



Multivariate Parametric Spatiotemporal Models for County Level Breast Cancer Survival Data

XIAOPING JIN

Division of Biostatistics, School of Public Health, University of Minnesota, Mayo Mail Code 303, Minneapolis, Minnesota 55455–0392, USA

BRADLEY P. CARLIN

brad@biostat.umn.edu

Division of Biostatistics, School of Public Health, University of Minnesota, Mayo Mail Code 303, Minneapolis, Minnesota 55455–0392, USA

Received March 5, 2003; Revised February 6, 2004; Accepted February 6, 2004

Abstract. In clustered survival settings where the clusters correspond to geographic regions, biostatisticians are increasingly turning to models with spatially distributed random effects. These models begin with spatially oriented frailty terms, but may also include further region-level terms in the parametrization of the baseline hazards or various covariate effects (as in a spatially-varying coefficients model). In this paper, we propose a multivariate conditionally autoregressive (MCAR) model as a mixing distribution for these random effects, as a way of capturing correlation across both the regions and the elements of the random effect vector for any particular region. We then extend this model to permit analysis of temporal cohort effects, where we use the term “temporal cohort” to mean a group of subjects all of whom were diagnosed with the disease of interest (and thus, entered the study) during the same time period (say, calendar year). We show how our spatiotemporal model may be efficiently fit in a hierarchical Bayesian framework implemented using Markov chain Monte Carlo (MCMC) computational techniques. We illustrate our approach in the context of county-level breast cancer data from 22 annual cohorts of women living in the state of Iowa, as recorded by the Surveillance, Epidemiology, and End Results (SEER) database. Hierarchical model comparison using the Deviance Information Criterion (DIC), as well as maps of the fitted county-level effects, reveal the benefit of our approach.

Keywords: cancer survival data, Geographic Information System (GIS), lattice data, Markov chain Monte Carlo methods, proportional hazards, random effects model

1. Introduction

With the advent of Markov Chain Monte Carlo (MCMC) computational methods, the Bayesian approach to fitting hierarchical survival models has become increasingly popular (see e.g., Ibrahim et al., 2001). The simplest approach is assume a parametric form for the baseline hazard. The Weibull is perhaps the most widely used parametric model among a variety of choices (gamma, lognormal, etc.), and seems to represent a good tradeoff between simplicity and flexibility.

To describe the basic approach, let t_{ij} be the time to death or censoring for subject j in stratum i , $j = 1, \dots, s_i$, $i = 1, \dots, p$. Let \mathbf{x}_{ij} be a vector of individual-specific covariates. With the usual assumption of proportional hazards $h(t_{ij}; \mathbf{x}_{ij})$, the same

covariate effects across strata, and a common Weibull baseline hazard in each stratum, the hazard is

$$h(t_{ij}; \mathbf{x}_{ij}) = h_0(t_{ij}) \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta}) = \rho t_{ij}^{\rho-1} \exp(\rho \beta_0 + \mathbf{x}_{ij}^T \boldsymbol{\beta}), \quad (1)$$

where h_0 is the baseline hazard, β_0 the log of scale parameter, and ρ is the shape parameter. We use the product of ρ and β_0 (instead of β_0 by itself) as the intercept in the log relative hazard since this reparametrization is computationally helpful. Specifically, in parametric survival models, ρ and β_0 tend to be less highly correlated in the posterior than ρ and $\beta_0^* \equiv \rho \beta_0$ (Louis, 1987), meaning that an MCMC algorithm using the former parametrization will converge more rapidly.

Carlin and Hodges (1999) consider several extensions of the basic model (1). First, they allow stratum-specific baseline hazards by replacing ρ by ρ_i , providing

$$h(t_{ij}; \mathbf{x}_{ij}) = h_{0i}(t_{ij}) \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta}) = \rho_i t_{ij}^{\rho_i-1} \exp(\rho_i \beta_0 + \mathbf{x}_{ij}^T \boldsymbol{\beta}). \quad (2)$$

MCMC fitting is routine given a distribution for these new random effects; say, $\rho_i \stackrel{iid}{\sim} G(\epsilon, 1/\epsilon)$, where $G(a, b)$ denotes the gamma distribution with mean ab and variance ab^2 . Thus, the ρ_i 's have mean 1 and variance $1/\epsilon$. Placing a flat prior on $\boldsymbol{\beta}$ and perhaps a vague $G(c, d)$ prior on ϵ completes our hierarchical specification.

The second extension of model (1) in Carlin and Hodges (1999) considers

$$h(t_{ij}; \mathbf{x}_{ij}) = h_{0i}(t_{ij}) \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta}) = \rho_i t_{ij}^{\rho_i-1} \exp(\rho_i \beta_{0i} + \mathbf{x}_{ij}^T \boldsymbol{\beta}), \quad (3)$$

which has unit-specific random scale parameters β_{0i} in addition to unit-specific random shape effects ρ_i . The β_{0i} are assumed to be exchangeable draws from $N(\mu_0, \sigma_0^2)$ with a flat prior on μ_0 and an $IG(c_0, d_0)$ prior on σ_0^2 , where IG denotes the inverse (reciprocal) gamma distribution.

In models (2) and (3), the stratum-specific baseline hazards are treated as draws from a population of hazard functions. When there are not too many strata, we may also carry out an ordinary stratified analysis, which assumes the strata to be completely unrelated; each stratum has its own fixed baseline hazard. Under the same assumption as before, stratum i 's hazard has the same form as the model (3), but in this model the shape parameters ρ_i and the log-scale parameters β_{0i} are fixed effects, precluding borrowing of strength across strata.

Another generalization of model (1) based on model (3) is

$$h(t_{ij}; \mathbf{x}_{ij}) = \rho_i t_{ij}^{\rho_i-1} \exp(\rho_i \beta_{0i} + \eta_i v_{ij} + \mathbf{x}_{ij}^T \boldsymbol{\beta}), \quad (4)$$

which adds a third set of unit-specific random effects corresponding to some covariate v_{ij} . This allows the size of the covariate effect to vary across strata. MCMC fitting is again routine given η_i 's as exchangeable draws from $N(0, \sigma^2)$ with an $IG(e, f)$ prior on σ^2 .

In this paper, we extend this modeling strategy to permit appropriate *spatial* and *multivariate* modeling of the random effects in survival models like (4). That is, it does not appear sensible to incorporate spatial association into the ρ_i , β_{0i} , and η_i mixing distributions, but ignore the correlation among these three components in any given stratum. Fortunately the multivariate conditionally autoregressive (MCAR)

distribution, introduced by Mardia (1988) and recently shown to be computationally feasible for hierarchical modeling by Gelfand and Vounatsou (2003) and Carlin and Banerjee (2003), emerges as ideal for rectifying this problem. In Section 2 we review CAR distributions in both the univariate and multivariate cases. Section 3 then describes our MCAR approach in the context of regionally referenced survival data, while the extension to the spatiotemporal case (as needed to capture cohort effects over time) is outlined in Section 4. After Section 5 briefly illuminates the likelihood and prior details and describes our approach to model choice in our hierarchical Bayesian formulation, Section 6 applies our approach to a data set of breast cancer survival rates taken from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program. Finally, Section 7 summarizes our findings and suggests avenues for future research in this burgeoning area.

2. Spatial Modeling

Sophisticated computer programs known as Geographic Information Systems (GISs) have allowed health science databases to incorporate geographical information about the subjects under study. Such databases have in turn generated interest among statisticians to develop and analyze models that can account for spatial clustering and variation. For instance, if the strata i in our Section 1 models correspond to county or other regional identifiers, we might suspect that the random effects $\{\rho_i\}$, $\{\beta_{0i}\}$, or $\{\eta_i\}$ corresponding to strata in geographic proximity to each other might also be similar in magnitude. This in turn means traditional exchangeable specifications for these random effects may be inappropriate.

Spatially varying random effects can be modeled in two general ways: *geostatistical approaches*, where we use the exact locations (e.g., latitude and longitude) of the strata, and *lattice approaches*, where we use only the positions of the strata relative to each other (e.g., which counties neighbor which others). While the former approach is quite sensible, it turns out to be difficult to fit due to the large amount of matrix inversion required. Moreover, the required assignment of each stratum to a single location (say, the regional centroid) is rather ad hoc. Banerjee et al. (2003) found that geostatistical models give almost the same results as lattice models in this setting, but at a roughly ten-fold increase in computer effort. As such, in this work we focus on lattice approaches. In Section 2.1 we review the general approach for a univariate spatial random process, while Section 2.2 describes the multivariate spatial random processes our data require.

2.1. Univariate CAR Modeling

Among lattice approaches for a univariate spatial random process ϕ_i , *conditionally autoregressive* (CAR) models (Besag, 1974) have been the most widely used. These models work equally well with regular or irregular lattice structures (e.g., checkerboard-like grids or arbitrarily sized and arranged geographical areas), and their

conditional specification makes them well-suited to model fitting using MCMC. In fact, this methodology has been automated in the GeoBUGS tools within the WinBUGS Bayesian software package (<http://www.mrc-bsu.cam.ac.uk/bugs/>).

The most popular CAR implementation is its pairwise difference formulation (Besag et al., 1995), which requires only the *adjacency* relationships among the various regions. To be specific, consider a vector $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^\top$ of p components that follows a multivariate Gaussian (normal) distribution with mean 0 and inverse dispersion matrix \mathbf{B} . Let W denote the adjacency matrix of the map (i.e., $w_{ii} = 0$, and $w_{i'i} = 1$ if i' is adjacent to i and 0 otherwise), m_i the number of neighbors of region i , and set $\mathbf{B} = \lambda(\text{Diag}(m_i) - W)$ where λ is a scale parameter. We may then write the joint distribution of a zero-centered CAR as

$$\boldsymbol{\phi} \sim N_p(\mathbf{0}, [\lambda(\text{Diag}(m_i) - W)]^{-1}), \quad (5)$$

which has corresponding full conditional distributions $\phi_i | \phi_{i' \neq i} \sim N(\bar{\phi}_i, 1/(\lambda m_i))$, where $\bar{\phi}_i = 1/m_i \sum_{i' \sim i} \phi_{i'}$, the average of the $\phi_{i'}$ in regions adjacent to region i . We denote this specification by the shorthand $\boldsymbol{\phi} | \lambda \sim \text{CAR}(\lambda)$.

In our pairwise difference setting, $\lambda(\text{Diag}(m_i) - W)$ is singular and its inverse does not exist, thus (5) is improper. To fix this, we can add the constraint $\sum_{i=1}^p \phi_i = 0$ (Carlin and Louis, 2000, Section 7.8.2), which can be imposed numerically by re-centering each sampled $\boldsymbol{\phi}$ vector around its own mean following each MCMC iteration. Another possible repair is to include a ‘‘propriety parameter’’ α in the precision matrix \mathbf{B} , i.e., set $\mathbf{B} = \lambda(\text{Diag}(m_i) - \alpha W)$. Carlin and Banerjee (2003) show that taking $|\alpha| < 1$ ensures that this \mathbf{B} is positive definite, resolving the impropriety. In fact, since negative smoothness parameters are not desirable (Wall, 2004), we usually take $0 < \alpha < 1$.

2.2. Multivariate CAR Modeling

The methodology described in Section 1 is for a single spatial random process. But in model (4), there are potentially three spatial random processes operating simultaneously (corresponding to the ρ_i , β_{0i} , and η_i). To handle this situation, consider the concatenated vector $\boldsymbol{\Phi}^\top = (\boldsymbol{\phi}_1^\top, \dots, \boldsymbol{\phi}_p^\top)$, so that $\boldsymbol{\Phi}$ is $np \times 1$ with each $\boldsymbol{\phi}_i$ being an n -dimensional vector (e.g., $\boldsymbol{\phi}_i = (\psi_i, \beta_{0i}, \eta_i)^\top$ where $\psi_i \equiv \log \rho_i$ and $n = 3$). How should $\boldsymbol{\Phi}$ be modeled?

An obvious first choice would be to use n separate univariate CAR models. That is,

$$\boldsymbol{\psi} | \lambda_1 \sim \text{CAR}(\lambda_1), \quad \boldsymbol{\beta}_0 | \lambda_2 \sim \text{CAR}(\lambda_2), \quad \text{and} \quad \boldsymbol{\eta} | \lambda_3 \sim \text{CAR}(\lambda_3) \quad .$$

Such an approach captures spatial variation in any one process, but ignores potential correlation among the random effects in the same region.

To remedy this, we propose a multivariate CAR (MCAR) model, which we present here following the notation of Carlin and Banerjee (2003). Analogous to the univariate situation, once again consider a multivariate Gaussian distribution for $\boldsymbol{\Phi}$ with mean 0 and inverse dispersion matrix \mathbf{B} , where now \mathbf{B} is a $np \times np$ symmetric positive definite matrix. Carlin and Banerjee (2003) show that a simple yet flexible specification for \mathbf{B} is obtained by setting

$$\mathbf{B} = (\text{Diag}(m_i) - W) \otimes \Lambda, \quad (6)$$

where \otimes denotes the Kronecker product and Λ is a $n \times n$ symmetric positive definite matrix. The joint distribution of a zero-centered MCAR is then

$$\boldsymbol{\phi} \sim N_{np}(\mathbf{0}, [(\text{Diag}(m_i) - W) \otimes \Lambda]^{-1})$$

and the corresponding full conditional distributions for the ϕ_i are

$$\phi_i | \phi_{i' \neq i} \sim N_n \left(\frac{1}{m_i} \sum_{i' \sim i} \phi_{i'}, \frac{1}{m_i} \Lambda^{-1} \right), \quad i = 1, \dots, p,$$

where $i' \sim i$ again denotes those regions i' spatially adjacent to region i . This latter expression reveals that the matrix Λ^{-1} describes the conditional variability and covariance relationships among the different spatial random process given the neighboring sites.

As with the univariate CAR model, there are some problems with the \mathbf{B} structure (6) due to its singularity. We can use the same repairs as in the univariate case to fix these problems. For example, we might add n centering constraints $\sum_{i=1}^p \phi_{ik} = 0$, for $k = 1, \dots, n$. We refer to this model as the MCAR(1, Λ). Alternatively, another repair would be to include a propriety parameter α in the precision matrix \mathbf{B} , that is,

$$\mathbf{B} = (\text{Diag}(m_i) - \alpha W) \otimes \Lambda.$$

We denote this specification as MCAR(α , Λ). As in the univariate case, if $|\alpha| < 1$, \mathbf{B} is positive definite (as long as Λ is), but we again typically take $0 < \alpha < 1$. Besides being a propriety parameter, α can be interpreted as a coefficient which measures spatial association for each spatial random process (Gelfand and Vounatsou, 2003). If $\alpha = 0$, the ϕ_i are independent across the strata.

Finally, Carlin and Banerjee (2003) observe that the above MCAR distribution imposes some rather unnatural symmetry properties on the $\boldsymbol{\Phi}$ covariance matrix. They thus suggest generalizing the model further by allowing n different α_k , corresponding to the n different spatial random processes. We denote this MCAR distribution as MCAR($\boldsymbol{\alpha}$, Λ) for $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^T$. However, this generalization comes at a significant price in terms of computing. Specifically, it requires rearranging the $\boldsymbol{\Phi}$ vector to $\boldsymbol{\Phi} = (\phi_{11}, \dots, \phi_{p1}, \dots, \phi_{1n}, \dots, \phi_{pn})^T$ (i.e., so that it has the region effects side by side), and then constructing the new \mathbf{B} matrix with the help of Cholesky factorizations of $\text{Diag}(m_i) - \alpha_k W$, $k = 1, \dots, n$. Since the α_k are unknown, this adds n factorizations to every MCMC iteration, a substantial increase in computational burden.

3. Multivariate Parametric Spatial Survival Models

As mentioned above, several papers have applied univariate CAR models in spatial survival settings. For instance, Banerjee et al. (2003) use a parametric proportional hazards model with a Weibull formulation for the baseline hazards, placing a univariate CAR structure on the frailty (intercept) terms (see below). Banerjee and

Carlin (2002) extend this approach to a semiparametric setup under the usual Cox proportional hazards model; Banerjee and Carlin (2003) further extend this model to the spatiotemporal case, but again using only a univariate CAR. Carlin and Banerjee (2003) employ the MCAR model we espouse, but the multivariate aspect in this paper lies only in the multiple counts on differing but related diseases (stomach cancer, pancreatic cancer, etc.); the various spatially-varying parameters (frailties, shape parameters and spatially varying coefficients) are still assumed to vary independently from each other.

In this section, we use the MCAR to model correlation across multiple *parameters* for the same disease (and in Section 4, extend this approach to the spatiotemporal setting). That is, we develop a spatiotemporal extension of Carlin and Hodges (1999) for use with our SEER data. The Banerjee et al. (2003) parametric model is

$$h(t_{ij}; x_{ij}) = h_0(t_{ij})\gamma_i \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta}) = \rho_0 t_{ij}^{\rho_0 - 1} \exp(\rho_0 \beta_0 + \omega_i + \mathbf{x}_{ij}^T \boldsymbol{\beta}), \quad (7)$$

where $\omega_i \equiv \log \gamma_i$ is the stratum-specific frailty term, designed to capture spatial variation among the strata, and modeled as $\boldsymbol{\omega} \sim \text{CAR}(\lambda)$.

Looking again at model (2), our first extension of model (7) is to allow stratum-specific baseline hazards, e.g., by replacing ρ by ρ_i , leading to

$$h(t_{ij}; \mathbf{x}_{ij}) = \rho_0 \rho_i t_{ij}^{\rho_0 \rho_i - 1} \exp(\rho_0 \rho_i \beta_0 + \omega_i + \mathbf{x}_{ij}^T \boldsymbol{\beta}). \quad (8)$$

Here we have two random effects, ρ_i and ω_i , and a fixed effect $\rho_0 > 0$ that allows centering of the ρ_i at some value other than 1. The ρ_i are random shape parameters and the ω_i are the random frailties. Traditionally, we would model $\psi_i = \log \rho_i$ and ω_i as i.i.d. mean-zero random variables across the strata, or using two separate, independent CARs on ψ_i and ω_i ; that is, $\boldsymbol{\psi} \sim \text{CAR}(\lambda_1)$ and $\boldsymbol{\omega} \sim \text{CAR}(\lambda_2)$. But following Section 2.2, an even more appropriate model here would be a bivariate spatial model for ψ_i and ω_i via the MCAR distribution. Several such distributions described in Section 2 can be used for this purpose, such as placing a $\text{MCAR}(1, \Lambda)$, $\text{MCAR}(\boldsymbol{\alpha}, \Lambda)$, or $\text{MCAR}(\boldsymbol{\alpha}, \Lambda)$ distribution on $\boldsymbol{\Phi} \equiv (\boldsymbol{\phi}_1^T, \dots, \boldsymbol{\phi}_p^T)^T$, where $\boldsymbol{\phi}_i = (\psi_i, \omega_i)^T$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$.

Next, based on model (4), another extension of model (7) is to consider a *spatially varying coefficients model* (Assunção, 2003). In this extension, some coefficients of covariates in model (7) are allowed to change in space. For example, in model (7), in addition to the spatial frailty term ω_i , we might also wish to model the coefficients of certain covariates (say, v_{1ij} and v_{2ij}) as both random and spatially varying. We might then write the hazard as

$$h(t_{ij}; \mathbf{x}_{ij}) = \rho_0 t_{ij}^{\rho_0 - 1} \exp(\rho_0 \beta_0 + \omega_i + \eta_{1i} v_{1ij} + \eta_{2i} v_{2ij} + \mathbf{x}_{ij}^T \boldsymbol{\beta}), \quad (9)$$

so that there are now three spatial random effects (ω_i , η_{1i} , and η_{2i}). We can again use three separate independent CAR priors here (i.e., $\boldsymbol{\omega} \sim \text{CAR}(\lambda_1)$, $\boldsymbol{\eta}_1 \sim \text{CAR}(\lambda_2)$, and $\boldsymbol{\eta}_2 \sim \text{CAR}(\lambda_3)$), or allow these three spatial random effects to themselves be correlated via $\text{MCAR}(1, \Lambda)$, $\text{MCAR}(\boldsymbol{\alpha}, \Lambda)$, or $\text{MCAR}(\boldsymbol{\alpha}, \Lambda)$ structures on $\boldsymbol{\Phi} \equiv (\boldsymbol{\phi}_1^T, \dots, \boldsymbol{\phi}_p^T)^T$, where $\boldsymbol{\phi}_i = (\omega_i, \eta_{1i}, \eta_{2i})^T$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T$.

Finally, we can combine models (8) and (9) to allow four spatial random processes. The hazard then becomes

$$h(t_{ij}; \mathbf{x}_{ij}) = \rho_0 \rho_i t_{ij}^{\rho_0 \rho_i - 1} \exp(\rho_0 \rho_i \beta_0 + \omega_i + \eta_{1i} v_{1ij} + \eta_{2i} v_{2ij} + \mathbf{x}_{ij}^T \boldsymbol{\beta}). \quad (10)$$

Again we would place four separate CARs on the four random effect collections, or else any of the above three MCAR distributions on the $\boldsymbol{\Phi}$ vector, where now $\boldsymbol{\phi}_i = (\psi_i, \omega_i, \eta_{1i}, \eta_{2i})^T$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^T$.

4. Spatiotemporal Survival Modeling

In a cohort study with K temporal cohorts (say, one for each year of study entry), subjects enter the study at different times, and are then followed until the occurrence of the primary endpoint or the end of study. As in spatiotemporal disease mapping (Waller et al., 1997), in our spatial survival setting we are often interested in temporal cohort effects as well as spatial effects. As such, in this section we extend our models to allow for cohort effects. As a first step, we might add a simple parametric function of time. For example, we can assume a linear time trend, and treat time as one of the continuous covariates in the spatial models discussed in Section 3. This would extend model (10) to

$$h(t_{ijk}; \mathbf{x}_{ijk}, z_{ijk}) = \rho_0 \rho_i t_{ijk}^{\rho_0 \rho_i - 1} \exp(\rho_0 \rho_i \beta_0 + \omega_i + \eta_{1i} v_{1ijk} + \eta_{2i} v_{2ijk} + \mathbf{x}_{ijk}^T \boldsymbol{\beta} + z_{ijk} \xi), \quad (11)$$

where $z_{ijk} = k - 1$, the year of diagnosis for j th patient living in county i and entering the database in year k , for $i = 1, \dots, p$, $k = 1, \dots, K$, and $j = 1, \dots, s_{ik}$.

Alternately, we might model time as a qualitative form that lets the data reveal any trend. Specifically, we could include dummy variables $z_{ijk}, k = 1, \dots, K - 1$, at most one of which will equal 1 for any given patient. This changes model (11) only slightly to

$$h(t_{ijk}; \mathbf{x}_{ijk}, \mathbf{z}_{ij}) = \rho_0 \rho_i t_{ijk}^{\rho_0 \rho_i - 1} \exp(\rho_0 \rho_i \beta_0 + \omega_i + \eta_{1i} v_{1ijk} + \eta_{2i} v_{2ijk} + \mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi}) \quad (12)$$

for $\mathbf{z}_{ij} = (z_{ij1}, z_{ij2}, \dots, z_{ij,K-1})^T$ and $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{K-1})^T$.

In both (11) and (12), time is modeled through one or more terms in the proportional hazards structure, which requires the hazard functions for different levels of a covariate to be proportional. If the different time periods produce hazard functions that deviate markedly from proportionality, these two approaches are no longer good choices. In this situation, we can treat time as a stratification factor and perform a stratified parametric analysis, as mentioned in Section 1. With the assumption of the same covariate effects across the strata, model (10) is extended to

$$h(t_{ijk}; \mathbf{x}_{ijk}) = \rho_{0k} \rho_i t_{ijk}^{\rho_{0k} \rho_i - 1} \exp(\rho_{0k} \rho_i \beta_{0k} + \omega_i + \eta_{1i} v_{1ijk} + \eta_{2i} v_{2ijk} + \mathbf{x}_{ijk}^T \boldsymbol{\beta}). \quad (13)$$

Each stratum k now has its own Weibull baseline hazard with shape parameter ρ_{0k} and log-scale parameter β_{0k} . The other models discussed in Section 3 can be

extended similarly by replacing ρ_0 by ρ_{0k} and β_0 by β_{0k} . In our data analyses we treat the ρ_{0k} and β_{0k} as fixed effects, but random mixing distributions (e.g., $\rho_{0k} \stackrel{iid}{\sim} G(a_0, b_0)$ for some a_0 and b_0 chosen so that the priors will have little impact relative to the data, or perhaps even an AR(1) structure) could be useful here, especially if data were scarce across cohorts.

Finally, in models (11)–(13) there is no spatiotemporal interaction. We can remedy this by incorporating spatially random cohort effects. For example, model (11) may be extended to

$$h(t_{ijk}; \mathbf{x}_{ijk}, z_{ijk}) = \rho_0 \rho_i t_{ijk}^{\rho_0 \rho_i - 1} \exp(\rho_0 \rho_i \beta_0 + \omega_i + \eta_{1i} v_{1ijk} + \eta_{2i} v_{2ijk} + \mathbf{x}_{ijk}^T \boldsymbol{\beta} + z_{ijk} \xi + \tau_i z_{ijk}),$$

where the τ_i capture the spatiotemporal interaction effects. For models (12) and (13) matters are more complicated, since now $\boldsymbol{\tau}_i$ is a vector so many more additional parameters are required.

5. Bayesian Implementation

The new MCAR models outlined above are straightforwardly implemented in a Bayesian framework using MCMC methods. Suppose we adopt model (13) in Section 4 (expressions for the other models in Sections 3 and 4 follow similarly). We can write the likelihood for this model as

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\delta}; \mathbf{t}, \mathbf{x}, \mathbf{v}, \boldsymbol{\gamma}) \\ \propto \prod_{k=1}^K \prod_{i=1}^I \prod_{j=1}^{s_{ik}} \left\{ \rho_{0k} \rho_i t_{ijk}^{\rho_{0k} \rho_i - 1} \exp(\rho_{0k} \rho_i \beta_{0k} + \omega_i + \eta_{1i} v_{1ijk} + \eta_{2i} v_{2ijk} + \mathbf{x}_{ijk}^T \boldsymbol{\beta}) \right\}^{\gamma_{ijk}} \\ \times \exp\left\{ -t_{ijk}^{\rho_{0k} \rho_i} \exp(\rho_{0k} \rho_i \beta_{0k} + \omega_i + \eta_{1i} v_{1ijk} + \eta_{2i} v_{2ijk} + \mathbf{x}_{ijk}^T \boldsymbol{\beta}) \right\}, \end{aligned} \quad (14)$$

where $\boldsymbol{\phi} = \{\phi_i\}$, $\phi_i = (\psi_i, \omega_i, \eta_{1i}, \eta_{2i})^T$, $\boldsymbol{\delta} = \{\delta_k\}$, $\delta_k = (\rho_{0k}, \beta_{0k})^T$, $\mathbf{t} = \{t_{ijk}\}$ denotes the collection of times to death, $\mathbf{x} = \{\mathbf{x}_{ijk}\}$ is the collection of covariate vectors corresponding to the fixed main effects, v_{1ijk} and v_{2ijk} are the covariates corresponding to the spatially varying coefficients, $\boldsymbol{\gamma} = \{\gamma_{ijk}\}$ is the collection of death indicators for all subjects in all strata, and s_{ik} is the number of patients in stratum i and cohort k . The joint posterior distribution of model (13) is given by

$$p(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\delta}, \boldsymbol{\Lambda}, \boldsymbol{\alpha}; \mathbf{t}, \mathbf{x}, \mathbf{v}, \boldsymbol{\gamma}) \propto L(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\delta}; \mathbf{t}, \mathbf{x}, \mathbf{v}, \boldsymbol{\gamma}) p(\boldsymbol{\phi} | \boldsymbol{\Lambda}, \boldsymbol{\alpha}) p(\boldsymbol{\beta}) p(\boldsymbol{\delta}) p(\boldsymbol{\Lambda}) p(\boldsymbol{\alpha}), \quad (15)$$

where the first term on the right-hand side is the likelihood (14), the second is $\text{MCAR}(\boldsymbol{\alpha}, \boldsymbol{\Lambda})$, and the remaining terms are prior distributions on the fixed effects $(\boldsymbol{\beta}, \boldsymbol{\delta})$ and MCAR hyperparameters $(\boldsymbol{\Lambda}, \boldsymbol{\alpha})$. Typically, flat priors are chosen for $\boldsymbol{\beta}$ and the β_{0k} , while vague $G(\epsilon, 1/\epsilon)$ priors (having mean 1 but variance $1/\epsilon$) are chosen for the ρ_{0k} . For the hyperpriors, we follow Carlin and Banerjee (2003) and place independent $\text{Unif}(0, 1)$ distributions on the components of $\boldsymbol{\alpha}$ and a $\text{Wishart}(R^{-1}, r)$

distribution on Λ . This latter distribution is parametrized to have mean $E(\Lambda) = rR^{-1}$ (Carlin and Louis, 2000, p. 328).

5.1. Computing

The Gibbs sampler (Gelfand and Smith, 1990; Carlin and Louis, 2000, Section 5.4.2) is used to update the parameters in this model, because it can take advantage of the MCAR's conditional specification. The Gibbs sampler requires drawing samples from the full conditional distributions of the model parameters, each of which must be proportional to (15). The complexity of the likelihood precludes closed-form full conditionals for most of the parameters in our models. However, the components of $\boldsymbol{\beta}$ may be conveniently updated using Metropolis–Hastings steps with univariate Gaussian proposals (see e.g., Carlin and Louis, 2000, Section 5.4.3). Moreover, Metropolis steps with multivariate Gaussian proposals may also be employed for updating the spatial effects ϕ_i by region. For the shape parameters ρ_{0k} , we reparametrize to the log scale (being careful to remember the Jacobian of this transformation), and then update the log ρ_{0k} rather than the ρ_{0k} directly, which is easily done again via Metropolis–Hastings steps with Gaussian proposals.

Finally, with regard to MCAR hyperparameters, our conjugate Wishart hyperprior for Λ results (after a bit of labor) in a Wishart full conditional; see Jin (unpublished) for details. For the propriety parameters α_k ($0 < \alpha_k < 1$), we again transform to a more convenient scale, this time using $\text{logit}(\alpha_k) = \log[\alpha_k/(1 - \alpha_k)]$, and update using Metropolis–Hastings steps with Gaussian proposals. In practice, we need to bound the α_k away from 1 (say, by insisting $0 < \alpha_k < 0.999$) to avoid identifiability (hence MCMC convergence) problems.

5.2. Hierarchical Model Choice

Bayesian comparison of a model M_1 versus another M_2 has historically been accomplished using the *Bayes factor*, the ratio of posterior to prior odds. However, Bayes factors can be difficult to compute using MCMC methods, and in any case are not well-defined for improper prior specifications such as ours. Thus, we are drawn to more informal model choice methods.

Penalized likelihood criteria, such as the Akaike Information Criterion (AIC; Akaike, 1973) and the Bayesian (Schwarz) Information Criterion (BIC; Schwarz, 1978), are computational shortcuts popular for use with traditional, nonhierarchical statistical models. In hierarchical models such as ours, however, it is difficult to determine the appropriate degrees of freedom to be used in the penalty term for each model. Spiegelhalter et al. (2002) provide a simple and intuitively appealing hierarchical modeling extension of the AIC criterion called the *Deviance Information Criterion*, or DIC. Thinking of $\boldsymbol{\theta}$ and \mathbf{y} as the entire collections of model parameters and data, respectively, DIC is based on the posterior distribution of the *deviance* function,

$$D(\boldsymbol{\theta}) = -2 \log f(\mathbf{y}|\boldsymbol{\theta}) + 2 \log h(\mathbf{y}) , \quad (16)$$

where $f(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood function for the observed data vector \mathbf{y} given the parameter vector $\boldsymbol{\theta}$, and $h(\mathbf{y})$ is some standardizing function of the data alone (which thus has no impact on model selection; see below). In the DIC approach, the *fit* of a model is summarized by the posterior expectation of the deviance, $\bar{D} = E_{\theta|\mathbf{y}}[D]$, while the *complexity* of a model is captured by the effective number of parameters, p_D , which is defined as

$$p_D = E_{\theta|\mathbf{y}}[D] - D(E_{\theta|\mathbf{y}}[\boldsymbol{\theta}]) = \bar{D} - D(\hat{\boldsymbol{\theta}}) ,$$

i.e., the expected deviance minus the deviance evaluated at the posterior expectations. Typically, this “effective” parameter total p_D will be less than the actual total number of parameters in the model, due to the borrowing of strength across random effects. The DIC is then defined analogously to the AIC as the expected deviance plus the effective number of parameters, i.e.,

$$\text{DIC} = \bar{D} + p_D = D(\hat{\boldsymbol{\theta}}) + 2p_D$$

with the latter expression clarifying the analogy with AIC. Since small values of \bar{D} indicate good fit while small values of p_D indicate a parsimonious model, small values of the sum (DIC) indicate preferred models. As with AIC and other penalized likelihood criteria, DIC is not intended for identification of the “correct” model, but merely as a method of comparing a collection of alternative formulations (all of which may be incorrect). Note also that DIC is scale-free; since any choice of standardizing function $h(\mathbf{y})$ in (16) is free of $\boldsymbol{\theta}$, it will contribute equally to each model’s DIC score (and thus have no impact on model selection). Spiegelhalter et al. (2002, Sections 6.2 and 8.1) remark that if one selects $h(\mathbf{y}) = f(\mathbf{y}|\hat{\boldsymbol{\theta}}(\mathbf{y}))$ where $\hat{\boldsymbol{\theta}}(\mathbf{y})$ is the usual maximum likelihood estimate of $\boldsymbol{\theta}$, then if the model is true we have $\bar{D} = E_{\theta|\mathbf{y}}[D] \approx p$, the number of free parameters in $\boldsymbol{\theta}$. So while this choice of $h(\mathbf{y})$ can be used as a check on overall model goodness of fit, to compare models and estimate their effective size p_D , it is sufficient simply to take $h(\mathbf{y}) = C$ for any (model-independent) constant C , and this is what we do. Thus our DIC values have no intrinsic meaning; only *differences* in DIC across models are meaningful.

Note that DIC is readily calculated as part of an MCMC sampling run. For example, again illustrating with model (13) in Section 4, to compute DIC we need only calculate the average deviance \bar{D} , and the deviance of the posterior mean, $D(\hat{\boldsymbol{\theta}})$. In our case the deviance function is $D(\boldsymbol{\theta}) \equiv D(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\delta}) = -2 \log L(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\delta}; \mathbf{t}, \mathbf{x}, \mathbf{v}, \gamma) + 2 \log C$, where L is the likelihood given in (14).

In view of DIC’s relatively short history of use in practice, one might well wonder whether other Bayesian model comparison tools might be available. Unfortunately, given the highly hierarchical framework of our models, the only possibility might be the predictive penalty approach of Gelfand and Ghosh (1998). This method works in a maximum expected utility framework, where the user supplies the utility function and his relative preference for maintaining fidelity to the data and controlling the model’s size (DIC instead works with a particular loss

function and equal weighting of the two components). While experience with this method is accruing, it remains relatively little-used due to its requirement of utility specification and its lack of inclusion in Bayesian software packages like WinBUGS. Moreover, DIC can be calculated for each model being considered without analytic adaptation, additional MCMC sampling (say, of predictive values), or any matrix inversion. For these reasons, in what follows we adopt DIC as our model choice criterion to compare the various parametric spatiotemporal models for the Iowa SEER breast cancer data set.

6. Application to SEER Breast Cancer Data

The National Cancer Institute's SEER program (<http://seer.cancer.gov>) is perhaps the most complete and authoritative source of cancer data in the US, offering public use county-level summaries on a yearly basis for several states in various parts of the country. We apply the methodology above to the analysis of Iowa SEER breast cancer survival data.

The SEER data we consider consist of 37,610 white women from the 99 counties of the state of Iowa who have been diagnosed with breast cancer between 1973 and 1994 (i.e., the women are classified into 22 temporal enrollment cohorts). Each woman received treatment and was progressively monitored until the end of 1999, so that each was followed for a minimum of five years. To reduce potential sources of bias, we excluded the 1075 patients (2.9% of the total) whose ages were greater than or equal to 100 at diagnosis or who had an undefined stage, leaving us 36,535 subjects for final analysis.

Death from breast cancer (rather than all causes) is often taken as the endpoint in studies like ours, but researchers' opinions on this issue seem to vary. Brinkley et al. (1984) found that about a third of breast cancer patients who died from other causes still had overt signs of breast cancer. Thus it may not be reasonable to assume that deaths from other causes are independent of those from breast cancer. As such, we follow Hemming and Shaw (2002) and many other authors in choosing death from any cause as our endpoint in the following analysis. Only those 16,088 patients (about 44% of the total) who dropped out of the study or survived until the end of the study period are considered censored.

For each individual, the SEER data set records the time in months that the patient survived, her county of residence at diagnosis, and the year of diagnosis. Several individual-level covariates are also available, including age at diagnosis, the number of primaries (i.e., the total number of primary tumors diagnosed for this patient), and the stage of breast cancer (in situ, local, regional, or distant). Other individual-level covariates are not available, but county-level covariates possibly associated with breast cancer risk, such as education (percentage with high school degree or above), median household income, and county population density, are obtained from 1980 Iowa census database, www.silo.lib.ia.us/specialized-services/datacenter/browse/estimates.html.

Since it is not a priori reasonable to assume that the effect of age on mortality is linear, age at diagnosis was not included as a continuous variable, but rather grouped into four categories: those under 50 years ($n = 6680$), those between 50 and 64 years ($n = 10,762$), those between 65 and 74 years ($n = 9153$), and those 75 years and over ($n = 9940$). We denote these as age groups I, II, III, and IV, respectively. We also denote local, in situ, regional, and distant as stage groups I, II, III, and IV, respectively. Patients in both the age I and stage I groups (i.e., youngest women with localized tumors) are treated as the baseline in the following analysis.

Turning to prior specifications, for all of the models described in Sections 3 and 4, we take a $\text{Unif}(0, 1)$ prior for α , and make the ρ_{0k} prior vague by setting $\epsilon = 0.001$ (so that the ρ_{0k} have mean 1 and variance 10^3). This latter specification essentially returns the ρ_{0k} to fixed effect status, since it essentially precludes shrinkage of the ρ_{0k} toward each other. With regard to the $\text{Wishart}(R^{-1}, r)$ prior for Λ , we choose r equal to the rank of the Λ matrix (i.e., the smallest value for which this prior is proper), ensuring it will be minimally informative. However, to ensure acceptable MCMC convergence we also choose appropriate and effect-specific values for R . For example, $R = 0.1I$ is chosen for our MCAR models without spatially random Weibull log-shape parameters ψ_i , since we found this prior was not overly influential relative to the information in the data. However, since our data encourage spatially random Weibull log-shape parameters ψ_i very close to 0, R specification for these parameters requires a trade-off. Larger R elements, such as 0.1 or 1.0, lead to a prior in conflict with the data, since such values imply rather disparate ψ_i , when in fact our data encourage them all to be near 0. On the other hand, very small R elements, such as 0, lead to MCMC convergence failure because such a prior cannot distinguish among the many very large values of Λ that are consistent with our data. As such, we choose the compromise value 0.01 to eliminate prior-data conflict while still preserving an acceptable degree of algorithm performance. Thus we take $R = \text{Diag}(0.01, 0.1, \dots, 0.1)$ when we include the ψ_i as the first component. See Jin (unpublished) for more details, as well as a brief robustness study of alternate R choices that supports this approach.

Note that we can use the `survreg` function in R or the `LIFEREG` procedure in SAS to fit a simple Weibull regression model that includes only fixed effects. The estimates from such a model can help us choose the initial values for these fixed effects in our MCMC sampler. We then simply set the random effects equal to zero. For our data set, the initial values turn out not to be critical for MCMC implementation; the results are robust to this choice.

For each model, we first ran a few initially overdispersed parallel MCMC chains, and monitored them using measurements of sample autocorrelations within the chains, cross-correlations between parameters, and plots of the sample traces. From these, we decided 5000 iterations was sufficient for the pre-convergence “burn-in” period. We then used a further 10,000 iterations as our “production” run for posterior summarization. Unfortunately, the complexity of our models and a large data set precluded us from running them in WinBUGS. Instead, we relied on programs written in C and executed in R (<http://www.r-project.org>) using the `.C` function.

Random number generation and posterior summarization were accomplished in R, while ArcView was used for mapping the results. A link to the C and R code for our chosen model can be found on the second author's website, www.biostat.umn.edu/~brad/software.html.

6.1. Model comparison

In this section, we use the DIC criterion described in Section 5.2 to compare the various parametric spatiotemporal models for the Iowa SEER breast cancer data set. In Table 1, the p_D and DIC scores listed for Case I correspond to models in which year of diagnosis is treated as a univariate continuous covariate z_{ijk} with value ranging from 0 to 21, as in (11). That is, we set $z_{ijk} = k - 1$, so that for data from diagnosis year 1973 ($k = 1$) we set $z_{ijk} = 0$, for year 1974 ($k = 2$) we set $z_{ijk} = 1$, and so on up to $z_{ijk} = 21$ for 1994 ($k = 22$). In Case II, we instead follow (12), and handle year of diagnosis using a vector of 21 dummy variables $\mathbf{z}_{ij} = (z_{ij1}, z_{ij2}, \dots, z_{ij,21})^T$, where $z_{ijk} = 1$ if the patient was diagnosed in year $1973 + k$ and 0 otherwise, $k = 1, \dots, 21$. In both cases, 1973 is the baseline year, and we assume a common baseline hazard for all years having a Weibull distribution with shape parameter ρ_0 and log-scale parameter β_0 . Finally, in Table 2, the years of diagnosis are now treated as strata, as in (13); we refer to this situation as Case III. Here each year has its own Weibull baseline hazard with shape parameter ρ_{0k} and log-scale parameter β_{0k} , $k = 1, \dots, 22$.

In Cases II and III, year of diagnosis is treated as a categorical variable with 22 levels. There are not enough events in some age-stage-year groups, some of which have all their observations censored. As a result, it is very hard to obtain MCMC convergence for certain fixed effects if we include terms for three-way interaction among age, stage, and year. Also as mentioned above, it is very time-consuming to explicitly model spatiotemporal interaction in Cases II or III. Thus, in order to compare across cases, the models in DIC Tables 1 and 2 assume that there is no three-way interaction and model only two-way interactions, namely age \times stage, age \times year, and year \times stage. We also assume that there is no spatiotemporal interaction (i.e., the parameters in our MCAR models need have only i and not ik subscripts). In both tables, the vector \mathbf{x}_{ijk} includes $4 \times 4 - 1 = 15$ dummy variables for the age-stage main effects, with category I coded as the baseline for both age and stage. For example, $x_{ijk1} = 1$ if the patient belongs to age-stage groups I–II, otherwise $x_{ijk1} = 0$; $x_{ijk2} = 1$ if the patient belongs to age-stage groups I–III, otherwise $x_{ijk2} = 0$; and so on. We found MCMC convergence is improved if we code the dummy variables in this way.

The vector \mathbf{x}_{ijk} also includes the number of primaries for each individual, and the education level, median household income, and population density for the individual's county of residence. For the county-level covariates we selected the 1980 Census data as most representative for our study period (1973–1994). To improve MCMC convergence, we center the covariates in the vector \mathbf{x}_{ijk} around their own means (Gelfand et al., 1995, 1996). Next, the vectors $\mathbf{y}_{ij} = (\mathbf{y}_{ij1}, \dots, \mathbf{y}_{ij6})^T$ are used for

Table 1. DIC comparison, parametric spatial models (Cases I and II) for the Iowa SEER breast cancer data.

Model	Log-relativehazard	Case I		Case II	
		p_D	DIC	p_D	DIC
1	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \beta_0$	27.9	251.2	167.8	292.1
2	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \beta_0 + \omega_i, \omega_i \sim \text{CAR}(\lambda)$	61.6	265.7	202.1	305.8
3	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \rho_i \beta_0 + \omega_i, (\psi_i, \omega_i)^T \sim \text{MCAR}(\alpha_1, \alpha_2, \Lambda)$	68.1	267.0	211.8	308.6
4	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \beta_0 + \omega_i + \delta_{1i} w_{1ijk}, (\omega_i, \delta_{1i})^T \sim \text{MCAR}(\alpha_1, \alpha_2, \Lambda)$	73.2	259.0	213.5	299.7
5	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \beta_0 + \delta_{1i} w_{1ijk} + \boldsymbol{\zeta}_i^T \mathbf{u}_{ijk}, (\delta_{1i}, \boldsymbol{\zeta}_i)^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \Lambda)$	131.4	254.9	264.1	294.2
6	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \beta_0 + \delta_{1i} w_{1ijk} + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk}, (\delta_{1i}, \boldsymbol{\eta}_i)^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \Lambda)$	145.1	192.9	280.3	235.0
7	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \beta_0 + \delta_{2i} w_{2ijk} + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk}, (\delta_{2i}, \boldsymbol{\eta}_i)^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \Lambda)$	135.0	200.6	264.5	232.1
8	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \beta_0 + \delta_{3i} w_{3ijk} + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk}, (\delta_{3i}, \boldsymbol{\eta}_i)^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \Lambda)$	134.9	197.0	265.6	229.1
9	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \rho_i \beta_0 + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk}, (\psi_i, \boldsymbol{\eta}_i)^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \Lambda)$	143.2	191.9	270.9	234.0
10	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \beta_0 + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk}, (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \Lambda)$	121.9	187.6	256.9	229.5
11	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \beta_0 + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk}, (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha, \Lambda)$	122.3	190.4	259.3	228.9
12	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \beta_0 + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk}, (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha = 1, \Lambda)$	120.1	204.4	256.1	236.8
13	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \boldsymbol{\xi} + \mathbf{y}_{ij}^T \boldsymbol{\gamma} + \rho_0 \beta_0 + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk}, (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim N_4(0, \Lambda)$	154.5	206.4	290.4	240.6

Table 2. DIC comparison, parametric spatial models (Case III) for the Iowa SEER breast cancer data.

Model	Log-relative hazard	Case III	
		p_D	DIC
1	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k$	188.7	275.2
2	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k + \omega_i, \omega_i \sim \text{CAR}(\lambda)$	220.7	286.6
3	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k + \omega_i, (\psi_i, \omega_i)^T \sim \text{MCAR}(\alpha_1, \alpha_2, \Lambda)$	227.6	288.1
4	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k + \omega_i + \delta_{1i} w_{1ijk}, (\omega_i, \delta_{1i})^T \sim \text{MCAR}(\alpha_1, \alpha_2, \Lambda)$	233.9	283.9
5	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k + \delta_{1i} w_{1ijk} + \zeta_i^T \mathbf{u}_{ijk},$ $(\delta_{1i}, \zeta_{1i}, \zeta_{2i}, \zeta_{3i}, \zeta_{4i})^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \Lambda)$	285.0	277.3
6	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k + \delta_{1i} w_{1ijk} + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk},$ $(\delta_{1i}, \eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \Lambda)$	290.7	216.0
7	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k + \delta_{2i} w_{2ijk} + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk},$ $(\delta_{2i}, \eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \Lambda)$	285.3	218.1
8	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k + \delta_{3i} w_{3ijk} + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk},$ $(\delta_{3i}, \eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \Lambda)$	285.6	219.7
9	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \rho_i \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk},$ $(\psi_i, \eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \Lambda)$	293.7	224.9
10	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk}, (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \Lambda)$	278.8	212.8
11	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk}, (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha, \Lambda)$	278.5	214.8
12	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk}, (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha = 1, \Lambda)$	275.8	218.1
13	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \rho_{0k} \beta_{0k} + \mathbf{y}_{ijk}^T \boldsymbol{\gamma}_k + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk}, (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim N_4(0, \Lambda)$	312.3	232.8

describing the interaction between age and year, and the interaction between stage and year. Recalling the baselines are groups I for age and stage, we use \mathbf{y}_{ij1} , \mathbf{y}_{ij2} and \mathbf{y}_{ij3} to model age-year interaction for age groups II, III, and IV, respectively. For example, $\mathbf{y}_{ij1} = \mathbf{z}_{ij}$ when the patient is in age group II, otherwise $\mathbf{y}_{ij1} = 0$, and so on. Similarly, \mathbf{y}_{ij4} , \mathbf{y}_{ij5} and \mathbf{y}_{ij6} will model stage-year interaction for stage groups II, III, and IV, respectively. For example, $\mathbf{y}_{ij4} = \mathbf{z}_{ij}$ when the patient is in stage group II, otherwise $\mathbf{y}_{ij4} = 0$, and so on. Like \mathbf{z}_{ij} , in Case I \mathbf{y}_{ijl} , $l = 1, \dots, 6$, is a scalar, while in Cases II and III it is a vector, namely $\mathbf{y}_{ijl} = (y_{ijl1}, y_{ijl2}, \dots, y_{ijl,21})^T$. In these two cases it is more convenient to group things by year, writing $\mathbf{y}_{ijk} = (y_{ijk1}, \dots, y_{ijk6})^T$ for $k = 1, \dots, 21$.

For the models with spatially varying coefficients, w_{1ijk} is the number of primaries, w_{2ijk} is county education level, and w_{3ijk} is county median household income. We denote the vector of indicator variables for the four stage groups by $\mathbf{v}_{ijk} = (v_{1ijk}, v_{2ijk}, v_{3ijk}, v_{4ijk})^T$. Similarly, we use the vector $\mathbf{u}_{ijk} = (u_{1ijk}, u_{2ijk}, u_{3ijk}, u_{4ijk})^T$ for the four age groups. Both of those quantities are also included in the fixed effect vector \mathbf{x}_{ijk} , which is sensible once we recall our MCAR prior is parametrized to have mean 0.

Tables 1 and 2 give the effective model size p_D and DIC scores for a variety of spatial survival models. Model 1 has fixed effects only, while Model 2 adds a standard univariate CAR frailty term ω_i . Models 3 adds the spatially random Weibull log-shape parameters ψ_i and a bivariate MCAR structure for the (ψ_i, ω_i) pairs.

Model 4 deletes the ψ_i and adds the random spatially varying coefficients δ_{1i} corresponding to the number of primaries w_{1ijk} .

Models 5–9 offer five-dimensional MCAR distributions with distinct smoothing parameters $\alpha_l, l = 1, \dots, 5$. In Model 5, we have spatial random effects for the number of primaries plus four age groups, while in Model 6 spatial random effects are for the number of primaries and the four stage groups. In Models 7–9, we delete the δ_{1i} from Model 6 and instead add random effects for education w_{2ijk} , income w_{3ijk} , and log-shape parameters ψ_i , respectively. Finally, Models 10–13 consider different priors for the four stage random effects $(\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})$, namely i.i.d. Gaussian, MCAR with $\alpha = 1$, MCAR with a single but unknown α , and MCAR with four different $\alpha_l, l = 1, \dots, 4$. Notice that for all of Models 5–13, the overall intercept ω_i is no longer needed, since there is a distinct frailty for every age (stage) group in the $\zeta(\boldsymbol{\eta})$ vector.

For Cases I–III, the patterns of change in p_D and DIC are generally similar across the 13 models. Higher model complexity leads to an increase in p_D , but not always to a decrease in DIC. For instance, DIC and p_D both increase in Models 2–5 relative to Model 1, and in the five-dimensional MCAR Models 6–9 relative to the four-dimensional MCAR Model 10. This suggests situations where the additional complexity does not produce a corresponding improvement in model fit. The effects of log-shape parameters ψ_i , number of primaries, age, education, and income also do not seem to significantly differ across the counties. However, comparing Model 10 to Model 1, the spatially varying coefficients for stage do lead to a decrease in DIC, despite the higher p_D score. Also, the decrease in both DIC and p_D comparing the MCAR Model 10 to the i.i.d. Gaussian Model 13 suggests that there is a spatial pattern to and spatial shrinkage for the stage effects.

The Case I DIC scores are significantly smaller than the corresponding ones in Case III, which are in turn smaller than those in Case II. The models treating year of diagnosis as a stratifying or dummy variable do not seem to offer a benefit in model fit, compared to the models treating year as a mere continuous covariate in the proportional hazards model. Note that in Cases II or III, to include the interaction between age and year, the interaction between stage and year, and the main effects of year requires a total of $22 \times 7 = 154$ parameters, whereas in Case I, only 7 parameters are necessary.

Overall, Model 10 in Table 1 (Case I) offers the lowest DIC score. Since in this model, year of diagnosis is treated as a univariate continuous covariate, we can reinsert the three-way (age-stage-year) interaction and the spatiotemporal interaction terms, to check our model assumptions. Model 1' in Table 3 takes account of spatiotemporal interaction by adding a spatially varying coefficient τ_i for year effects into Model 10. Now we have five spatial random effects $(\tau_i, \eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})$. Building on Model 1', Model 2' includes the three-way interaction, and thus becomes the “full model.” To improve MCMC convergence, we do this by defining a new dummy vector $\mathbf{T}_{ijk} = (t_{ijk1}, \dots, t_{ijk,16})^T$. For example, $t_{ijk1} = z_{ijk}$ if the patient is in age–stage group I–I, otherwise $t_{ijk1} = 0$; $t_{ijk2} = z_{ijk}$ if the patient is in age–stage groups I–II, otherwise $t_{ijk2} = 0$; and so on.

Table 3. DIC comparison, parametric spatial models for the Iowa SEER breast cancer data.

Model	Log-relative hazard	p_D	DIC
1'	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + z_{ijk} \zeta + \mathbf{y}_{ijk}^T \boldsymbol{\gamma} + \tau_i z_{ijk} + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk},$ $(\tau_i, \eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \Lambda)$	158.4	179.1
2'	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{T}_{ijk}^T \boldsymbol{\gamma} + \tau_i z_{ijk} + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk},$ $(\tau_i, \eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \Lambda)$	168.7	166.9
3'	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{T}_{ijk}^T \boldsymbol{\gamma} + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk},$ $(\boldsymbol{\eta}_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \Lambda)$	132.9	183.3
4'	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{T}_{ijk}^T \boldsymbol{\gamma}$	36.8	236.3
5'	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{T}_{ijk}^T \boldsymbol{\gamma} + \tau_i z_{ijk} + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk},$ $(\tau_i, \eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha, \Lambda)$	143.8	178.6
6'	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{T}_{ijk}^T \boldsymbol{\gamma} + \tau_i z_{ijk} + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk},$ $(\tau_i, \eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim \text{MCAR}(\alpha = 1, \Lambda)$	162.9	182.5
7'	$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{T}_{ijk}^T \boldsymbol{\gamma} + \tau_i z_{ijk} + \boldsymbol{\eta}_i^T \mathbf{v}_{ijk},$ $(\tau_i, \eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})^T \sim N_5(0, \Lambda)$	215.1	202.4

Next we want to see whether the spatiotemporal interactions in the full Model 2' are significant. That is, in Model 3' we delete the spatially varying coefficient τ_i for year but retain the three-way interaction among age, stage, and year. Model 4' considers only fixed effects with all fixed interaction terms. Finally, in Models 5'–7' we alter the full model by simplifying the distribution on the spatially varying coefficient vector; namely, to an MCAR with a single α , an MCAR with $\alpha = 1$, and an i.i.d. Gaussian. Comparing Model 2' with Models 1', 3', and 4' suggests significant three-way interaction, spatiotemporal interaction, and spatially distributed stage effects. Overall Model 2' has the lowest DIC score, and thus a general MCAR model seems worth fitting to our data set. We thus conclude that Model 2' (the full model in Case I) is the best, and as such, in what follows we present results from this model.

6.2. Results from the Selected Model

In this section, we begin by summarizing our results from the full MCAR Model 2' in Table 3, which includes all interactions among age, stage, and year, spatially varying coefficients η_{1i} , η_{2i} , η_{3i} and η_{4i} for the four stage groups, and spatiotemporal interactions τ_i . Table 4 provides 2.5, 50, and 97.5 posterior percentiles for the fixed effects (components of $\boldsymbol{\beta}$) in this model.

All of these effects except those for the number of primaries, household median income, and county population density covariates have 95% central Bayesian credible intervals that exclude 0. For a given stage, higher age at diagnosis increases the hazard. Interestingly, the rate of this increase is different from one stage to another. For example, the hazard for the distant stage does not increase much with age group, while the hazard for the in situ stage increases dramatically. Specifically, comparing age group IV (>74 years) with age group I (<50 years), the posterior median hazard rate increases by a factor of $e^{1.777 - (-1.128)} = e^{2.905} = 18.26$ for the in situ group. Also, breast cancer stage itself is an important covariate; with “local” as the reference group, women with in situ diagnoses have lower risk, but women with regional and distant diagnoses have higher and much higher risks, respectively. Finally, the increase from in situ to distant becomes less dramatic with increasing age, motivating the more aggressive use of screening methods (mammogram, etc.) at earlier ages.

Table 4. Posterior quantiles for the fixed effects β in the selected MCAR model.

Covariate	2.5%	50%	97.5%
Number of primaries	-0.0517	-0.0244	0.00198
Percentage with high school degree or above	-0.0133	-0.00762	-0.00256
Median household income (10^5 dollars)	-1.262	-0.429	0.271
County population density (persons/square mile)	-0.000124	0.0000774	0.000266
Age <50 years, local stage (baseline)			
Age <50 years, in situ stage	-1.803	-1.128	-0.535
Age <50 years, regional stage	0.678	0.842	0.997
Age <50 years, distant stage	2.117	2.444	2.732
Age 50–64 years, local stage	0.397	0.535	0.670
Age 50–64 years, in situ stage	-0.896	-0.470	-0.0500
Age 50–64 years, regional stage	1.090	1.236	1.365
Age 50–64 years, distant stage	2.239	2.461	2.649
Age 65–74 years, local stage	1.132	1.270	1.401
Age 65–74 years, in situ stage	0.785	1.096	1.467
Age 65–74 years, regional stage	1.545	1.689	1.832
Age 65–74 years, distant stage	2.509	2.707	2.922
Age >74 years, local stage	1.763	1.894	2.033
Age >74 years, in situ stage	1.496	1.777	2.054
Age >74 years, regional stage	2.105	2.248	2.387
Age >74 years, distant stage	2.730	2.975	3.202

Table 5. Posterior quantiles for the fixed effects of year γ in the selected MCAR model.

year effect	2.5%	50%	97.5%
Age <50 years, local stage	-0.0297	-0.0180	-0.00630
Age <50 years, in situ stage	-0.109	-0.0525	-0.00102
Age <50 years, regional stage	-0.0111	-0.00101	0.00846
Age <50 years, distant stage	-0.0251	-0.00352	0.0190
Age 50–64 years, local stage	-0.0466	-0.0382	-0.0300
Age 50–64 years, in situ stage	-0.0575	-0.0257	0.00570
Age 50–64 years, regional stage	-0.0234	-0.0154	-0.00743
Age 50–64 years, distant stage	0.00907	0.0210	0.0346
Age 65–74 years, local stage	-0.0543	-0.0475	-0.0399
Age 65–74 years, in situ stage	-0.0892	-0.0615	-0.0383
Age 65–74 years, regional stage	-0.0358	-0.0277	-0.0194
Age 65–74 years, distant stage	-0.00917	0.00540	0.0205
Age >74 years, local stage	-0.0272	-0.0209	-0.0148
Age >74 years, in situ stage	-0.0595	-0.0392	-0.0197
Age >74 years, regional stage	-0.0269	-0.0195	-0.0118
Age >74 years, distant stage	-0.00662	0.00701	0.0217

Table 5 provides 2.5, 50, and 97.5 posterior percentiles for the fixed effects of year at each age-stage group (components of γ) in this model. For the distant stage group, there is no significant change in survival rate over the diagnosis years for 3 of the 4 age groups (all but age 50–64). For most of the other age-stage groups, the hazard rate decreases significantly with diagnosis year, as we would expect with improve-

ments in breast cancer treatment and screening (especially via mammography). Women diagnosed at the regional stage have their hazard rate decreasing less than women in the local and in situ stage groups. For example, for women aged 65–74 years and the in situ group, the posterior median hazard rate decreases by a factor of $e^{0.0615 \times 21} = 3.64$ from 1973 to 1994. These findings are consistent with the observed survival rates calculated directly from the data set.

The interaction between stage and age group is also apparent from Figure 1, which plots the fitted 5-year survival rates versus diagnosis year for the four stages and four age groups for a typical patient (i.e., having average values for number of primaries and the various county-level covariates, and setting the county-specific random effects to zero). Notice that survival rates have been relatively stable over time in the youngest age group, while in the other age groups they are decreasing for those with the most severe (distant) diagnosis and increasing for the other three stage groups.

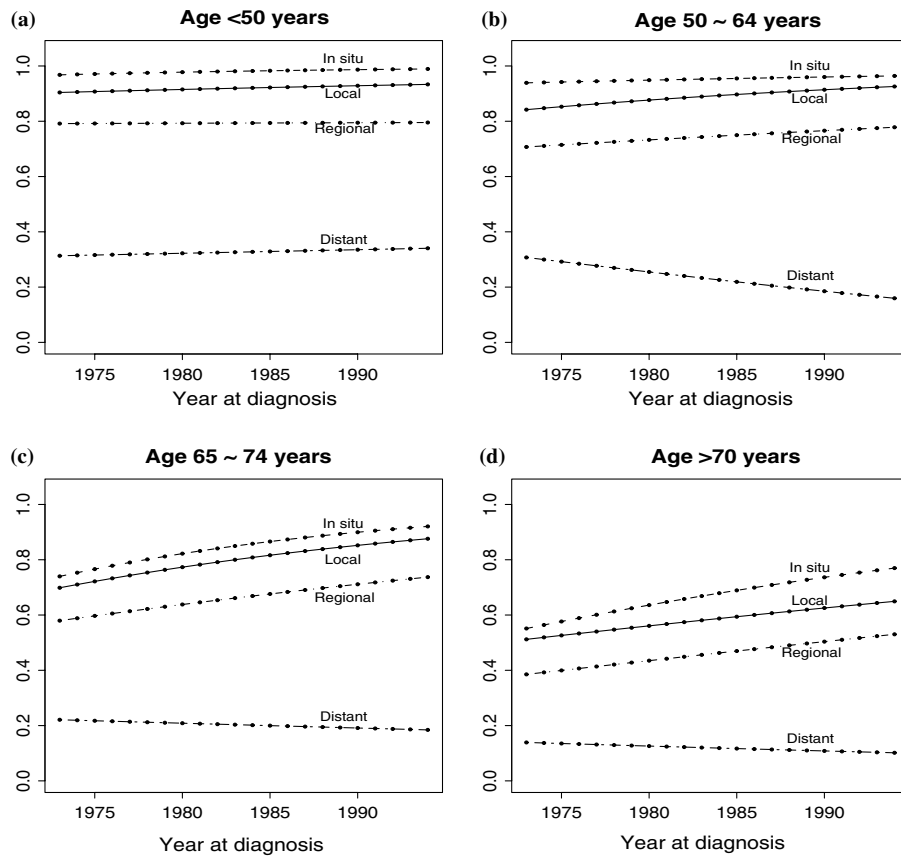


Figure 1. Posterior medians for the 5-year survival rates versus diagnosis year by cancer stage and age at diagnosis for a typical patient, using the chosen MCAR model.

This may reflect an increased likelihood of earlier diagnosis over the study period thanks to more aggressive and improved screening techniques, and a corresponding “antiselection” effect among those diagnosed late (i.e., this group is getting smaller over time, retaining only women having the poorest prognosis).

Turning to geographical summaries, Figure 2 maps the posterior means of the spatially varying coefficient η_{1i} , η_{2i} , η_{3i} , η_{4i} , and τ_i for our chosen MCAR model. For each panel, the Moran’s I statistic (a common measure of spatial association) as well as a p -value for its significance using the usual normal approximation is provided in the caption. Note that significant spatial association (at level 0.05) exists for 3 of the figure’s 5 panels; namely, for effects corresponding to stage II (in situ), stage III (regional), and spatiotemporal interaction.

Visually, the level of spatial pattern is modest, consistent with the estimated α values (roughly 0.4) for our chosen MCAR model. Still, all four of panels (a)–(d)



Figure 2. Maps of posterior means of the spatially varying coefficients chosen MCAR models for the Iowa SEER breast cancer survival data: (a) η_{1i} for local stage (Moran’s $I = 0.095$; $p = 0.107$); (b) η_{2i} for in situ stage (Moran’s $I = 0.132$; $p = 0.029$); (c) η_{3i} for regional stage (Moran’s $I = 0.241$; $p < 0.001$); (d) η_{4i} for distant stage (Moran’s $I = 0.083$; $p = 0.152$); and (e) τ_i for the spatiotemporal interaction (Moran’s $I = 0.251$; $p < 0.001$).

indicate an area of higher fitted values in the northwest portion of the state. Interestingly, panel (e) seems to show a contrasting pattern, with primarily lower values in the northwest region. This may suggest some deficits in access to cancer screening or care in this part of the state, and seems to merit further investigation. We also note that many of the effects in all five panels seem practically significant, in the sense that their magnitudes are reflective of potentially meaningful differences in relative risk. For instance, a county with a fitted value of 0.2 in panel (b) multiplies the hazard by $\exp(0.2) \approx 1.22$, i.e., a 22% increase in the hazard rate from baseline for patients diagnosed at the in situ stage. The fitted values seem trivially small in panel (e) until we recall they must be multiplied by the time covariate (in this case, an integer between 0 and 21).

A corresponding set of 5 maps (not shown) for the corresponding i.i.d. model (Model 7' in Table 3) shows less spatial pattern, with no significant Moran statistics. This is unsurprising, since this prior is not expecting spatial shrinkage. This is also consistent with the p_D and DIC values for Models 2' and 7' in Table 3: the significantly larger effective parameter count (p_D) for Model 7' suggests less shrinkage among the random effects.

7. Summary and Future Research

In this paper, we have developed a multivariate spatial parametric survival model for geographically indexed time to event data. Our model uses an MCAR distribution to capture correlations across both geographic regions and the multiple random effects (frailties, shape parameters, and perhaps spatially varying coefficients) for a given region. A temporal extension also permits application to spatial survival data that are collected in cohorts over time. Implemented using MCMC computational techniques in a hierarchical Bayesian fashion, our approach permits borrowing of strength across regions, offering a compromise between the extremes of a stratified analysis (forcing the sometimes sparse data from each region to stand on its own) and a pooled analysis (which unrealistically assumes the regions to be statistically identical). We illustrated our approach with the analysis of a SEER breast cancer data set, challenging due to its high dimension, spatial and temporal aspects, large number of potential covariates, and manifold interactions. We found the approach to be reasonably easy to use, as well as productive of results that should be helpful for statewide cancer control experts and decisionmakers.

Future avenues for research in this area are many. First, we could imagine extending our Case II and III models to incorporate spatiotemporal interactions, as might be necessary to capture general yet evolving geographic patterns of mortality over time. Second, we could switch from a fully parametric to a semiparametric approach, where the baseline hazard is modeled using not a Weibull but a Cox proportional hazards structure, as most commonly used in applied survival analysis. Banerjee and Carlin (2003) provide such methods in the univariate CAR setting, so extension to the MCAR setting should be possible. Relatedly, MCAR models are

needed for simultaneously modeling *multiple* cancers (say, breast, lung, colon, and prostate), for which correlation may be thought of as arising at either the individual or county level. Some notion of stochastic order may need to be employed here, for pairs of cancers we suspect are likely (though not required) to be temporally ordered at the patient level – say, lung cancers that are induced by previous radiation treatment for breast cancer.

Acknowledgments

The work of both authors was supported in part by NIH Grant 2–R01–ES07750. The authors are grateful to Profs. Sudipto Banerjee and Beth Virnig for extensive discussions and assistance with computing issues and the SEER database, all of which were essential to this project’s completion.

References

- H. Akaike, “Information theory and an extension of the maximum likelihood principle,” *2nd International Symposium on Information Theory*, B. N. Petrov and F. Csáki (eds.) Budapest: Akadémiai Kiadó, pp. 267–281, 1973.
- R. M. Assunção, “Space-varying coefficient models for small area data,” *Environmetrics* vol. 14 pp. 453–473, 2003.
- S. Banerjee and B. P. Carlin, “Spatial semiparametric proportional hazards models for analyzing infant mortality rates in Minnesota counties,” *Case Studies in Bayesian Statistics, Volume VI*, C. Gatsonis, et al., (eds.) Springer-Verlag: New York, pp. 137–151, 2002.
- S. Banerjee and B. P. Carlin, “Semiparametric spatiotemporal frailty modeling,” *Environmetrics* vol. 14 pp. 523–535, 2003.
- S. Banerjee, M. M. Wall, and B. P. Carlin, “Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota,” *Biostatistics* vol. 4 pp. 123–142, 2003.
- J. Besag, “Spatial interaction and the statistical analysis of lattice systems (with discussion),” *Journal of Royal Statistical Society Series B* vol. 36 pp. 192–236, 1974.
- J. Besag, P. Green, D. Higdon, and K. Mengersen, “Bayesian computation and stochastic systems (with discussion),” *Statistical Science* vol. 10 pp. 3–66, 1995.
- D. Brinkley, J. L. Haybittle, and M. R. Alderson, “Death certification in cancer of the breast,” *British Medical Journal* vol. 288 pp. 465–467, 1984.
- B. P. Carlin and S. Banerjee, “Hierarchical multivariate CAR models for spatiotemporally correlated survival data (with discussion),” *Bayesian Statistics 7*, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West (eds.) Oxford University Press: Oxford, pp. 45–63, 2003.
- B. P. Carlin and J. S. Hodges, “Hierarchical proportional hazards regression models for highly stratified data,” *Biometrics* vol. 55 pp. 1162–1170, 1999.
- B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis* (Second Edition), Chapman & Hall/CRC Press: Boca Raton FL, 2000.
- A. E. Gelfand and S. K. Ghosh, “Model choice: a minimum posterior predictive loss approach,” *Biometrika* vol. 85 pp. 1–11, 1998.
- A. E. Gelfand, S. K. Sahu, and B. P. Carlin, “Efficient parametrizations for normal linear mixed models,” *Biometrika* vol. 82 pp. 479–488, 1995.

- A. E. Gelfand, S. K. Sahu, and B. P. Carlin, "Efficient parametrizations for generalized linear mixed models (with discussion)," *Bayesian Statistics 5*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds.) Oxford University Press: Oxford, pp. 165–180, 1996.
- A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of American Statistical Association* vol. 85 pp. 398–409, 1990.
- A. E. Gelfand and P. Vounatsou, "Proper multivariate conditional autoregressive models for spatial data analysis," *Biostatistics* vol. 4 pp. 11–25, 2003.
- K. Hemming and J. Shaw, "A parametric dynamic survival model applied to breast cancer survival times," *Journal of Royal Statistical Society Series C (Applied Statistics)* vol. 51 pp. 421–435, 2002.
- J. G. Ibrahim, M.-H. Chen, and D. Sinha, *Bayesian Survival Analysis*, Springer-Verlag: New York, 2001.
- X. Jin, "Multivariate conditional autoregressive models for spatiotemporal disease data," Unpublished PhD dissertation, Division of Biostatistics, University of Minnesota, 2005.
- T. A. Louis, "Discussion of 'Parameter orthogonality and appropriate conditional inference', by D. R. Cox and N. Reid," *Journal of Royal Statistical Society, Series B* vol. 49 pp. 31, 1987.
- K. V. Mardia, "Multi-dimensional multivariate Gaussian Markov random fields with application to image processing," *Journal of Multivariate Analysis* vol. 24 pp. 265–284, 1988.
- G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics* vol. 6 pp. 461–464, 1978.
- D. J. Spiegelhalter, N. Best, B. P. Carlin, and A. van der Linde, "Bayesian measures of model complexity and fit (with discussion)," *Journal of Royal Statistical Society Series B* vol. 64 pp. 583–639, 2002.
- M. M. Wall, "A close look at the spatial structure implied by the CAR and SAR models," *Journal of Statistical Planning and Inference* vol. 121 pp. 311–324, 2004.
- L. A. Waller, B. P. Carlin, H. Xia, and A. E. Gelfand, "Hierarchical spatiotemporal mapping of disease rates," *Journal American Statistical Association* vol. 92 pp. 607–617, 1997.