gamma}_{1,2}, \boldsymbol{\gamma}_{2,1}, \boldsymbol{\gamma}_{2,2}, \boldsymbol{\gamma}_{2,3}, \boldsymbol{\gamma}_{2,4}\}$. Let $\mathbf{X}$ be the $n \times (q+1)$ design matrix of covariates $\tilde{\mathbf{x}}_{ij}$s, where $n = \sum_{i=1}^m n_i$. Following

Jara and Hanson (2011), each $\boldsymbol{\gamma}_{l,k}$ is assigned an independent normal g-prior (Zellner, 1983), $\boldsymbol{\gamma}_{l,k} \sim N_{q+1}\left(\mathbf{0}, \frac{2n}{\alpha\rho(l+1)}(\mathbf{X}'\mathbf{X})^{-1}\right)$, where $\rho(l) = l^2$. Furthermore, the LDTFP is specified by setting $\boldsymbol{\gamma}_{0,1} \equiv \mathbf{0}$, such that for every $\mathbf{x} \in \mathcal{X}$, $G_{\mathbf{x}}$ is almost surely a median-zero probability measure. This is important to avoid identifiability issues.

The precision parameter $\alpha \in \mathbb{R}^+$ controls how closely the random distribution $G_{\mathbf{x}}$ follows Φ_σ in terms of L_1 distance (Hanson et al., 2008). Large values of α indicate a strong belief that ϵ_{ij}s are closely iid from Φ_σ. Smaller values of α, on the other hand, allow more pronounced deviations of $G_{\mathbf{x}}$ from Φ_σ. We consider a gamma prior on α, say, $\alpha \sim \Gamma(a_0, b_0)$, as suggested in Jara and Hanson (2011); the full conditional distribution for α is also a gamma distribution. Here $\Gamma(a, b)$ refers to the gamma distribution with shape a and rate b. As for the choice of L, we typically consider $L \approx \log_2(n/n_0)$, where n_0 (usually from 5 to 10) is a "typical" number of observations falling into each set at level L (Hanson, 2006). However, this choice is conservative for the LDTFP relative to Polya trees (Jara and Hanson, 2011). Hanson (2006) observed that the LPML remains essentially unchanged or gets slightly worse after a certain level L; this is confirmed for the LDTFP in Zhou et al. (2015). In addition, the LDTFP depends on the partition $\Pi^{L,\sigma}$ which is further determined by the centering distribution Φ_σ. To mitigate this partition effect on posterior inferences, we specify a mixture of LDTFP by putting an inverse gamma prior on σ, say $\sigma^{-2} \sim \Gamma(a_\sigma, b_\sigma)$.

Based on the above LDTFP specification, given $\boldsymbol{\gamma}$, σ and $\mathbf{x}$, $G_{\mathbf{x}}$ is known. Define the function $k_\sigma(x)$ to be the index $k \in \{1, \ldots, 2^L\}$ such that x falls into set $B_{L,k}$, i.e. $k_\sigma(x) = \lceil 2^L \Phi_\sigma(x) \rceil$, where $\lceil x \rceil$ is the ceiling function, the smallest integer greater than or equal to x. Further define probability $p_{\mathbf{x}}(k)$ for $k = 1, \ldots, 2^L$ as

$$p_{\mathbf{x}}(k) = G_{\mathbf{x}}\{B_{L,k}|\boldsymbol{\gamma}, \sigma\} = \prod_{l=1}^{L} Y_{l,\lceil k2^{l-L} \rceil}(\mathbf{x}).$$

The resulting density of $\epsilon_{ij}|\mathbf{x}_{ij}$ is given by

$$g_{\mathbf{x}_{ij}}(e) = 2^L \phi_\sigma(e) p_{\mathbf{x}_{ij}}\{k_\sigma(e)\}, \tag{2.7}$$

where ϕ_σ is the density function corresponding to Φ_σ. The cumulative distribution function associated with $g_{\mathbf{x}_{ij}}(e)$ is given by

$$G_{\mathbf{x}_{ij}}(e) = p_{\mathbf{x}_{ij}}\{k_\sigma(e)\}\left\{2^L \Phi_\sigma(e) - k_\sigma(e)\right\} + \sum_{k=1}^{k_\sigma(e)} p_{\mathbf{x}_{ij}}(k). \tag{2.8}$$

As shown by Jara and Hanson (2011), the LDTFP has appealing theoretical properties such as continuity as a function of the covariates, large support on the space of conditional density functions, straightforward posterior computation relying on algorithms for fitting generalized linear models, and the

process closely matches conventional Polya tree priors (see, e.g., Hanson, 2006) at each value of the covariate, which justify its choice here.

Regarding the priors for regression coefficients $\tilde{\boldsymbol{\beta}}$ and τ^2, typically independent priors $\tilde{\boldsymbol{\beta}} \sim N_p(\mathbf{m}_0, \mathbf{S}_0)$ and $\tau^{-2} \sim \Gamma(a_\tau, b_\tau)$ are considered; these can both be chosen to be relatively vague. In summary the proposed GAFT spatial frailty model takes the following hierarchical structure:

$$\begin{aligned}
(1-\delta_{ij})|t_{ij} &= I(l_{ij} < t_{ij} \leq u_{ij}), j = 1, \ldots, n_i, i = 1, \ldots, m \\
y_{ij} = \log(t_{ij}) &= \tilde{\mathbf{z}}'_{ij}\tilde{\boldsymbol{\beta}} + v_i + \epsilon_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, m \\
\epsilon_{ij}|\boldsymbol{\gamma}, \sigma &\overset{ind.}{\sim} g_{\mathbf{x}_{ij}}(\cdot), j = 1, \ldots, n_i, i = 1, \ldots, m \\
\boldsymbol{\gamma}_{l,k}|\alpha &\overset{ind.}{\sim} N_{q+1}(\mathbf{0}, \frac{2n}{\alpha(l+1)^2}(\mathbf{X}'\mathbf{X})^{-1}), k = 1, \ldots, 2^l, l = 1, \ldots, L-1 \\
\alpha &\sim \Gamma(a_0, b_0) \\
\sigma^{-2} &\sim \Gamma(a_\sigma, b_\sigma) \\
v_i|\{v_j\}_{j\neq i}, \tau &\sim N\left(\sum_{j=1}^{m} w_{ij}v_j/w_{i+},\ \tau^2/w_{i+}\right), i = 1, \ldots, m \\
\tau^{-2} &\sim \Gamma(a_\tau, b_\tau) \\
\tilde{\boldsymbol{\beta}} &\sim N_{p+1}(\mathbf{m}_0, \mathbf{S}_0).
\end{aligned} \tag{2.9}$$

We next describe a Markov chain Monte Carlo (MCMC) scheme for obtaining posterior inference, which can be implemented using the function `frailtyGAFT` available in the R package `spBayesSurv`.

2.4 Posterior computation

In this section we describe an MCMC sampling algorithm for the proposed model, arrived at after a considerable amount of trial and error, to determine the conditions that provide reasonable mixing and speed. By now, MCMC schemes are a standard part of the Bayesian statistician's toolbox, we refer the reader to Robert and Casella (2005) for an overview. Posterior samples for the model parameters are used for all inference of interest.

Let $\Omega = (\mathbf{y}_c, \tilde{\boldsymbol{\beta}}, \mathbf{v}, \tau^2, \sigma^2, \boldsymbol{\gamma}, \alpha)$ denote collectively the model parameters to be updated, where $\mathbf{y}_c = \{y_{ij} : \delta_{ij} = 0\}$ are censored log-survival times. The $y_{ij} \in \mathbf{y}_c$, each component of $\tilde{\boldsymbol{\beta}}$, v_i and σ are all sampled using the single-variable slice sampling method (Neal, 2003). Note the median regression coefficients $\tilde{\boldsymbol{\beta}}$ may be conveniently updated using adaptive Metropolis-Hastings (Haario et al., 2001) with multivariate Gaussian proposals, but this method suffers extremely low acceptance rates when the LDTFP is far from its centering normal distribution. For the LDTFP regression parameters $\gamma_{l,k}$, we utilize Metropolis-Hastings steps with Gaussian proposals based on iterative weighted least squares (Gamerman, 1997), recognizing that the $\gamma_{l,k}$ full conditionals are

proportional to logistic regression likelihoods. The hyperparameter τ^2 and α are sampled according to their conjugate full conditional distributions. A complete description of updating steps is available in the online material.

Given a set of posterior samples $\{\boldsymbol{\Omega}^{(s)}, s = 1, \ldots, S\}$, all the inference targets can be easily estimated. For example, the baseline survival function $S_{0,\mathbf{x}}(t) = P(e^{\beta_0+\epsilon} > t|\mathbf{x})$ given the covariate $\mathbf{x}$ is estimated by

$$S_{0,\mathbf{x}}(t) = \frac{1}{S}\sum_{s=1}^{S}\left\{1 - G_{\mathbf{x}}^{(s)}\left(\log t - \beta_0^{(s)}\right)\right\}, \tag{2.10}$$

where $G_{\mathbf{x}}^{(s)}(\cdot)$ is given in (2.8) with all unknown parameters replaced by corresponding posterior values in the sth iterate.

2.5 Bayesian hypothesis testing

The proposed GAFT spatial frailty model includes the following as important special cases: an AFT spatial frailty model with nonparametric baseline where $G_{\mathbf{x}} = G_{\mathbf{x}'}$ for all $\mathbf{x} = \mathbf{x}'$ and parametric baseline model $G_{\mathbf{x}} = N(0, \sigma^2)$ for all $\mathbf{x} \in \mathcal{X}$. Hypothesis tests can be constructed based on the LDTFP coefficients $\{\boldsymbol{\gamma}_{l,k} : k = 1, \ldots, 2^l, l = 1, \ldots, L-1\}$, where $\boldsymbol{\gamma}_{l,k} = (\gamma_{l,k,0}, \ldots, \gamma_{l,k,q})'$. Let $\boldsymbol{\gamma}_{l,k,-j}$ denote the subvector of $\boldsymbol{\gamma}_{l,k}$ without element $\gamma_{l,k,j}$ for $j = 0, \ldots, q$. Set $\boldsymbol{\Upsilon}_j = (\gamma_{l,k,j}, k = 1, \ldots, 2^l, l = 1, \ldots, L-1)'$, $\boldsymbol{\Upsilon}_{-j} = (\boldsymbol{\gamma}'_{l,k,-j}, k = 1, \ldots, 2^l, l = 1, \ldots, L-1)'$ and $\boldsymbol{\Upsilon} = (\boldsymbol{\gamma}'_{l,k}, k = 1, \ldots, 2^l, l = 1, \ldots, L-1)'$. Testing the hypotheses $H_0 : \boldsymbol{\Upsilon}_{-0} = \mathbf{0}$ and $H_0 : \boldsymbol{\Upsilon} = \mathbf{0}$ leads to global comparisons of the proposed model with the above two special cases respectively. Similarly, we may also test the null hypothesis $H_0 : \boldsymbol{\Upsilon}_j = \mathbf{0}$ for the jth covariate effect of $\mathbf{x}$ on the baseline survival, $j = 1, \ldots, q$. We use Bayes factors to accomplish these hypotheses.

Suppose we wish to test $H_0 : \boldsymbol{\Upsilon}_j = \mathbf{0}$ versus $H_1 : \boldsymbol{\Upsilon}_j \neq \mathbf{0}$, for fixed $j \in \{1, \ldots, q\}$. The Bayes factor between hypotheses H_1 and H_0 is defined as

$$BF_{10} = \frac{\int \mathcal{L}(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) d(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})}{\int \mathcal{L}(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}) p_0(\boldsymbol{\psi}) d\boldsymbol{\psi}}, \tag{2.11}$$

where $\boldsymbol{\psi}$ is the remaining model parameters under the alternative, $p_0(\boldsymbol{\psi})$ and $p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})$ are the prior probability densities under H_0 and H_1 respectively, $\mathcal{D}$ is the observed data, and $\mathcal{L}(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})$ is the likelihood function. According to the Savage-Dickey density ratio expression (Dickey, 1971), if

$$p(\boldsymbol{\psi}|\boldsymbol{\Upsilon}_j = \mathbf{0}) = p_0(\boldsymbol{\psi}), \tag{2.12}$$

then BF_{10} can be written as

$$BF_{10} = \frac{p(\boldsymbol{\Upsilon}_j = \mathbf{0})}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D})}, \tag{2.13}$$

where $p(\boldsymbol{\Upsilon}_j) = \int p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) d\boldsymbol{\psi}$ and $p(\boldsymbol{\Upsilon}_j|\mathcal{D}) = \int p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}|\mathcal{D}) d\boldsymbol{\psi}$ are the marginal prior and posterior density of $\boldsymbol{\Upsilon}_j$ respectively under H_1. We show in Proposition 1 of Appendix that, when the precision parameter α is fixed, the assumption (2.12) holds and $p(\boldsymbol{\Upsilon}_j = \mathbf{0})$ is given by

$$p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha) = \prod_{l=1}^{L-1} \prod_{k=1}^{2^l} \phi\left(0 \middle| 0, \frac{2n}{\alpha\rho(l+1)} (\mathbf{X}'\mathbf{X})^{-1}_{jj}\right). \tag{2.14}$$

where $(\mathbf{X}'\mathbf{X})^{-1}_{jj}$ is the $(j+1, j+1)$th element of $(\mathbf{X}'\mathbf{X})^{-1}$, and $\phi(\cdot|\mu, \sigma^2)$ denotes the normal density with mean μ and variance σ^2.

Now consider putting a prior π and π_0 on α under H_1 and H_0 respectively, assuming the same priors on $\boldsymbol{\gamma}_{l,k}$ as Proposition 1. Then we have

$$\begin{aligned} p(\boldsymbol{\Upsilon}_{-j}|\boldsymbol{\Upsilon}_j = 0) &= \int p(\boldsymbol{\Upsilon}_{-j}|\boldsymbol{\Upsilon}_j = 0, \alpha) \frac{p(\boldsymbol{\Upsilon}_j = 0|\alpha)\pi(\alpha)}{p(\boldsymbol{\Upsilon}_j = 0)} d\alpha \\ &= \int p_0(\boldsymbol{\Upsilon}_{-j}|\alpha) \frac{p(\boldsymbol{\Upsilon}_j = 0|\alpha)\pi(\alpha)}{p(\boldsymbol{\Upsilon}_j = 0)} d\alpha \end{aligned} \tag{2.15}$$

and $p_0(\boldsymbol{\Upsilon}_{-j}) = \int p_0(\boldsymbol{\Upsilon}_{-j}|\alpha)\pi_0(\alpha) d\alpha$. To satisfy (2.12), we need equation (2.15) equal to $p_0(\boldsymbol{\Upsilon}_{-j})$, which holds when

$$\pi_0(\alpha) = \frac{p(\boldsymbol{\Upsilon}_j = 0|\alpha)\pi(\alpha)}{p(\boldsymbol{\Upsilon}_j = 0)}.$$

Let $\Gamma(\cdot|a, b)$ denote the density of $\Gamma(a, b)$ distribution. Taking $\pi(\cdot) = \Gamma(\cdot|a_0, b_0)$ yields that $\pi_0(\cdot) = \Gamma(\cdot|a_0 + 2^{L-1} - 1, b_0)$. However, π_0 puts too much probability on H_0 against H_1, so that H_0 can be hardly rejected. To avoid this undesirable situation, we take $\pi_0 = \pi$ and apply the generalized Savage-Dickey density ratio approach, proposed by Verdinelli and Wasserman (1995), which does not rely on the assumption (2.12). We show in Proposition 2 of Appendix that BF_{10} can be written as a product of two quantities and both can be estimated from posterior simulation, that is,

$$BF_{10} = \{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D})\}^{-1} \left\{ E\left[\frac{1}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)} \right] \right\}^{-1}, \tag{2.16}$$

where the expectation is with respect to $p(\alpha|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D})$.

As noted by Raftery (1996), only a crude approximation of Bayes factor is needed. Thus we estimate the marginal posterior density $p(\boldsymbol{\Upsilon}_j|\mathcal{D})$ by an $M = 2^L - 2$ dimensional multivariate normal distribution using MCMC posterior samples. That is, we assume $p(\boldsymbol{\Upsilon}_j|\mathcal{D}) \approx \phi_M(\boldsymbol{\Upsilon}_j; \hat{\mathbf{m}}_j, \hat{\mathbf{S}}_j)$, where $\phi_M(\cdot; \mathbf{m}, \mathbf{S})$ is a multivariate normal density with mean $\mathbf{m}$ and covariance matrix $\mathbf{S}$, and $\hat{\mathbf{m}}_j$ and $\hat{\mathbf{S}}_j$ are estimated by the sample mean and covariance of the posterior sample for $\boldsymbol{\Upsilon}_j$. To avoid drawing a sample from $p(\alpha|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D})$, we assume that $p(\alpha|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D}) \approx p(\alpha|\mathcal{D})$ under H_0. We then approximate

$\{E\,[1/p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)]\}^{-1}$ by $p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\hat{\alpha})$, where $\hat{\alpha}$ is the posterior mean of α. Thus, we can estimate the Bayes factor by

$$\hat{BF}_{10} = \frac{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\hat{\alpha})}{\phi_M(\boldsymbol{\Upsilon}_j = \mathbf{0}; \hat{\mathbf{m}}_j, \hat{\mathbf{S}}_j)}, \tag{2.17}$$

where $p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)$ is given in (2.14).

3 Simulation studies

We performed simulations to illustrate and assess the proposed approach using the provided `R` package `spBayesSurv`. The data were simulated from the GAFT model (2.6), where $\tilde{\mathbf{z}}_{ij} = (1, z_{1ij}, z_{2ij})'$ with $z_{1ij} \overset{iid}{\sim} N(0,1)$ and $z_{2ij} \overset{iid}{\sim}$ Bernoulli(0.5), $\tilde{\boldsymbol{\beta}} = (\beta_0, \beta_1, \beta_2)' = (-1, 1, -0.5)'$, v_i follows the CAR model (2.3) with $\tau^2 = 0.1$ and W being the Louisiana proximity matrix used in the SEER data analysis, and $i = 1, \ldots, 64$, $j = 1, \ldots, 10$. We considered three different distribution settings for the error term ϵ_{ij}:

Scenario I: $\epsilon_{ij}|x_{ij} \overset{ind.}{\sim} \begin{cases} N(0, 0.8^2) & \text{if } x_{ij} = 0 \\ 0.5N(-1, 0.5^2) + 0.5N(1, 0.5^2) & \text{if } x_{ij} = 1 \end{cases}$

Scenario II: $\epsilon_{ij} \overset{i.i.d}{\sim} 0.5N(-1, 0.5^2) + 0.5N(1, 0.5^2)$

Scenario III: $\epsilon_{ij} \overset{i.i.d}{\sim} N(0, 0.8^2)$

where $x_{ij} = z_{2ij}$. Note that the first scenario is not a particular case of the proposed model; the second and third scenarios are included to examine the behaviour of the proposed approach when a standard parametric or semiparametric AFT model (with covariate-free error term) is correct. Non-informative right-censoring times were simulated from a Uniform$(0.5, 1)$ distribution so that the censoring rate is around 25%. For each simulation scenario, a total of 500 replicates of the dataset were generated. We then fit the proposed model with both covariates included in the LDTFP modeling part, i.e. $\tilde{\mathbf{x}}_{ij} = \tilde{\mathbf{z}}_{ij}$, using the following prior settings: $L = 4$, $a_0 = 5$, $b_0 = 1$, $\mathbf{m}_0 = \mathbf{0}_3$, $\mathbf{S}_0 = 10^5\mathbf{I}_3$, $a_\sigma = b_\sigma = 2.001$, $a_\tau = b_\tau = 0.1$. For each MCMC algorithm, $10,000$ scans were thinned from $100,000$ after a burn-in period of $10,000$ iterations.

Table 1 presents the proportions of Bayes factor BF_{10} greater than 3, 10 and 30 for the hypotheses discussed in Section 2.5. The results demonstrate that the proposed Bayes factor is able to identify which covariate truly affects the shape of the error term and which one does not, with very low wrong-decision rates across all the scenarios. For example, using the cutpoint 3, the type I error rates are around 0.05 in Scenarios I and II, and are around 0.1 in Scenario III. If more conservative tests are preferred, one could use the cutpoint 10 or 30, which gives much smaller type I errors.

Table 2 summaries the inference results for regression parameters, including the averaged bias (BIAS) and posterior standard deviation (PSD) of each point estimate, the Monte Carlo standard deviation of the point estimate (SD-Est) and the coverage probability (CP) of the 95% highest posterior density

Table 1 Simulation data. The proportions (across Monte Carlo simulations) of BF_{10} greater than 3, 10 and 30.

Scenario	H_0	$BF_{10} > 3$	$BF_{10} > 10$	$BF_{10} > 30$
I	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.056	0.004	0.004
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	1	1	1
II	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.052	0.016	0.008
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	0.018	0.008	0.004
	$\boldsymbol{\Upsilon}_{-0} = \mathbf{0}$	0.022	0.004	0.002
	$\boldsymbol{\Upsilon} = \mathbf{0}$	1	1	1
III	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.096	0.016	0
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	0.080	0.010	0.002
	$\boldsymbol{\Upsilon}_{-0} = \mathbf{0}$	0.124	0.018	0
	$\boldsymbol{\Upsilon} = \mathbf{0}$	0.116	0.030	0.008

Table 2 Simulation data. True value, averaged bias (BIAS) and posterior standard deviation (PSD) of each point estimate (i.e. posterior mean), standard deviation (across Monte Carlo simulations) of the point estimate (SD-Est) and coverage probability (CP) for the 95% credible interval.

Scenario	Parameter	True	BIAS	PSD	SD-Est	CP
I	β_0	-1.0	0.011	0.069	0.061	0.966
	β_1	1.0	0.005	0.056	0.050	0.958
	β_2	-0.5	0.010	0.129	0.104	0.976
	τ^2	0.1	0.034	0.055	0.045	0.978
II	β_0	-1.0	0.019	0.126	0.105	0.966
	β_1	1.0	-0.005	0.078	0.068	0.970
	β_2	-0.5	0.005	0.171	0.150	0.966
	τ^2	0.1	0.029	0.052	0.041	0.982
III	β_0	-1.0	0.004	0.062	0.055	0.980
	β_1	1.0	0.004	0.048	0.046	0.960
	β_2	-0.5	-0.001	0.085	0.076	0.968
	τ^2	0.1	0.039	0.061	0.047	0.982

interval. The results show that the point estimates (i.e. posterior means) of $\tilde{\boldsymbol{\beta}}$ are almost unbiased under all three scenarios, while the point estimates of τ^2 are positively biased. This is not surprising, as the posterior distribution of τ^2 is grossly right-skewed, in which case the mode would be an ideal choice for the point estimate. By using the posterior mode (calculated from a kernel-smoothed density) as a point estimate of τ^2, the averaged biases reduce to 0.007, 0.006 and 0.010 under Scenarios I, II and III, respectively. The PSD values are all slightly greater than, but fairly close to the SD-Est values, suggesting that the posterior standard deviation is an appropriate estimator

of the frequentist standard error. The CP values are all close to the nominal 95% level or slightly greater. Here all covariates are included in the LDTFP baseline survival function regardless of their significance. Note that the use of Bayes factors allows us to detect which covariates need to be stratified on and how significant they are. Based on the test results, we can always remove the covariates that have $BF_{10} < 1$ from the LDTFP modeling. For this reason, in the online material we also report the parameter inferences with the LDTFP specified according to the truth, i.e., $\tilde{\mathbf{x}}_{ij} = (1, z_{2ij})'$ under Scenario I and $\tilde{\mathbf{x}}_{ij} = 1$ under Scenarios II & III. The results reveal that the CP values for $\tilde{\boldsymbol{\beta}}$ are now around the nominal 95% except the CP for β_2 in Scenario I. It seems that our approach tends to slightly overestimate the standard deviation of the covariate coefficient when that covariate highly affects the baseline function. Figure 1 presents the average, across the 500 simulated data sets, of the fitted density and survival functions of log survival times for some specific covariate values. The results reveal that the proposed model is capable to capture the crossing behaviour of survival curves very well.

In Section 3 of the online material, additional simulation results are presented for the sensitivity analysis on the prior of α and the choice of L. The parameter estimates and hypothesis tests are essentially not affected by the choices of hyperparameters in the prior of α, although we observe that the curve estimates under $\alpha \sim \Gamma(2, 2)$ are slightly closer to the truth than those under $\alpha \sim \Gamma(20, 2)$. Regarding the impact of L, we find that parameter and curve estimates are not sensitive to the choice of L, but hypothesis tests are more sensitive. For instance, when L increases from 4 to 5, the Bayes factor values become larger overall, especially for Scenario III (i.e. when the log-normal AFT is the truth), making our tests less conservative. For further comparison, we also fit the censored quantile regression model (Portnoy, 2003) using the function `crq` available in the `quantreg` R package (Koenker, 2008) under Scenario I, but with the data generated without frailties. Note that `crq` does not allow spatial information. In comparison to our approach, the results show that `crq` provides almost two times greater standard deviation estimates for non-intercept coefficients, and more troubling, the coverage probability for estimating β_2 is much lower than 95%. These findings inform us that ignoring heteroscedastic errors could result in badly overestimated standard deviations and low coverage probabilities.

4 Application to SEER prostate cancer data

We apply the proposed model to the prostate cancer survival data from the SEER program of the National Cancer Institute (see `http://seer.cancer.gov/`). The data set we consider consists of a cohort of 2999 men from the 64 counties of Louisiana, who have been diagnosed with prostate cancer in 2002, with follow-up continued through the end of 2011. In our analysis, the observed survival time is calculated as the amount of years from diagnosis to either death or the last follow-up, where death can be from any cause. By

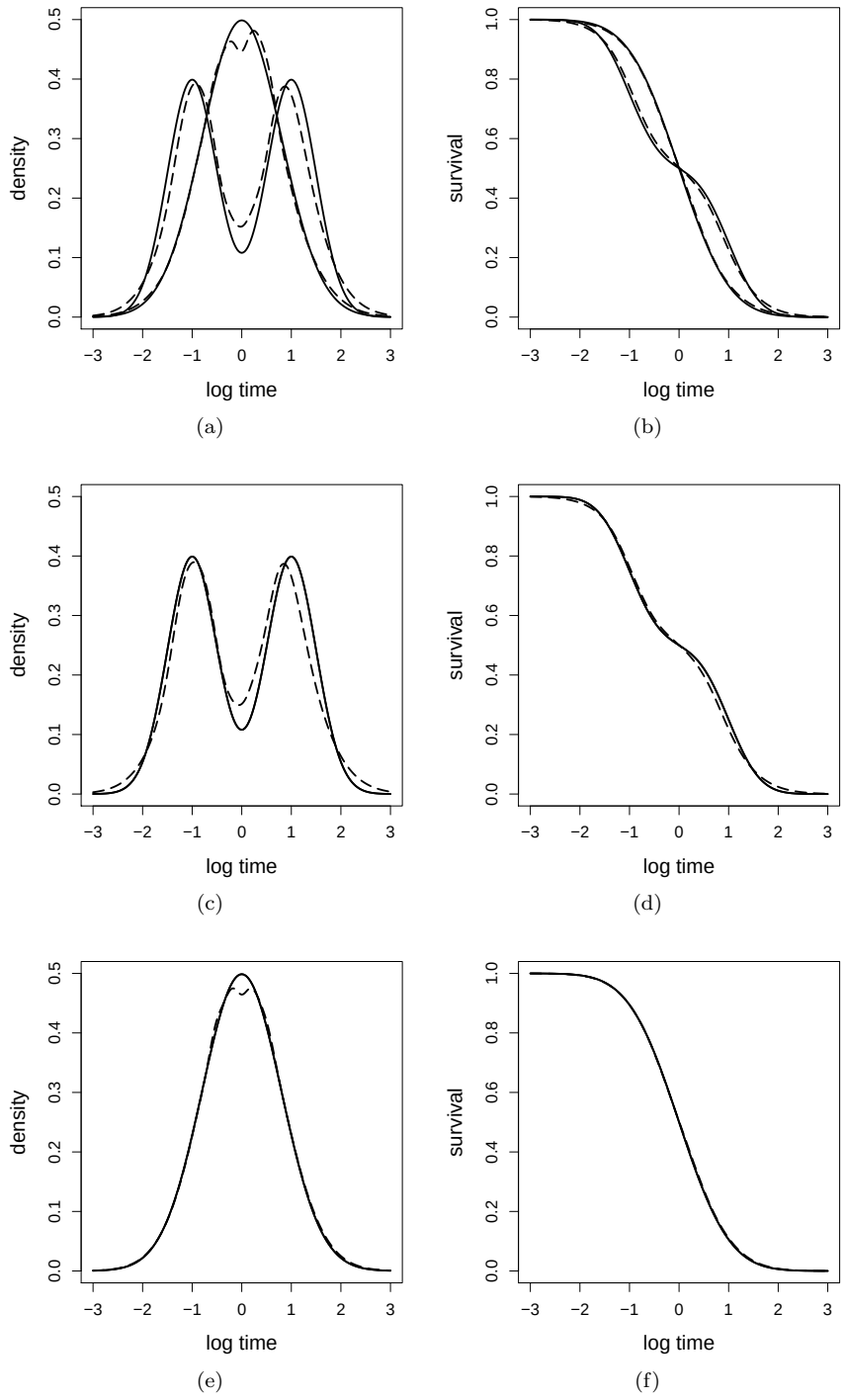


Fig. 1 Simulated data. Mean, across simulations, of the posterior mean of the density functions (left three panels) and survival functions (right three panels) of log survival times under Scenario I (panel a and b), Scenario II (panel c and d) and Scenario III (panel e and f). The curves in each panel of a and b are for $(z, x) = (1.5, 1)$ (initially left curve) and $(z, x) = (1, 0)$. The other curves are for $(z, x) = (1, 0)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines.

Table 3 SEER data. Summary characteristics for Louisiana prostate cancer patients diagnosed in 2002 with follow-up continued through the end of 2011.

Continuous	Mean	Median	Std. Dev.
follow-up (yrs)	7.63	9.17	2.87
Age (yrs)	67.52	68.00	9.45
Categorical	**Level**	**Count**	**Proportion**
Race	White	2159	0.72
	Black	840	0.28
Stage	Local/Regional	2875	0.96
	Distant	124	0.04
Marital	Married	2330	0.78
	Other	669	0.22
Grade	well/moderately differentiated	2220	0.74
	poorly/not differentiated	779	0.26

the end of 2011, 62.1% of patients who survived until the last follow-up are treated as right-censored. The observed survival time in years and county of residence at diagnosis are available for each individual. The individual-specific covariates at diagnosis include: age, race (white and black), SEER summary stage (localized/regional and distant), marital status at diagnosis (married and other), grade of tumor differentiation (well/moderately differentiated and poorly/not differentiated), where the first category in each above parenthesis is treated as reference. Table 3 presents several summary statistics for the data. Since the effects of stage, marital status and grade on baseline survival functions are not significant (per-variable Bayes factor is less than 1) based our initial model fitting via the proposed approach, we exclude them in modeling baseline functions presented below. Thus we have 5-dimensional $\mathbf{z}_{ij}$ and 2-dimensional $\mathbf{x}_{ij}$.

We fit the proposed GAFT spatial frailty model using the corresponding variants of the algorithm described in Section 2.4 and similar prior specifications used in the simulation study. Based upon examination of trace plots for model parameters, we run a single chain of 350,000, where 10,000 scans are thinned after a burn-in period of 50,000. In Section 4 of the online material, we present the posterior trace plots for $\tilde{\boldsymbol{\beta}}$, α and σ^2, and their autocorrelation function (ACF) plots together with effective samples sizes. The Markov chain mixed reasonably well regardless of the high dimension of parameters in our model.

For comparison, we further fit a semiparametric AFT spatial frailty model, where the baseline survival is fitted using the LDTFP but with intercept only, and a GAFT model without frailties. The three models are compared using the log pseudo marginal likelihood (LPML) developed by Geisser and Eddy (1979). Let $f_{\mathbf{z}_{ij}}(\cdot)$ be the density function corresponding to $S_{\mathbf{z}_{ij}}(\cdot)$. In the context of the proposed model, the LPML is defined as LPML =

Table 4 SEER data. Posterior means (95% credible intervals) of fixed effects $\tilde{\boldsymbol{\beta}}$ from fitting the proposed model and spatial frailty semiparametric AFT. The LPML is also shown for each model. Results are based on standardized ages.

Covariates	GAFT/CAR (LPML=-4110.5)	AFT/CAR (LPML=-4115.6)	GAFT (LPML=-4116.3)
Intercept	2.85(2.74, 2.96)	2.87(2.77, 2.96)	2.87(2.79, 2.96)
Age	-0.55(-0.62, -0.48)	-0.56(-0.62, -0.49)	-0.55(-0.61, -0.49)
Race	-0.23(-0.35, -0.11)	-0.29(-0.40, -0.19)	-0.23(-0.34, -0.14)
Marital	-0.33(-0.45, -0.22)	-0.32(-0.43, -0.20)	-0.30(-0.40, -0.21)
Grade	-0.23(-0.33, -0.12)	-0.25(-0.36, -0.14)	-0.21(-0.31, -0.12)
Stage	-1.55(-1.76, -1.34)	-1.52(-1.75, -1.29)	-1.57(-1.78, -1.37)

Table 5 SEER data. Bayes factors for testing each covariate effect on the baseline survival.

Covariates	Intercept	Age	Race	Overall
BF_{10}	> 1000	24.5	4.7	61.0

$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\log(\text{CPO}_{ij})$, where CPO_{ij}, referred to as conditional predictive ordinate, is given by $[f_{\mathbf{z}_{ij}}(t_{ij})^{\delta_{ij}}\left\{S_{\mathbf{z}_{ij}}(l_{ij})-S_{\mathbf{z}_{ij}}(u_{ij})\right\}^{1-\delta_{ij}}|\mathcal{D}_{(ij)}]$ with $\mathcal{D}_{(ij)}$ denoting the remaining data after excluding the ijth data point $\mathcal{D}_{ij}$. A larger value of LPML indicates better predictive ability for the corresponding model. Furthermore, Geisser and Eddy (1979) discussed the exponentiated difference in LPML values from two models to obtain the so called pseudo Bayes factor (PBF). The PBF is a surrogate for the traditional Bayes factor, and can be interpreted similarly, but is more analytically tractable, less sensitive to prior specifications, and does not suffer from Lindley's paradox. The method suggested by Gelfand and Dey (1994) can be used to estimate the CPO statistics from MCMC output.

Table 4 summarizes the results. The proposed GAFT model with CAR frailties has the largest LPML compared to the semiparametric AFT frailty model and non-frailty GAFT model, indicating that both allowing baseline survival function varying with covariates and taking into account spatial correlation improve model fit according to LPML. Note that the pseduo Bayes factor for the proposed model versus the AFT model is $e^{4115.6-4110.5}\approx 164$, indicating a decisive win for the GAFT/CAR model. Regarding the regression coefficient estimates, we see that all the three models give very similar covariate effects except for race, where GAFT/CAR model gives an estimate that is 26% higher than that from AFT model. This can be explained by the moderate significance of race on the baseline survival function based on the Bayes factors for per-value tests, as shown in Table 5. The age also highly affects the baseline survival according to the Bayes factor value.

For sake of visualization, Figure 2 presents the posterior mean curves under the proposed model. Panel (a) shows the baseline survival curves for white

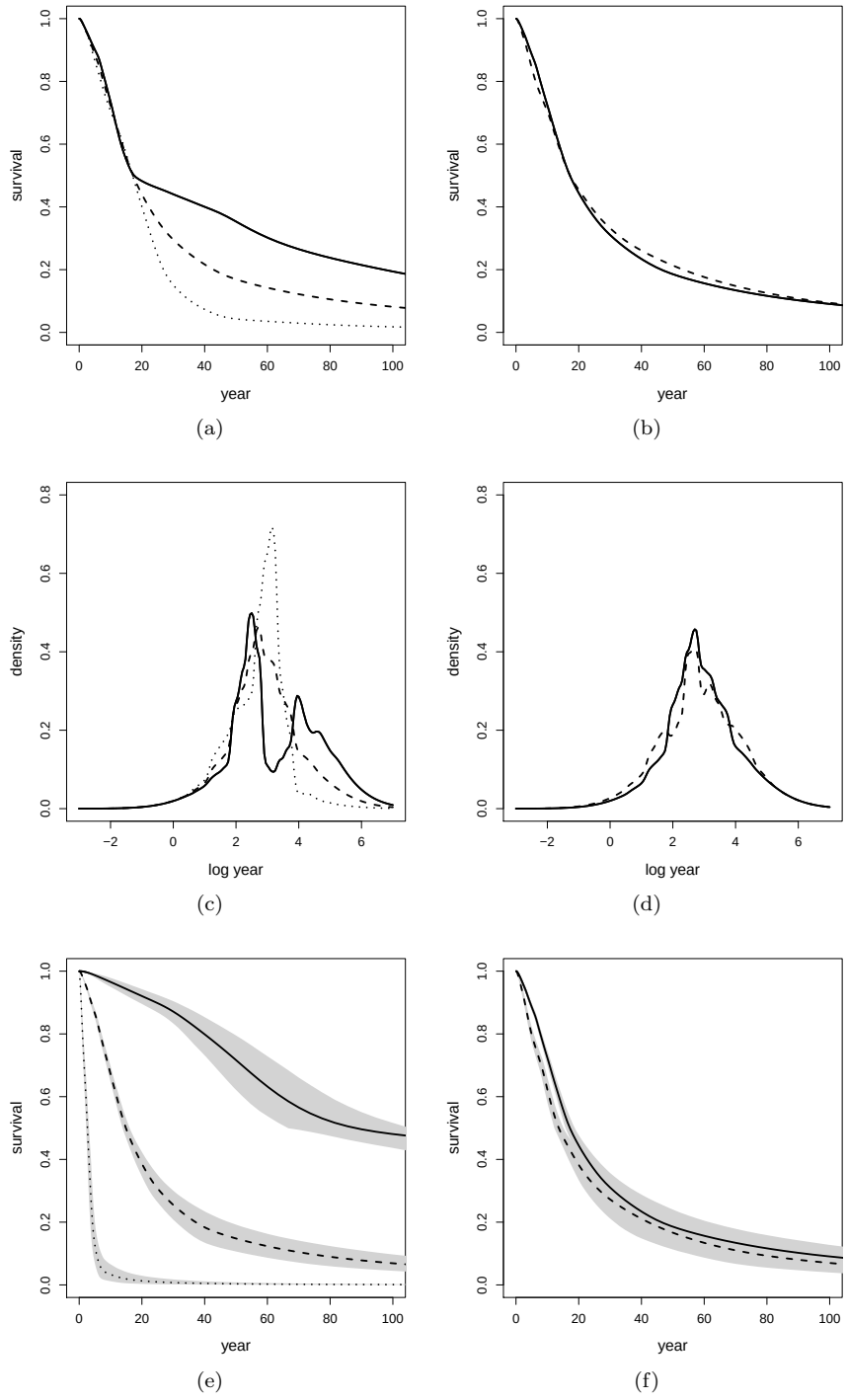
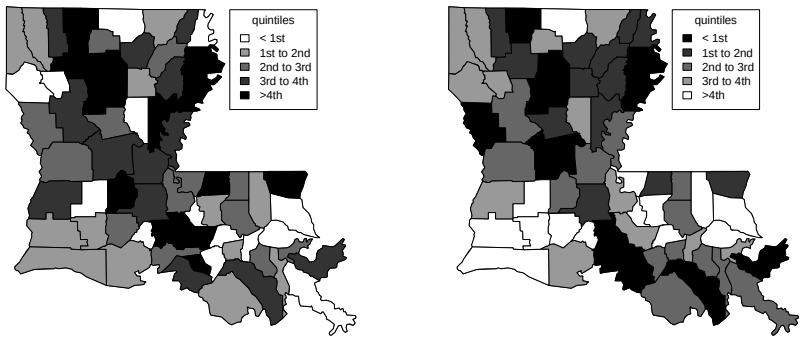


Fig. 2 SEER data. Posterior mean curves under the proposed model. Panel (a) shows the baseline survival curves for white patients with age=40 (solid), age=70 (dashed) and age=100 (dotted). Panel (b) shows the baseline survival curves for white (solid) versus black (dashed) patients with age at the population mean (67.5). Panels (c) and (d) show their corresponding baseline densities. Panel (e) presents the final survival curves for patients with age=40 (solid), age=70 (dashed) and age=100 (dotted), holding other covariates at the reference levels. Panel (f) presents the final suvival curves for white (solid) versus black (dashed) patients, holding other categorical covariates at the reference levels and the age at the population mean (67.5). The 90% point-wise credible intervals are shown in gray areas.



(a) (b)

Fig. 3 SEER data. Maps of the 2002 mortality rate (panel a) and CAR frailties (panel b) in the proposed model for Louisiana counties in 2002.

patients at three different ages. We see that the baseline curves coincide up to 18 years and then become very different, indicating that patients diagnosed at different age tend to have different baseline survival curves. Panel (b) shows that white and black patients have crossing baseline survival curves although the effect of race on baseline survival is not that significant. Panels (c) and (d) present the corresponding baseline densities, where we see that the baseline density changes from one mode to two as age increases. These results indicate that the traditional AFT assumption is violated for this data set, and the proposed model provides more valid inference. Finally, the final covariate-adjusted survival curves are presented for patients with three different ages in panel (e) and for white versus black patients in panel (f), where we see that race has a significant impact on survival up to 20 years after the diagnosis date.

The posterior means of spatial frailties for each county are mapped in Figure 3. The map shows that northern counties have relatively lower frailties and several counties in the southeast also exhibit lower spatial frailties. Since the frailties are additive to the logarithm of survival times in the proposed model, survival times are expected to be shorter in the regions with lower frailties. The mortality rates (percentages of death) based on our data set are also mapped for comparison, where we see that the counties showing lower frailties typically correspond higher mortality rates, providing support for the frailty modeling approach.

Based on the results from fitting the proposed model with CAR frailties as shown in Table 4, we see that all age, race, stage, marital status and grade are significant risk factors for survival of prostate cancer. Finally, note that $e^{-0.23} \approx 0.8$. The median lifetime for blacks is about 80% of that for whites, adjusting for other covariates and county of residence. The GAFT/CAR model

retains the easy interpretability in terms of typical lifetime as traditional AFT models, yet allows for a smoothly changing, heteroscedastic baseline survival. In Section 4 of the online material, the covariate effects are also compared with those obtained under the censored quantile regression model (Portnoy, 2003), where we note that the standard deviation for race effect is four times as large as that under GAFT, and consequently race becomes insignificant.

5 Discussion

We have proposed a general median-regression survival model with CAR frailties that includes the traditional semiparametric and parametric AFT models as special cases. The overall strength of the proposed model is that robust, flexible modeling assumptions make it appropriate for use in most real data applications, leading to significant improvement of the prediction of cancer survival in epidemiological cancer studies. We have also developed an efficient `R` function `frailtyGAFT` in the package `spBayesSurv`, as well as plotting functions, LPML estimation, etc., so others can easily use the proposed model.

Along with a general, flexible spatial survival model that retains easy interpretability but allows for crossing survival curves, we also offer simple tests for adequacy of the traditional AFT model with static nonparametric S_0, as well as per-variable tests for whether covariates impact S_0, implemented in the `frailtyGAFT` function. Crossing or partially coinciding survival curves is fairly common (Bouliotis and Billingham, 2011, Logan et al., 2008) in clinical studies. Often survival curves coincide or are negligibly different during the initial period and differences start to occur after treatments take effect. Panel (a) in Figure 2 shows coinciding survivals in the baseline $S_{0,\mathbf{x}}(t)$ for about 18 years post-diagnosis in the Louisiana SEER data, then marked differences for white patients with different diagnosed ages, holding other covariates constant. Our analysis of the Louisiana SEER data shows a highly significant difference in survival between blacks and whites; adjusting for county and other covariates, whites have a median lifetime that is about 26% greater than blacks. However, the traditional censored quantile regression of Portnoy (2003) fails to detect such significant racial difference.

A referee has made the following observation. It is possible to consider a multivariate CAR on the $(p+1)$-dimensional $\mathbf{v}_i$ (Gelfand and Vounatsou, 2003), yielding the model $\log(t_{ij}) = \tilde{\mathbf{z}}'_{ij}(\tilde{\boldsymbol{\beta}} + \mathbf{v}_i) + \epsilon_{ij}$, thus allowing for county-level changes in how predictors affect survival. The $\tilde{\boldsymbol{\beta}}$ are interpreted as overall regression effects and $\mathbf{v}_i$ are county-level deviations from $\tilde{\boldsymbol{\beta}}$.

Acknowledgements This work was supported by NCI grant 5R03CA176739. The authors would like to thank the editor, the associate editor, and the two referees for their valuable comments, which led to great improvements to the paper.

6 Appendix

Proposition 1 *Assume that* $\boldsymbol{\gamma}_{l,k}|\alpha \overset{ind.}{\sim} N_{q+1}\left(\mathbf{0}, \frac{2n}{\alpha\rho(l+1)}(\mathbf{X}'\mathbf{X})^{-1}\right)$ *under* H_1 *and* $\boldsymbol{\gamma}_{l,k,-j}|\alpha \overset{ind.}{\sim} N_q\left(\mathbf{0}, \frac{2n}{\alpha\rho(l+1)}(\mathbf{X}'_{-j}\mathbf{X}_{-j})^{-1}\right)$ *under* H_0*, where* α *is fixed and* $\mathbf{X}_{-j}$ *is the design matrix* $\mathbf{X}$ *excluding the* $(j+1)$*th column. Then the assumption (2.12) holds, and*

$$p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha) = \prod_{l=1}^{L-1}\prod_{k=1}^{2^l}\phi\left(0\middle|0, \frac{2n}{\alpha\rho(l+1)}(\mathbf{X}'\mathbf{X})^{-1}_{jj}\right).$$

where $(\mathbf{X}'\mathbf{X})^{-1}_{jj}$ *is the* $(j+1, j+1)$*th element of* $(\mathbf{X}'\mathbf{X})^{-1}$*, and* $\phi(\cdot|\mu, \sigma^2)$ *denotes the normal density with mean* μ *and variance* σ^2*.*

Proof Since $\boldsymbol{\gamma}_{l,k}|\alpha$ follows a multivariate normal, $(\boldsymbol{\gamma}_{l,k,-j}|\boldsymbol{\gamma}_{l,k,j} = 0, \alpha)$ still follows a multivariate normal distribution

$$\begin{aligned}
p(\boldsymbol{\gamma}_{l,k,-j}|\boldsymbol{\gamma}_{l,k,j} = 0, \alpha) &\propto \exp\left\{-\frac{\alpha\rho(l+1)}{4n}\boldsymbol{\gamma}'_{l,k}(\mathbf{X}'\mathbf{X})\boldsymbol{\gamma}_{l,k}\right\} \\
&\propto \exp\left\{-\frac{\alpha\rho(l+1)}{4n}\boldsymbol{\gamma}'_{l,k,-j}(\mathbf{X}'_{-j}\mathbf{X}_{-j})\boldsymbol{\gamma}_{l,k,-j}\right\} \\
&\propto N_q\left(\mathbf{0}, \frac{2n}{\alpha\rho(l+1)}(\mathbf{X}'_{-j}\mathbf{X}_{-j})^{-1}\right).
\end{aligned}$$

This implies that $p(\boldsymbol{\gamma}_{l,k,-j}|\boldsymbol{\gamma}_{l,k,j} = 0, \alpha) = p_0(\boldsymbol{\gamma}_{l,k,-j}|\alpha)$ and by independence $p(\boldsymbol{\Upsilon}_{-j}|\boldsymbol{\Upsilon}_j = 0, \alpha) = p_0(\boldsymbol{\Upsilon}_{-j}|\alpha)$. In addition, α is fixed and $\boldsymbol{\Upsilon}_{-j}$ is independent of all other parameters in $\boldsymbol{\psi}$, thus the assumption (2.12) holds. It is easy to evaluate $p(\boldsymbol{\Upsilon}_j = 0)$ by noting the properties of multivariate normal.

Proposition 2 *Assume the same priors on* $\boldsymbol{\gamma}_{l,k}$ *as Proposition 1 and additional prior on* α *as* $\pi(\alpha) = \Gamma(\alpha|a_0, b_0)$ *under both* H_1 *and* H_0*. Then given existence of all involved expectations,* BF_{10} *can be written as*

$$BF_{10} = \{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D})\}^{-1}\left\{E\left[\frac{1}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)}\right]\right\}^{-1},$$

where the expectation is with respect to $p(\alpha|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D})$*.*

Proof First note that $\boldsymbol{\psi}$ represents all remaining model parameters but $\boldsymbol{\Upsilon}_j$ and the prior for $\boldsymbol{\Upsilon}_j$ only depend on the precision parameter α, so we have $p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\boldsymbol{\psi}) = p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)$. Also note that $\mathcal{L}(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})$ is the likelihood function, so we could denote it by $p(\mathcal{D}|\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})$. It follows that

$$\begin{aligned}
&p(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}|\mathcal{D})\int \mathcal{L}(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})d(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}|\mathcal{D})\int p(\mathcal{D}|\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})d(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}|\mathcal{D})p(\mathcal{D}) \\
&= p(\mathcal{D}|\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi})p(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}) \\
&= \mathcal{L}(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi})p(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}).
\end{aligned}$$

Then we have

$$
\begin{aligned}
BF_{10}^{-1} &= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \frac{\mathcal{L}(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}) p(\boldsymbol{\psi})}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \mathcal{L}(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) d(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})} d\boldsymbol{\psi} \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \frac{\mathcal{L}(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}) p(\boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D})}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}|\mathcal{D}) \int \mathcal{L}(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) d(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})} d\boldsymbol{\psi} \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \frac{\mathcal{L}(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}) p(\boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D})}{\mathcal{L}(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}) p(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi})} d\boldsymbol{\psi} \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \frac{p(\boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D})}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi})} d\boldsymbol{\psi} \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \frac{p(\boldsymbol{\psi}|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D})}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\boldsymbol{\psi})} d\boldsymbol{\psi} \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \int \frac{1}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)} p(\boldsymbol{\psi}_{-\alpha}, \alpha|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D}) d\boldsymbol{\psi}_{-\alpha} d\alpha \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \frac{1}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)} p(\alpha|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D}) d\alpha,
\end{aligned}
$$

where $(\boldsymbol{\psi}_{-\alpha}, \alpha) = \boldsymbol{\psi}$.

References

Banerjee, S. and Carlin, B. P. (2003). Semiparametric spatio-temporal frailty modeling. *Environmetrics*, 14(5):523–535.

Banerjee, S. and Dey, D. K. (2005). Semiparametric proportional odds models for spatially correlated survival data. *Lifetime Data Analysis*, 11(2):175–191.

Banerjee, S., Wall, M. M., and Carlin, B. P. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*, 4(1):123–142.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B*, 36(2):192–236.

Bouliotis, G. and Billingham, L. (2011). Crossing survival curves: alternatives to the log-rank test. *Trials*, 12(Suppl 1):A137.

Chiou, S. H., Kang, S., and Yan, J. (2015). Semiparametric accelerated failure time modeling for clustered failure times from stratified sampling. *Journal of the American Statistical Association*, 110(510):621–629.

Christensen, R. and Johnson, W. (1988). Modeling accelerated failure time with a Dirichlet process. *Biometrika*, 75(4):693–704.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

De Iorio, M., Johnson, W. O., Müller, P., and Rosner, G. L. (2009). Bayesian nonparametric nonproportional hazards survival modeling. *Biometrics*, 65(3):762–771.

Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, 42(1):204–223.

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7(1):57–68.

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365):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(3):501–514.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4(1):11–15.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242.

Hanson, T. and Johnson, W. O. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association*, 97(460):1020–1033.

Hanson, T. and Johnson, W. O. (2004). A Bayesian semiparametric AFT model for interval-censored data. *Journal of Computational and Graphical Statistics*, 13(2):341–361.

Hanson, T., Kottas, A., and Branscum, A. (2008). Modelling stochastic order in the analysis of receiver operating characteristic data: Bayesian nonparametric approaches. *Journal of the Royal Statistical Society: Series C*, 57(2):207–225.

Hanson, T. E. (2006). Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association*, 101(476):1548–1565.

Hanson, T. E. and Jara, A. (2013). Surviving fully Bayesian nonparametric regression models. In *Bayesian Theory and Applications*, pages 592–615. Oxford University Press, Oxford.

Hanson, T. E., Jara, A., Zhao, L., et al. (2012). A Bayesian semiparametric temporally-stratified proportional hazards model with spatial frailties. *Bayesian Analysis*, 7(1):147–188.

Henderson, R., Shimakura, S., and Gorst, D. (2002). Modeling spatial variation in leukemia survival data. *Journal of the American Statistical Association*, 97(460):965–972.

Hennerfeind, A., Brezger, A., and Fahrmeir, L. (2006). Geoadditive survival models. *Journal of the American Statistical Association*, 101(475):1065–1075.

Jara, A. and Hanson, T. E. (2011). A class of mixtures of dependent tailfree processes. *Biometrika*, 98(3):553–566.

Koenker, R. (2008). Censored quantile regression redux. *Journal of Statistical Software*, 27(6):1–25.

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, 96(456):1458–1468.

Kuo, L. and Mallick, B. (1997). Bayesian semiparametric inference for the accelerated failure-time model. *Canadian Journal of Statistics*, 25(4):457–472.

Li, Y. and Ryan, L. (2002). Modeling spatial survival data using semiparametric frailty models. *Biometrics*, 58(2):287–297.

Logan, B. R., Klein, J. P., and Zhang, M.-J. (2008). Comparing treatments in the presence of crossing survival curves: an application to bone marrow transplantation. *Biometrics*, 64(3):733–740.

Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, 31(3):705–767.

Pang, L., Lu, W., and Wang, H. J. (2015). Local Buckley-James estimation for heteroscedastic accelerated failure time model. *Statistica Sinica*, 25(3):863–877.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, 98(464):1001–1012.

Raftery, A. E. (1996). Hypothesis testing and model selection via posterior simulation. In *Markov Chain Monte Carlo in Practice*, pages 163–187. Springer.

Robert, C. and Casella, G. (2005). *Monte Carlo Statistical Methods*. Springer.

Verdinelli, I. and Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association*, 90(430):614–618.

Walker, S. G. and Mallick, B. K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics*, 55(2):477–483.

Wang, S., Zhang, J., and Lawson, A. B. (2012). A Bayesian normal mixture accelerated failure time spatial model and its application to prostate cancer. *Statistical Methods in Medical Research*, http://dx.doi.org/10.1177/0962280212466189.

Zellner, A. (1983). Applications of Bayesian analysis in econometrics. *The Statistician*, 32(1/2):23–34.

Zhang, J. and Lawson, A. B. (2011). Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. *Journal of Applied Statistics*, 38(3):591–603.

Zhao, L. and Hanson, T. E. (2011). Spatially dependent Polya tree modeling for survival data. *Biometrics*, 67(2):391–403.

Zhao, L., Hanson, T. E., and Carlin, B. P. (2009). Mixtures of Polya trees for flexible spatial frailty survival modelling. *Biometrika*, 96(2):263–276.

Zhou, H., Hanson, T., Jara, A., and Zhang, J. (2015). Modeling county level breast cancer survival data using a covariate-adjusted frailty proportional hazards model. *The Annals of Applied Statistics*, 9(1):43–68.