Biostatistics (2003), **4**, 1, *pp.* 123–142
Printed in Great Britain

Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota

SUDIPTO BANERJEE, MELANIE M. WALL, BRADLEY P. CARLIN*

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

brad@biostat.umn.edu

SUMMARY

The use of survival models involving a random effect or 'frailty' term is becoming more common. Usually the random effects are assumed to represent different clusters, and clusters are assumed to be independent. In this paper, we consider random effects corresponding to clusters that are spatially arranged, such as clinical sites or geographical regions. That is, we might suspect that random effects corresponding to strata in closer proximity to each other might also be similar in magnitude. Such spatial arrangement of the strata can be modeled in several ways, but we group these ways into two general settings: *geostatistical approaches*, where we use the exact geographic 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). We compare our approaches in the context of a dataset on infant mortality in Minnesota counties between 1992 and 1996. Our main substantive goal here is to explain the pattern of infant mortality using important covariates (sex, race, birth weight, age of mother, etc.) while accounting for possible (spatially correlated) differences in hazard among the counties. We use the GIS `ArcView` to map resulting fitted hazard rates, to help search for possible lingering spatial correlation. The DIC criterion (Spiegelhalter *et al.*, Journal of the Royal Statistical Society, Series B 2002, to appear) is used to choose among various competing models. We investigate the quality of fit of our chosen model, and compare its results when used to investigate neonatal versus post-neonatal mortality. We also compare use of our time-to-event outcome survival model with the simpler dichotomous outcome logistic model. Finally, we summarize our findings and suggest directions for future research.

Keywords: Markov chain Monte Carlo methods; Proportional hazards; Random effects model.

1. INTRODUCTION

Survival models have a long history in the biostatistical and medical literature (see e.g. Cox and Oakes (1984)). Very often, time-to-event data will be grouped into *strata* (or *clusters*), such as clinical sites, geographic regions, and so on. In this setting, a hierarchical modeling approach using stratum-specific *frailties* is often appropriate. Introduced by Vaupel *et al.* (1979), this is a mixed model with random effects (the frailties) that correspond to a stratum's overall health status.

*To whom correspondence should be addressed

To illustrate, let t_{ij} be the time to death or censoring for subject j in stratum i, $j = 1, \ldots, n_i$, $i = 1, \ldots, I$. Let $\mathbf{x}_{ij}$ be a vector of individual-specific covariates. The usual assumption of proportional hazards $h(t_{ij}; \mathbf{x}_{ij})$ enables models of the form

$$h(t_{ij}; \mathbf{x}_{ij}) = h_0(t_{ij}) \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij}) \ , \tag{1}$$

where h_0 is the *baseline hazard*, which is affected only multiplicatively by the exponential term involving the covariates. In the frailty setting, model (1) is extended to

$$\begin{aligned} h(t_{ij}; x_{ij}) &= h_0(t_{ij})\, \omega_i \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij}) \\ &= h_0(t_{ij}) \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij} + W_i) \ , \end{aligned} \tag{2}$$

where $W_i \equiv \log \omega_i$ is the stratum-specific frailty term, designed to capture differences among the strata. Typically a simple i.i.d. specification for the W_i is assumed, e.g.

$$W_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2) \ . \tag{3}$$

At $\sigma^2 = 0$, model (2) reduces to model (1); in practice, σ^2 (like $\boldsymbol{\beta}$ and h_0) is often estimated from the data. We remark that nonnormal (e.g. power variance function family) distributions are often used to model frailties; see e.g. Hougaard (2000, Chapter 7). However, the normal distribution facilitates modeling correlation structure between them, as we shall describe in Section 2.

With the advent of Markov chain Monte Carlo (MCMC) computational methods, the Bayesian approach to fitting hierarchical frailty models such as these has become increasingly popular (see e.g. Carlin and Louis (2000), Section 7.6). The simplest approach perhaps is to assume a parametric form for the baseline hazard h_0. While a variety of choices (gamma, lognormal, etc.) have been explored in the literature, in this paper we adopt the Weibull, which seems to represent a good tradeoff between simplicity and flexibility. This then produces

$$h(t_{ij}; x_{ij}) = \rho t_{ij}^{\rho - 1} \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij} + W_i). \tag{4}$$

Now, placing prior distributions on ρ, $\boldsymbol{\beta}$, and σ^2 completes the Bayesian model specification. Such models are by now a standard part of the literature, and easily fitted using the `WinBUGS` software (Spiegelhalter *et al.*, 1995a,b). Carlin and Hodges (1999) consider further extending model (4) to allow stratum-specific baseline hazards, i.e. by replacing ρ by ρ_i. MCMC fitting is again routine given a distribution for these new random effects – say, $\rho_i \overset{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha, 1/\alpha)$, so that the ρ_i have mean 1 (corresponding to a constant hazard over time) but variance $1/\alpha$.

In this paper, we consider hierarchical survival models for datasets which are *spatially* arranged. That is, we might suspect that frailties W_i corresponding to strata in closer proximity to each other might also be similar in magnitude. This could arise if, say, the strata corresponded to hospitals in a given region, to counties in a given state, and so on. Such spatial arrangement of the strata can be modeled in several ways, but we group these ways into two general settings: *geostatistical approaches*, where we use the exact geographic 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).

The remainder of our paper is organized as follows. In Section 2 we lay out our various approaches for spatial frailty modeling, discussing possible advantages and disadvantages of each. Implementational details for Bayesian computing and model selection are also discussed. Section 3 then applies these models to a challenging dataset on infant mortality in Minnesota counties between 1992 and 1996. Our main substantive goal is to explain the pattern of infant mortality using important covariates (sex, race,

birth weight, mother's age, etc.) while accounting for possible (spatially correlated) differences in hazard among the counties. The GIS `ArcView` is used to create maps of the resulting fitted hazard rates, which in turn are useful for detecting any lingering spatial correlation. After using the DIC criterion to compare models, we investigate the quality of fit of our chosen model, and compare results from separate analyses of neonatal and post-neonatal mortality. In Section 4 we investigate the relationship between our spatial frailty approach and a simpler spatial logistic regression model. Finally, Section 5 summarizes our findings and offers directions for future research in this area.

2. Spatial frailty modeling

2.1 *Geostatistical models*

In this section we adopt the traditional approach to modeling spatial association among observations at a fixed set of spatial locations, referred to by Cressie (1993) as *geostatistical modeling*. Such models assume that the random process of interest $Y(s)$ is indexed continuously by s throughout a space D representing the geographic region being studied. Such models are often used to predict the unobserved value $Y(\mathbf{t})$ at some *target* location $\mathbf{t}$, given observations $\mathbf{Y} \equiv \{Y(\mathbf{s}_i)\}$ at known *source* locations $\mathbf{s}_i,\ i = 1, \ldots, I$. A common model is the Gaussian (normal) one, wherein we might assume

$$\mathbf{Y} \mid \mu, \boldsymbol{\theta} \sim N_I(\mu, H(\boldsymbol{\theta})) \ , \tag{5}$$

where N_I denotes the I-dimensional normal distribution, μ is the (stationary) mean level, and $(H(\boldsymbol{\theta}))_{ii'}$ gives the covariance between $Y(\mathbf{s}_i)$ and $Y(\mathbf{s}_{i'})$. The simplest form for H is an *isotropic* one, where we assume spatial correlation to be a function solely of the Euclidean distance $d_{ii'}$ between $\mathbf{s}_i$ and $\mathbf{s}_{i'}$. For example, we might set $\boldsymbol{\theta} = (\sigma^2, \phi)'$ and take the exponential form

$$(H(\boldsymbol{\theta}))_{ii'} = \sigma^2 \exp(-\phi d_{ii'}),\ \sigma^2 > 0,\ \phi > 0 \ . \tag{6}$$

We hasten to add that while this is a simple and intuitive for other choices, such as the powered exponential,

$$(H(\boldsymbol{\theta}))_{ii'} = \sigma^2 \exp(-\phi d_{ii'}^{\kappa}),\ \sigma^2 > 0,\ \phi > 0,\ \kappa \in (0, 2] \ ,$$

the spherical, the Gaussian, and the Matérn (see e.g. Cressie (1993), or Stein (1999)) are also possible. In particular, while the latter requires calculation of a Bessel function, Stein (1999, p. 51) illustrates its ability to capture a broader range of local correlation behavior despite having no more parameters than the powered exponential.

In our spatial survival context, we apply the geostatistical model (5) and (6) to our random frailties W_i. Thus, writing $\mathbf{W} = \{W_i\}$ we replace the i.i.d. assumption for $\mathbf{W}$ in model (4) with the spatial structure

$$\mathbf{W} \mid \boldsymbol{\theta} \sim N_I(\mathbf{0}, H(\boldsymbol{\theta})) \ . \tag{7}$$

Adding prior distributions for ρ, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$ completes a Bayesian specification using (4) and (7).

The isotropic model (6) works well for a broad class of datasets, but may need to be extended for those exhibiting *anisotropy* (i.e. when the strength of spatial association in the frailties is higher along one direction than another—say, due to disease spread that followed a prevailing wind direction). Ecker and Gelfand (1999) describe Bayesian approaches for handling geometric anisotropy in traditional geostatistical models.

2.2 *Lattice models*

In this section we replace the geostatistical frailty distribution that assumes the random process $\mathbf{W}$ is indexed continuously throughout the space D, with a model which assumes that $\mathbf{W}$ is defined only on

discretely indexed regions such that the regions form a partition of the space D. This type of model is often referred to as a *lattice* model, where the partition of the space D is the 'lattice.' Models of this type usually incorporate information about the adjacency of regions rather than any type of continuous distance metric. We will consider

$$\mathbf{W} \mid \lambda \sim \text{CAR}(\lambda) \,, \tag{8}$$

where CAR stands for a *conditionally autoregressive* structure (Besag *et al.*, 1991). The most common form of this prior (Bernardinelli and Montomoli, 1992) has joint distribution proportional to

$$\lambda^{I/2} \exp\left[-\frac{\lambda}{2} \sum_{i \text{ adj } i'} (W_i - W_{i'})^2\right] \propto \lambda^{I/2} \exp\left[-\frac{\lambda}{2} \sum_{i=1}^{I} m_i W_i (W_i - \overline{W}_i)\right] ,$$

where i adj i' denotes that regions i and i' are adjacent, $\overline{W}_i$ is the average of the $W_{i' \neq i}$ that are adjacent to W_i, and m_i is the number of these adjacencies. This CAR prior is a member of the class of *pairwise difference priors* (Besag *et al.*, 1995), which are identified only up to an additive constant. To permit the data to identify an intercept term β_0 in the hazard function (2), we also add the constraint $\sum_{i=1}^{I} W_i = 0$. A consequence of our prior specification is that

$$W_i \mid W_{i' \neq i} \sim N(\overline{W}_i \,,\, 1/(\lambda m_i)) \,. \tag{9}$$

A vague (but proper) gamma hyperprior distribution for λ completes this portion of the model specification.

We remark that it would certainly be possible to include both spatial and non-spatial frailties, as is now common practice in spatial lattice modeling (see e.g. Besag *et al.* (1991)). In our case, this would mean supplementing our spatial frailties W_i with a collection of non-spatial frailties, say $V_i \overset{\text{i.i.d.}}{\sim} N(0, 1/\tau)$. One problem with this approach is that the frailties now become identified only by the prior, and so the proper choice of priors for τ and λ (or $\boldsymbol{\theta}$) becomes problematic (see e.g. Eberly and Carlin (2000)). Another problem is the resultant decrease in algorithm performance wrought by the addition of so many additional, weakly identified parameters. While this latter problem actually thwarted us in our Section 3 data analysis, we mention the possibility here since it may well pay dividends in other spatial frailty settings.

2.3 *Bayesian implementation*

As already mentioned, the models outlined above are straightforwardly implemented in a Bayesian framework using MCMC methods. Suppose we adopt the geostatistical frailty model (7) (the expressions for the CAR frailty model (8) follow similarly). Letting γ_{ij} be a death indicator (0 if alive, 1 if dead), the joint posterior distribution of interest is given by

$$p\,(\boldsymbol{\beta}, \mathbf{W}, \rho, \boldsymbol{\theta} \mid \mathbf{t}, \mathbf{x}, \boldsymbol{\gamma}) \propto L\,(\boldsymbol{\beta}, \mathbf{W}, \rho\,;\, \mathbf{t}, \mathbf{x}, \boldsymbol{\gamma})\; p\,(\mathbf{W}|\boldsymbol{\theta})\; p(\boldsymbol{\beta}) p\,(\rho)\; p(\boldsymbol{\theta}) \,, \tag{10}$$

where the first term in the right-hand side is the Weibull likelihood, the second is the joint distribution of the random frailties, and the remaining terms are prior distributions. In (10), $\mathbf{t} = \{t_{ij}\}$ denotes the collection of times to death, $\mathbf{x} = \{\mathbf{x}_{ij}\}$ the collection of covariate vectors, and $\boldsymbol{\gamma} = \{\gamma_{ij}\}$ the collection of death indicators for all subjects in all strata. Marginal posteriors of interest are obtained from (10) by integrating out any unwanted parameters.

For our investigations, we retain the parametric form of the baseline hazard given in (4). Thus

$$L(\boldsymbol{\beta}, \mathbf{W}, \rho\ ;\ \mathbf{t}, \mathbf{x}, \boldsymbol{\gamma}) \propto \prod_{i=1}^{I}\prod_{j=1}^{n_i}\{\rho t_{ij}^{\rho-1}\exp(\boldsymbol{\beta}^T\mathbf{x}_{ij}+W_i)\}^{\gamma_{ij}}\exp\{-t_{ij}^{\rho}\exp(\boldsymbol{\beta}^T\mathbf{x}_{ij}+W_i)\}. \quad (11)$$

The model specification in the Bayesian setup is completed by assigning prior distributions for $\boldsymbol{\beta}$, ρ, and $\boldsymbol{\theta}$. Typically, a flat (improper uniform) prior is chosen for $\boldsymbol{\beta}$, while vague but proper priors are chosen for ρ and $\boldsymbol{\theta}$—say, a $G(\alpha, 1/\alpha)$ prior for ρ, a $G(a, b)$ prior for ϕ, and an $IG(c, d)$ prior for σ^2, where G and IG denote the gamma and inverse (reciprocal) gamma distributions, respectively.

The Gibbs sampler (Gelfand and Smith, 1990) is used to update the parameters in the model. This requires drawing samples from the full conditional distributions derived from (10). The complexity of the likelihood in (11) precludes closed-form full conditionals, but the parameters may be updated conveniently using Metropolis–Hastings substeps (Carlin and Louis, 2000, Section 5.4.4). The regression parameters may also be updated using the adaptive rejection sampling algorithm of Gilks and Wild (1992), making use of the log-concavity of their full conditionals.

2.4 *Bayesian model choice*

Bayesian comparison of a model M_1 versus another M_2 has historically been accomplished using the *Bayes factor*. 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. Recently, Spiegelhalter *et al.* (2002) have provided a simple and intuitively appealing extension of the AIC criterion called the *deviance information criterion*, or DIC. This criterion is based on the posterior distribution of the *deviance* statistic,

$$D(\boldsymbol{\theta}) = -2\log f(\mathbf{y}|\boldsymbol{\theta}) + 2\log h(\mathbf{y})\ , \quad (12)$$

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). In this approach, the *fit* of a model is summarized by the posterior expectation of the deviance, $\overline{D} = E_{\theta|y}[D]$, while the *complexity* of a model is captured by the effective number of parameters, p_D. Spiegelhalter *et al.* (2002) show that a reasonable definition of p_D is

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

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 (in our case, the W_i). The DIC is then defined analogously to the AIC as the expected deviance plus the effective number of parameters, i.e.

$$DIC = \overline{D} + p_D\ .$$

Since small values of $\overline{D}$ 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; the

choice of standardizing function $h(\mathbf{y})$ in (12) is arbitrary. Thus, values of DIC have no intrinsic meaning; only *differences* in DIC across models are meaningful.

Besides its generality, an attractive aspect of DIC is that it may be readily calculated during an MCMC run by monitoring both $\boldsymbol{\theta}$ and $D(\boldsymbol{\theta})$, and at the end of the run simply taking the sample mean of the simulated values of D, minus the plug-in estimate of the deviance using the sample means of the simulated values of $\boldsymbol{\theta}$. This quantity can be calculated for each model being considered without analytic adaptation, complicated loss functions, additional MCMC sampling (say, of predictive values), or any matrix inversion. For these reasons, we adopt it as our informal model choice criterion in the sequel.

3. APPLICATION TO MINNESOTA INFANT MORTALITY

3.1 *Model fitting*

We apply the methodology above to the analysis of infant mortality in Minnesota. The data were obtained from the linked birth–death records data registry kept by the Minnesota Department of Health. The data comprise 267 646 live births occurring during the years 1992–1996 followed through the first year of life, together with relevant covariate information such as birth weight, sex, race, mother's age, and the mother's total number of previous births. Because of the careful linkage connecting infant death certificates with birth certificates (even when the death occurs in a separate state), we assume that each baby in the data set that is not linked with a death must have been alive at the end of one year. Of the live births only 1547 babies died before the end of their first year. The number of days they lived is treated as the response t_{ij} in our models, while the remaining survivors were treated as 'censored', or in other words, alive at the end of the study period. In addition to this information, the mother's Minnesota county of residence prior to the birth is provided. The contiguous county neighbor structure as well as the latitude and longitude of the centroids of these counties are available from `ArcView`. This information enabled us to implement both the geostatistical and the lattice models discussed in Section 2. In addition, we investigate the non-spatial frailty model (3), as well as a simple nonhierarchical ('no frailty') model, which simply sets $W_i = 0$ for all i.

For all of our models, we adopt a flat prior for $\boldsymbol{\beta}$, and a $G(\alpha, 1/\alpha)$ prior for ρ which we make vague by setting $\alpha = 0.01$. Metropolis random walk steps with Gaussian proposals were used for sampling from the full conditionals for $\boldsymbol{\beta}$, while Hastings independence steps with gamma proposals were used for updating ρ.

For the geostatistical modeling of the W_i (Section 2.1), we use the isotropic exponential correlation function (6). The inter-county distances (i.e. the $d_{ii'}$ in (6)) are computed using the coordinates of the centroids of the counties. For the exponential correlation function, the quantity $3/\phi$ may be thought of as a measure of the effective isotropic range, i.e. the distance beyond which the correlation between the observations drops to less than 0.05 (Ecker and Gelfand, 1997). Here we adopt a vague $IG(2, 0.01)$ prior for σ^2, ensuring a mean of 100 but infinite variance. For ϕ we take a vague $G(0.01, 100)$ prior, having mean 1 but variance 100.

For the lattice model of Section 2.2, we use the CAR distribution for the W_i. Here we require a prior for the smoothness parameter λ. Note that this prior will implicitly determine the prior variability of the random frailties W_i (see e.g. (9)), but that it will do so differently for each county, depending on the number of adjacencies m_i. Moreover, (9) is a *conditional* specification, in contrast to the marginal specifications in (3) and (7), further complicating the selection of a prior 'comparable' to those already selected for σ^2 and ϕ. While many authors (Bernardinelli *et al.*, 1995; Best *et al.*, 1999; Eberly and Carlin, 2000) have studied this comparability issue, in our case we are fortunate to have a dataset that is large relative to the number of random effects to be estimated. As such, we simply select a vague $G(0.001, 1000)$ (mean 1, variance 1000) specification for λ, and rely on the data to overwhelm the priors.

Table 1. *DIC and effective number of parameters p_D for the competing models*

Model	p_D	DIC
No frailty	8.72	511
Non-spatial frailty	39.35	392
CAR frailty	34.52	371
Geostat frailty	35.02	360

For each of the above models, we ran five initially overdispersed parallel MCMC chains, and monitored them using measurements of sample autocorrelations within the chains, cross-correlations between the parameters, and plots of the sample traces. For the nonhierarchical and non-spatial frailty models, fairly rapid mixing and convergence to the stationary distribution was observed within 10 000 iterations. For the spatial models, however, convergence was much slower. Here our diagnostic tools suggested discarding the first 20 000 iterations from each chain as pre-convergence burn-in. Retaining every 10th of the remaining $5 \times 10\,000 = 50\,000$ iterations yielded a final sample of size 5000 for posterior analysis. We also note that the geostatistical model took much longer to run than the other two. This can be attributed to the computations involving matrix inversions and determinant evaluations of dimension 87×87 within each iteration of the sampler (there are 87 counties in Minnesota). The LU decomposition algorithm (Golub and Van Loan, 1996) was used for this purpose. The CAR model, on the other hand, avoids this problem since it directly models the weight (inverse dispersion) matrix. The very large size of our dataset precluded running our models in `WinBUGS`, so most of our computations were carried out in `Visual C++`. Posterior summarization was accomplished in `S-plus`, while `ArcView` was used for mapping the results.

Table 1 compares our four models in terms of two of the criteria discussed in Section 2.4, DIC and effective model size p_D. For the no-frailty model, we see a p_D of 8.72, very close to the actual number of parameters, 9 (8 components of β plus the Weibull parameter ρ). The other three models have substantially larger p_D values, though much smaller than their actual parameter counts (which would include the 87 random frailties W_i); apparently there is substantial shrinkage of the frailties toward their grand mean. The DIC values suggest that each of these models is substantially better than the no-frailty model, despite their increased size. Of these, the two spatial frailty models have the best DIC values, though plots of the full estimated posterior deviance distributions (not shown) suggest substantial overlap. On the whole we seem to have modest support for the spatial frailty models over the ordinary frailty model.

Tables 2–4 provide 2.5, 50, and 97.5 posterior percentiles for the main effects in our three frailty models, respectively. In all three models, all of the predictors are significant at the 0.05 level. Since the reference group for the sex variable is boys, we see that girls have a lower hazard of death during the first year of life. The reference group for the race variables is white; the native American beta coefficient is rather striking. In the CAR model, this covariate increases the posterior median hazard rate by a factor of $e^{0.782} = 2.19$. The effect of 'unknown' race is also significant, but more difficult to interpret: in this group, the race of the infant was not recorded on the birth certificate. Separate terms for Hispanics, Asians, and Pacific Islanders were also originally included in the model, but were eliminated after emerging as not significantly different from zero. Note that the estimate of ρ is quite similar across models, and suggests a decreasing baseline hazard over time. This is consistent with the fact that a high proportion (495, or 32%) of the infant deaths in our dataset occurred in the first *day* of life: the force of mortality (hazard rate) is very high initially, but drops quickly and continues to decrease throughout the first year.

Evidence of the modest amount of spatial similarity in our dataset is provided by the posterior median for ϕ in the geostatistical model (Table 4); its value of 0.043 implies a median effective spatial range of

Table 2. *Posterior summaries for the nonspatial frailty model*

Covariate	2.5%	50%	97.5%
Intercept	−2.135	−2.024	−1.976
Sex (boys = 0)			
girls	−0.271	−0.189	−0.105
Race (white = 0)			
black	−0.209	−0.104	−0.003
native American	0.457	0.776	1.004
unknown	0.303	0.871	1.381
Mother's age	−0.005	−0.003	−0.001
Birth weight in kg	−1.820	−1.731	−1.64
Total births	0.064	0.121	0.184
ρ	0.411	0.431	0.480
σ	0.083	0.175	0.298

Table 3. *Posterior summaries for the CAR frailty model*

Covariate	2.5%	50%	97.5%
Intercept	−2.585	−2.461	−2.405
Sex (boys = 0)			
girls	−0.224	−0.183	−0.096
Race (white = 0)			
black	−0.219	−0.105	−0.007
native American	0.455	0.782	0.975
unknown	0.351	0.831	1.165
Mother's age	−0.005	−0.004	−0.003
Birth weight in kg	−1.953	−1.932	−1.898
Total births	0.088	0.119	0.151
ρ	0.470	0.484	0.497
λ	12.62	46.07	100.4

Table 4. *Posterior summaries for the geostatistical frailty model*

Covariate	2.5%	50%	97.5%
Intercept	−2.531	−2.436	−2.394
Sex (boys = 0)			
girls	−0.248	−0.192	−0.104
Race (white = 0)			
black	−0.220	−0.104	−0.009
native American	0.464	0.753	0.962
unknown	0.318	0.807	1.136
Mother's age	−0.005	−0.004	−0.003
Birth weight in kg	−1.970	−1.929	−1.889
Total births	0.082	0.111	0.146
ρ	0.460	0.484	0.497
ϕ	0.011	0.043	0.079
σ	0.080	0.147	0.264

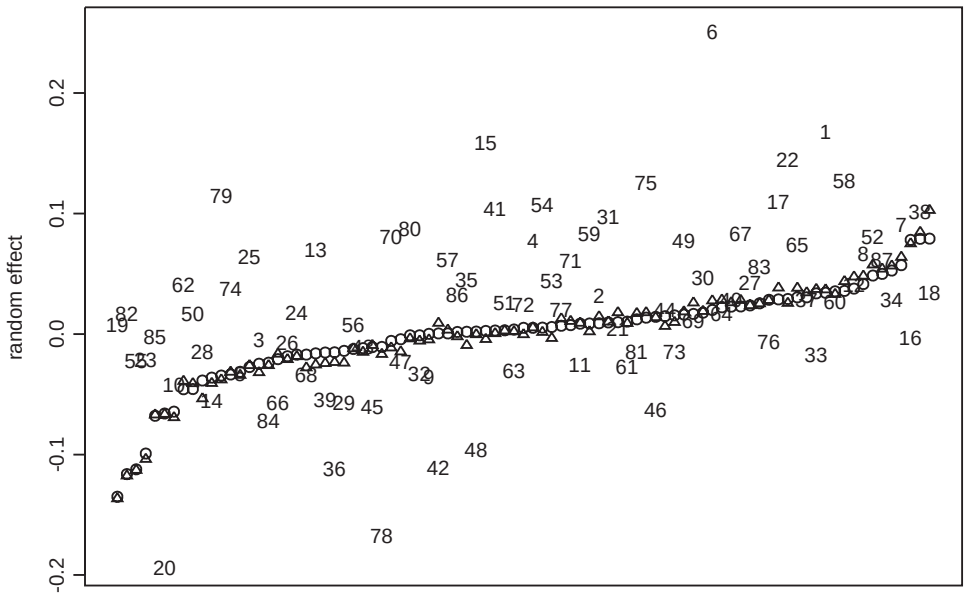

Fig. 1. Plot of posterior median frailties for the CAR (using circles as plotting characters), geostatistical (triangles), and i.i.d. (county numbers) models, with covariates. Counties are sorted horizontally in increasing order of their CAR median frailty.

$3/0.043 \approx 70$ km. This is larger than the typical distance separating centroids of contiguous counties, but far smaller than the maximum distance observed in our dataset (about 700 km). Indeed, this provides some reason why our geostatistical and CAR results should be so similar, since in most cases, borrowing strength from counties having centroids within 70 km will be nearly the same as borrowing strength from adjacent counties. This similarity is quite apparent from Figure 1, which plots the posterior median frailties for our three models. Model-distinct plotting characters are used: circles for CAR, triangles for geostatistical, and the actual county number for non-spatial frailty. Clearly there are many counties having radically different fitted frailties under the spatial and non-spatial models, but the fitted frailties are virtually identical for the two spatial models. Given the geostatistical model's roughly tenfold increase in computer effort, we decided at this point not to consider it further in our analysis.

A benefit of fitting the spatial CAR structure is seen in the reduction of the length of the 95% credible intervals for the covariates in the spatial models compared to the i.i.d. model. As we might expect, there are modest efficiency gains when the model that better specifies the covariance structure of its random effects is used. That is, since the spatial dependence priors for the frailties are in better agreement with the likelihood than is the independence prior, the prior-to-posterior learning afforded by Bayes' rule leads to smaller posterior variances in the former cases. Most notably, the 95% credible set for the effect of 'unknown' race is (0.303, 1.381) under the non-spatial frailty model (Table 2), but (0.351, 1.165) under the CAR frailty model (Table 3), a reduction in length of roughly 25%.

The solid curves in Figure 2 demonstrate the prior-to-posterior learning for two parameters (ρ and λ) in our spatial CAR model, and two parameters (ϕ and σ) in our geostatistical model. Note that in all four cases, the priors are barely visible as nearly flat lines just above the horizontal axes, showing the relative vagueness of the priors and the high degree of Bayesian learning. While this is itself suggestive of a high degree of robustness to the prior, we checked this further by rerunning these two models under a prior that

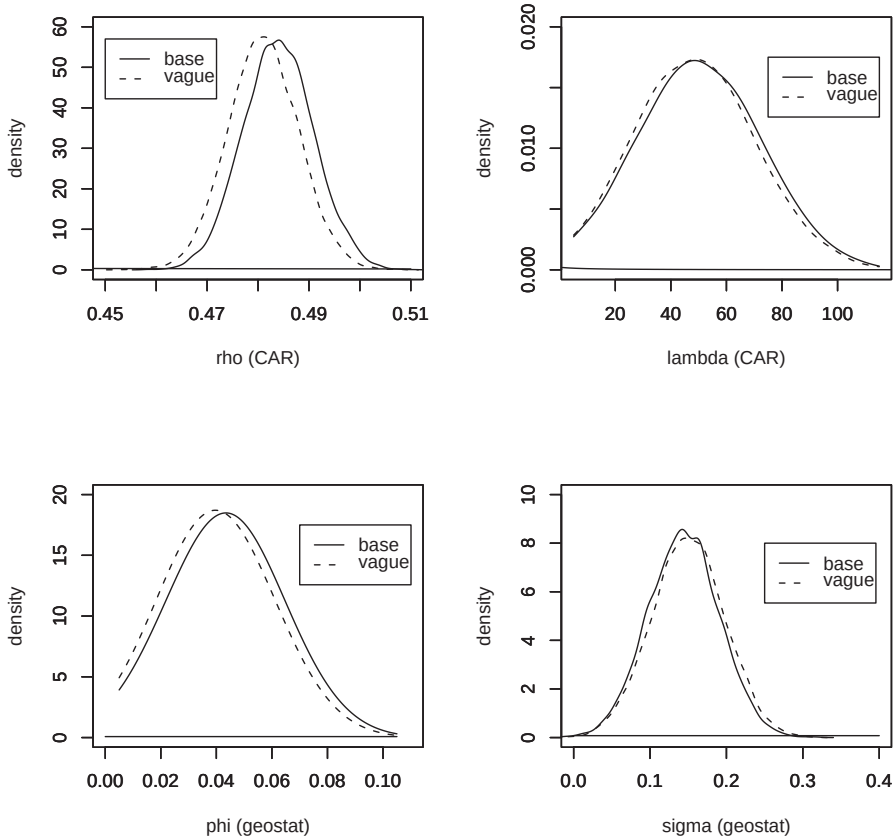

Fig. 2. Prior to posterior learning under base (solid) and 'vague' (dashed) priors.

is even more vague than the 'base' prior described above; namely $\rho \sim G(0.001, 1000)$ (mean still 1 but variance now 1000), $\lambda \sim G(0.0001, 10, 000)$ (mean still 1 but variance now 10,000), $\phi \sim G(0.001, 1000)$ (mean still 1 but variance now 1000), and $\sigma^2 \sim IG(2, 0.001)$ (mean now 1000; variance still infinite). The dashed curves in Figure 2 show little change in the marginal posteriors under these new priors. While this confirms only the stability of our results under even less informative priors than ours, the only 'genuine' prior information available for this problem comes in the form of restrictions to known ranges (say, the known maximal intercentroidal distance in the Minnesota map). In any case, in this analysis we prefer to avoid overly informative priors where possible.

3.2 *Mapping summaries*

We now proceed to mapping summaries of our results. First, to further motivate inclusion of our covariates, Figures 3 and 4 map the posterior medians of the W_i under the non-spatial (i.i.d. frailties) and CAR models, respectively, in the case where *no* covariates **x** are included in the model. The fitted i.i.d. model indicates excess mortality in the north, which is accentuated and extended to a generally increasing pattern from south to north by the CAR model. This trend, combined with the clear emergence of the Minneapolis (county 27) and St Paul (county 62) urban area, strongly suggests the need for fitting covariates in our model, most of which vary spatially.

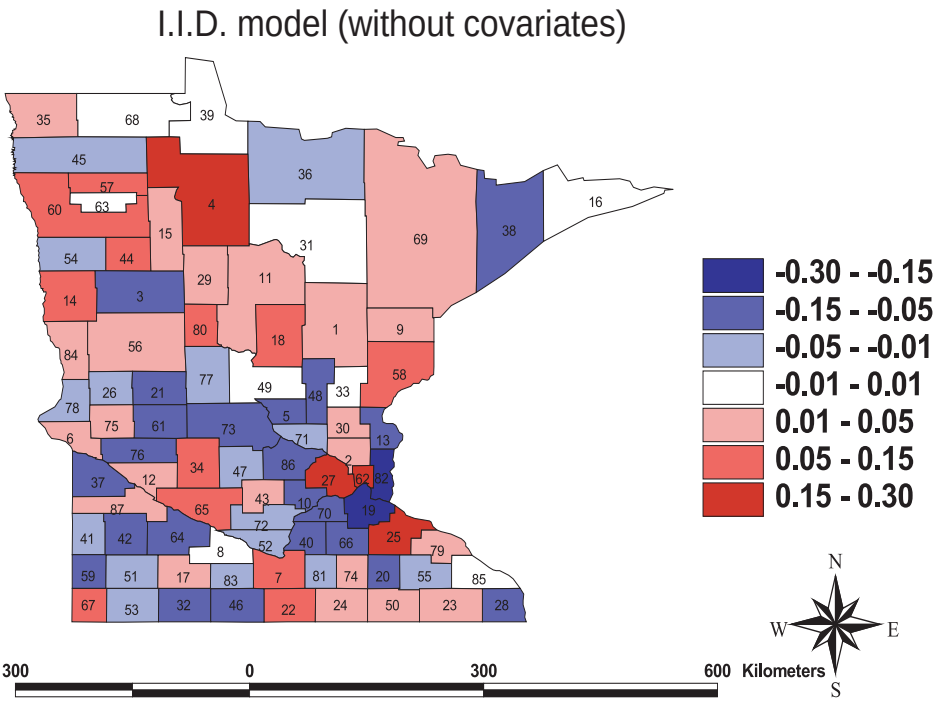

Fig. 3. Posterior median frailties, i.i.d. model without covariates, Minnesota county-level infant mortality data

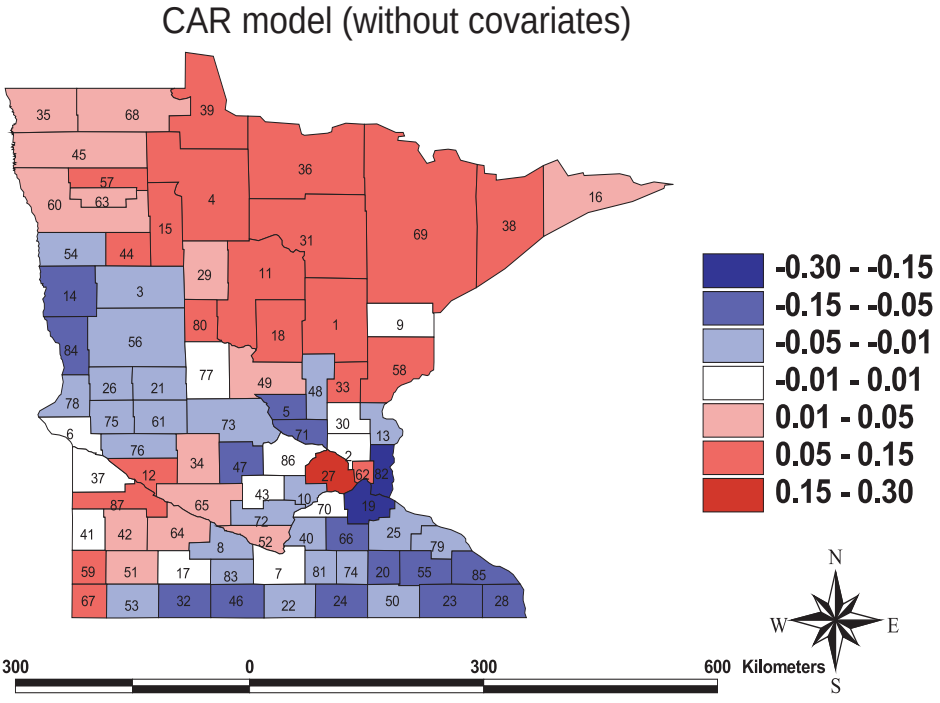

Fig. 4. Posterior median frailties, CAR model without covariates, Minnesota county-level infant mortality data

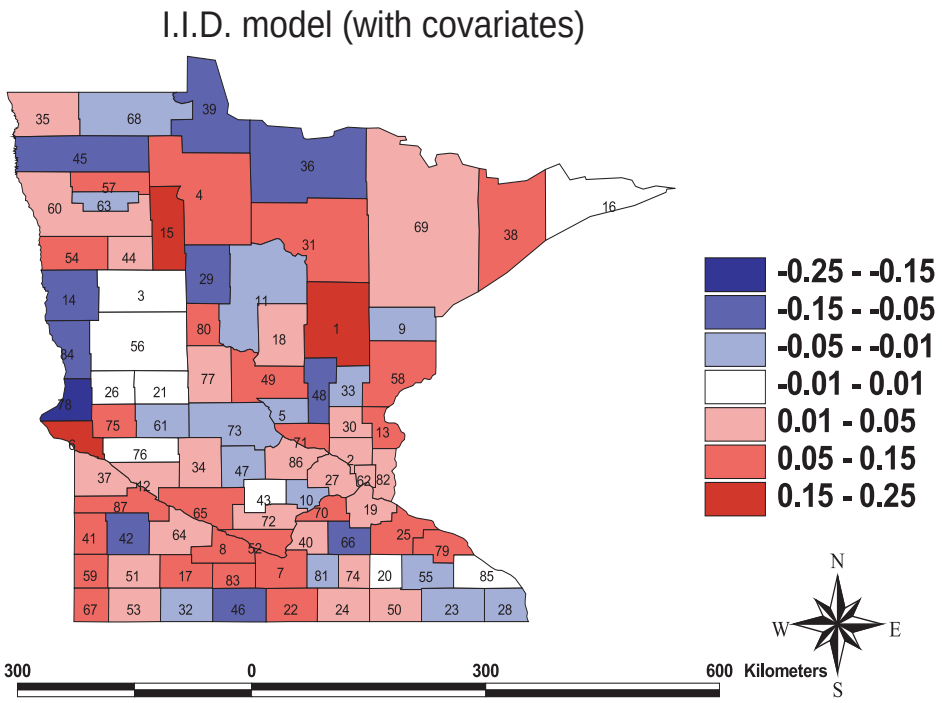

Fig. 5. Posterior median frailties, i.i.d. model with covariates, Minnesota county-level infant mortality data

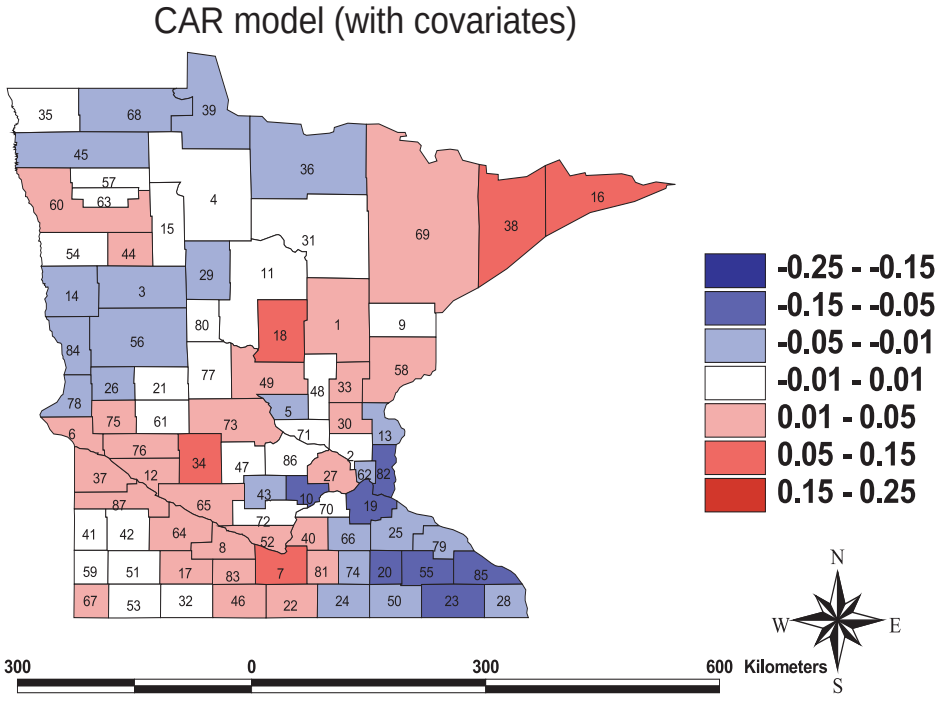

Fig. 6. Posterior median frailties, CAR model with covariates, Minnesota county-level infant mortality data

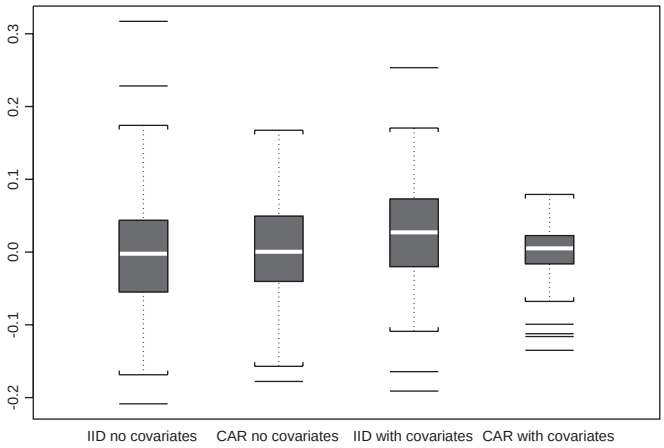

Fig. 7. Boxplots of posterior median frailties, i.i.d. and CAR models with and without covariates.

Figures 5 and 6 then repeat this mapping exercise for the models employing all the covariates listed in Tables 2 and 3. As expected, no clear spatial pattern is evident in the i.i.d. map, but from the CAR map we are able to identify two clusters of counties having somewhat higher hazards (in the southwest following the Minnesota River, and in the northeast 'arrowhead' region), and two clusters with somewhat lower hazards (in the northwest, and the southeastern corner). Thus, despite the significance of the covariates now in these models, Figure 6 suggests the presence of some still-missing, spatially varying covariate(s) relevant for infant mortality. Such covariates might include location of birth (home or hospital), overall quality of available health or hospital care, mother's economic status, and mother's number of prior abortions or miscarriages.

3.3 *Model checking*

In addition to the improved appearance and epidemiological interpretation of Figure 6, another reason to prefer the full (with covariates) CAR model is provided in Figure 7, which shows boxplots of the posterior median frailties for the four cases corresponding to Figures 3–6. The tightness of the full CAR boxplot suggests this model is best at reducing the need for the frailty terms. This is as it should be, since these terms are essentially spatial residuals, and represent lingering lack of fit in our spatial model (although they may well also account for some excess *non*-spatial variability, since our current models do not include non-spatial frailty terms). Note that all of the full CAR residuals are in the range (–0.15, 0.10), or (0.86, 1.11) on the hazard scale, suggesting that missing spatially varying covariates have only a modest (10–15%) impact on the hazard; from a practical standpoint, this model fits quite well.

We now consider several additional plots and diagnostics designed to check the fit of our Weibull-lognormal CAR frailty model with covariates to the observed survival times. First, the solid curve in Figure 8 traces the observed Kaplan–Meier curve (a nonparametric summary of the data), while the dot-dashed curve gives the Kaplan–Meier curve based on a sample from the posterior predictive distribution under the no-frailty model. The dashed curve is similar, except that it is under the CAR frailty model; for this model, 95% equal-tail pointwise credible intervals are also shown as dotted curves. Note that these intervals capture the data-based solid curve beginning near Day 20, whereas more than 100 days elapse before the dot-dashed, no-frailty curve is reliably within the spatial frailty model bounds. Overall, the agreement between the solid (data) and dashed (CAR frailty) curves is quite good, with the latter nearly

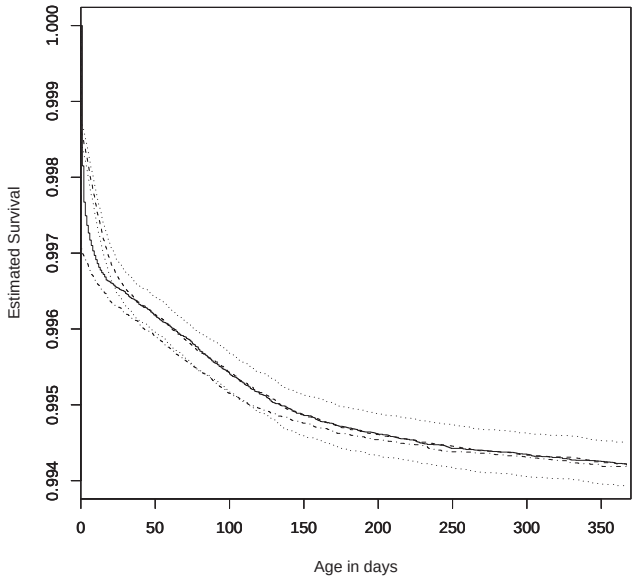

Fig. 8. Kaplan–Meier plots for model checking: solid, as computed from the data itself; dot-dashed, from posterior predictive samples, no-frailty model; dashed, from posterior predictive samples, CAR frailty model (with dotted curves indicating 95% pointwise confidence limits).

perfectly capturing the initial drop during Day 1, as well as the slower decline in the post-neonatal period. The area of disagreement lies in the period between Days 1 and 35. Banerjee and Carlin (2002) show that a semiparametric (Cox) model is required to further improve the fit in this range for these data.

Next, we considered plots (not shown) of Bayesian standardized residuals, defined as $r_{ij} = [t_{ij} - E(t_{ij}|\mathbf{t})]/\sqrt{\text{Var}(t_{ij}|\mathbf{t})}$, and found all to be less than 1 in absolute value. Finally, we computed a Bayesian p-value (Gelman *et al.*, 1995, p. 169) based on the goodness-of-fit statistic $D = \sum_i \sum_j [g(t_{ij}) - E(g(t_{ij})|\boldsymbol{\theta})]^2/\text{Var}(g(t_{ij})|\boldsymbol{\theta})$. This is computed as the empirical proportion of D^* replicates, each arising from a posterior predictive sample, that exceed the actual observed value, D_{obs}. Taking g equal to the identity function and the hazard function, the values obtained (0.264 and 0.312, respectively) also indicate acceptable fit on both scales.

3.4 *Neonatal versus post-neonatal mortality*

Since the risk factors associated with neonatal (death within the first 28 days) and post-neonatal (death during Days 29–365) mortality can be very different, infant mortality cases are frequently divided into these two classes in the epidemiology literature. To check if a similar difference in risk factors would be borne out in our setting, we refit our CAR frailty model to two modifications of our dataset: one considering only neonatal mortality (where all deaths after 28 days are considered censored), and another considering only post-neonatal mortality (where all deaths prior to 28 days are eliminated from the analysis).

In our dataset, 934 (60%) of the 1547 deaths are neonatal; of these, 495 (32% of the total, and 53% of the neonatal deaths) occur during the first day. Tables 5 and 6 give posterior summaries of the CAR frailty model fits to the neonatal and post-neonatal groups separately. Note that sex, birthweight and total births are significant for both groups, while mother's age and native American race are significant only for the

Table 5. *Posterior summaries, CAR frailty model applied to neonatal deaths*

Covariate	2.5%	50%	97.5%
Intercept	−0.378	0.003	0.385
Sex (boys = 0)			
girls	−0.393	−0.264	−0.134
Race (white = 0)			
black	−0.738	−0.464	−0.190
native American	−0.256	0.146	0.549
unknown	0.301	0.899	1.495
Mother's age	−0.011	0.000	0.011
Birth weight in kg	−2.487	−2.501	−2.524
Total births	0.0270	0.074	0.122
ρ	0.215	0.228	0.242
λ	18.58	52.14	109.70

Table 6. *Posterior summaries, CAR frailty model applied to post-neonatal deaths*

Covariate	2.5%	50%	97.5%
Intercept	−6.455	−5.884	−5.314
Sex (boys = 0)			
girls	−0.724	−0.554	−0.385
Race (white = 0)			
black	−0.021	0.277	0.576
native American	0.709	1.068	1.429
unknown	−0.350	0.533	1.415
Mother's age	−0.064	−0.049	−0.032
Birth weight in kg	−1.063	−0.936	−0.808
Total births	0.1597	0.2149	0.2701
ρ	0.7342	0.7923	0.8603
λ	14.45	48.01	104.2

post-neonatal group, and black and unknown race are significant only for the neonatal group. Thus the two groups differ in ways that are both intuitive and substantively intriguing. Children of older mothers do better in Days 28–365, suggesting that this group of children may receive better post-neonatal care. This effect may, however, be somewhat confounded with birth weight, which while significant in both tables plays a more dramatic role in the neonatal group. Native Americans have elevated post-neonatal mortality (possibly due to distance from health care facilities with NICUs), while blacks have elevated neonatal mortality. Previous studies of this area could of course suggest many other potential social, demographic, and economic covariates for investigation. We also see an increase in the fitted ρ value from Table 5 to 6, consistent with the less steep decrease in the survival curve during the post-neonatal period, previously seen in Figure 8. Finally, maps (not shown) of the posterior median frailties for the two groups differ little from the corresponding one for the combined group shown in Figure 6.

4. COMPARISON OF SPATIAL FRAILTY AND LOGISTIC REGRESSION MODELS

We are not aware of previous attempts at considering infant mortality as a time-to-event problem, with or without spatial correlation. In more general contexts (say, a clinical trial enrolling and following patients at spatially proximate clinical centers), a spatial survival model like ours may be the only appropriate model. However, since our dataset does not have any babies censored because of loss to follow-up, competing risks, or any reason other than the end of the study, there is no ambiguity in defining a *binary* survival outcome for use in a random effects logistic regression model. That is, we replace the event time data t_{ij} with an indicator of whether the subject did ($Y_{ij} = 0$) or did not ($Y_{ij} = 1$) survive the first year. Letting $p_{ij} = \Pr(Y_{ij} = 1)$, our model is then

$$\text{logit}(p_{ij}) = \widetilde{\boldsymbol{\beta}}^T \mathbf{x}_{ij} + \widetilde{W}_i \ ,$$

with the usual flat prior for $\widetilde{\boldsymbol{\beta}}$ and an i.i.d. , CAR, or geostatistical prior for the $\widetilde{W}_i$. Other authors (Doksum and Gasko, 1990; Ingram and Kleinman, 1989) have shown that in this case of no censoring before follow-up (and even in cases of equal censoring across groups), it is possible to get results for the $\widetilde{\boldsymbol{\beta}}$ parameters in the logistic regression model very similar to those obtained in the proportional hazards model (1), except of course for the differing interpretations (log odds versus log relative risk, respectively). Moreover, when the probability of death is very small, as it is in the case of infant mortality, the log odds and log relative risk become even more similar. Since it uses more information (i.e. time to death rather than just a survival indicator), intuitively, the proportional hazards model should make gains over the logistic model in terms of power to detect significant covariate effects. Yet, consistent with the simulation studies performed by (Ingram and Kleinman, 1989), our experience with the infant mortality data indicate that only a marginal increase in efficiency (decrease in variance) is exhibited by the posterior distributions of the parameters.

On the other hand, we did find some difference in terms of the estimated random effects in the logistic model compared to the proportional hazards model. Figure 9 shows a scatterplot of the estimated posterior medians of W_i versus $\widetilde{W}_i$ for each county obtained from the models where there were no covariates, and the random effects were assumed to i.i.d. The sample correlation of these estimated random effects is 0.81, clearly indicating that they are quite similar. Yet there are still some particular counties that result in rather different values under the two models. One way to explain this difference is that the hazard functions are not exactly proportional across the 87 counties of Minnesota. A close examination of the counties that had differing $\widetilde{W}_i$ versus W_i shows that they had different average times at death compared to other counties with similar overall death rates. Consider for example County 70, an outlier circled in Figure 9, and its comparison to circled Counties 73, 55, and 2, which have similar death rates (and hence roughly the same horizontal position in Figure 9). We find County 70 has the smallest mean age at death, implying that it has more early deaths, explaining its smaller frailty estimate. Conversely, County 14 has a higher average time at death but overall death rates similar to Counties 82, 48, and 5 (again note the horizontal alignment in Figure 9), and as a result has higher estimated frailty. A lack of proportionality in the baseline hazard rates across counties thus appears to manifest as a departure from linearity in Figure 9.

The detection and implications of non-proportional hazards for the frailty terms has received little attention in the literature to date. In this simple case where logistic regression is also an appropriate model for the data, this comparison of estimated random effects appears to offer a useful diagnostic tool.

5. CONCLUDING REMARKS AND FUTURE DIRECTIONS

In this paper we have described several hierarchical approaches to frailty modeling for spatially correlated survival data, and illustrated the methods with a dataset on infant deaths by county in Minnesota. Our decision to adopt Bayesian methods, implemented using MCMC computational algorithms, enables

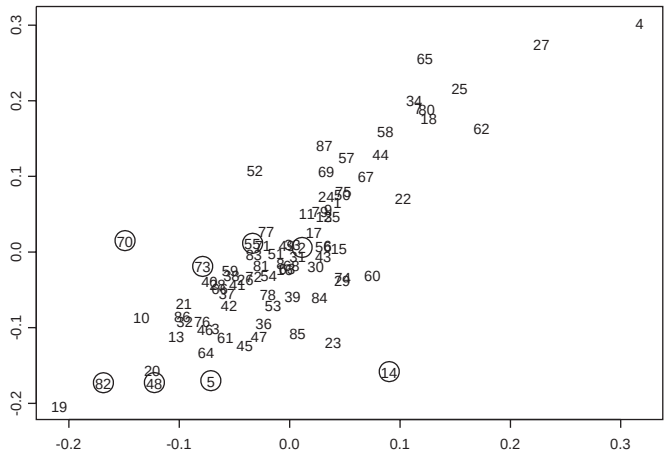

Fig. 9. Posterior medians of the frailties W_i (horizontal axis) versus posterior medians of the logistic random effects $\widetilde{W}_i$ (vertical axis). Plotting character is county number; significance of circled counties is described in the text.

full posterior inference on the resulting main effects and county-level frailties, and without resort to model simplifications or approximations. Our approach appears to pay dividends in the specific application. In particular, the GIS maps of the smoothed frailties resulting from our models suggest overall trends in infant mortality, as well as potential regions to target for further investigation (say, to uncover an as yet unidentified spatially varying covariate) or public health intervention efforts. We have also clarified a possible benefit to using spatial frailty models instead of simpler spatial logistic regression models (when applicable) in spatially oriented survival analyses, and suggested a related diagnostic for departures from proportional hazards.

The CAR prior, when implemented with the common 0–1 adjacency weighting we have used, is well known to have awkward theoretical properties due to noninvertibility and edge effects; see e.g. Wall (2000). That is, the adjacency weights implicitly determine a precision matrix that does not correspond to a proper covariance structure H, as in (7); see e.g. Best *et al.* (1999) and Carlin (1999) for further discussion of this point. The CAR model can be made invertible by using weights satisfying a certain condition (Besag and Kooperberg, 1995; Conlon and Waller, 1999) investigate the use of such a covariance specification based on distance, and plot the induced CAR weights versus distance in an attempt to calibrate these two approaches. However, the calibration is difficult, and in any case the approach forfeits the computational simplicity of the weight-specification approach, since it now demands inversion of the (typically large) H at each iteration, just as in the geostatistical approach. Since in our experience the smoothed frailty maps are similar under the two procedures for comparable prior specifications, we tend to favor the computationally simpler (weight-based) CAR approach.

Previous work by Carlin and Hodges (1999) suggests a generalization of our basic model (4) to

$$h(t_{ij}; x_{ij}) = \rho_i t_{ij}^{\rho_i - 1} \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij} + W_i) \; .$$

That is, we allow two sets of random effects: the existing frailty parameters W_i, and a new set of shape parameters ρ_i. This then allows both the overall level and the shape of the hazard function over time to vary from county to county. Either i.i.d. or CAR priors could be assigned to these two sets of random effects, which could themselves be correlated within county. In the latter case, this might be fit using the so-called '2-fold CAR' model (Kim *et al.*, 2001).

Finally, one might wonder about nonparametric alternatives to our parametric (Weibull) baseline hazard function. Clayton (1994) formulates the Cox model using counting process notation, and shows how to estimate the baseline hazard and regression parameters using MCMC methods. While somewhat contrived, frailty terms may be incorporated in this approach, and so could be exploited in our spatial frailty setting. An alternative would be to use the mixture of monotone functions approach of Gelfand and Mallick (1995), as illustrated for stratified data by Carlin and Hodges (1999); here again, the strata become geographic regions, connected via a spatially correlated mixing distribution. We hope to explore these and other extensions in a future manuscript.

ACKNOWLDEGEMENTS

Sudipto Banerjee and Melanie M. Wall are Assistant Professors and Bradley P. Carlin is Professor in the Division of Biostatistics, School of Public Health at the University of Minnesota, Minneapolis, MN, 55455. The work of the first and third authors was supported in part by NSF/EPA grant SES 99-78238, while that of the second author was supported in part by National Center for Health Statistics grant NCHS-UR6/CCU517477-01. The authors are grateful to the Minnesota Center for Excellence in Health Statistics for providing and permitting analysis of the Minnesota infant mortality dataset.

REFERENCES

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Csáki, F. (eds), *2nd International Symposium on Information Theory*, Budapest: Akadémiai Kiadó, pp. 267–281.

BANERJEE, S. AND CARLIN, B. P. (2002). Spatial semiparametric proportional hazards models for analyzing infant mortality rates in Minnesota counties. In Gatsonis, C., Carniguity, A. L., Gelman, A., Higdon, D., Kass, R. E., Paulen, D. and Verinelli, I. (eds), To appear in *Case Studies in Bayesian Statistics*, Volume VI. New York: Springer.

BERNARDINELLI, L., CLAYTON, D. G. AND MONTOMOLI, C. (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine* **14**, 2411–2431.

BERNARDINELLI, L. AND MONTOMOLI, C. (1992). Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in Medicine* **11**, 983–1007.

BESAG, J., GREEN, P., HIGDON, D. AND MENGERSEN, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* **10**, 3–66.

BESAG, J. AND KOOPERBERG, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* **82**, 733–746.

BESAG, J., YORK, J. C. AND MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann. Inst. Statist. Math.* **43**, 1–59.

BEST, N. G., ARNOLD, R. A., THOMAS, A., WALLER, L. A. AND CONLON, E. M. Bayesian models for spatially correlated disease and exposure data (with discussion). In Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds), *Bayesian Statistics 6*, Oxford: Oxford University Press, **1999**, pp. 131–156.

CARLIN, B. P. (1999). Discussion of 'Bayesian models for spatially correlated disease and exposure data,' by N. G. Best, L. A. Waller, A. Thomas, E. M. Conlon, and R. A. Arnold. In Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds), *Bayesian Statistics 6*, Oxford: Oxford University Press, pp. 147–150.

CARLIN, B. P. AND HODGES, J. S. (1999). Hierarchical proportional hazards regression models for highly stratified data. *Biometrics* **55**, 1162–1170.

CARLIN, B. P. AND LOUIS, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd edition. Boca Raton, FL: Chapman and Hall/CRC Press.

CLAYTON, D. (1994). Some approaches to the analysis of recurrent event data. *Statistics in Medical Research* **3**, 244–262.

CONLON, E. M. AND WALLER, L. A. (1999). Flexible spatial hierarchical models for mapping disease rates. *Proceedings of A.S.A. Section on Statistics and the Environment*, Washington, DC: American Statistical Association, pp. 82–87.

COX, D. R. AND OAKES, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*, revised edition. New York: Wiley.

DOKSUM, K. A. AND GASKO, M. (1990). On a correspondence between models in binary regression analysis and in survival analysis. *International Statistical Review* **58**, 243–252.

EBERLY, L. E. AND CARLIN, B. P. (2000). Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statistics in Medicine* **19**, 2279–2294.

ECKER, M. D. AND GELFAND, A. E. (1997). Bayesian variogram modeling for an isotropic spatial process. *J. Agr. Biol. Env. Statist.* **2**, 347–369.

ECKER, M. D. AND GELFAND, A. E. (1999). Bayesian modeling and inference for geometrically anisotropic spatial data. *Mathematical Geology* **31**, 67–83.

GELFAND, A. E. AND MALLICK, B. K. (1995). Bayesian analysis of proportional hazards models built from monotone functions. *Biometrics* **51**, 843–852.

GELFAND, A. E. AND SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

GELMAN, A., CARLIN, J., STERN, H. AND RUBIN, D. B. (1995). *Bayesian Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC Press.

GILKS, W. R. AND WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society* Series C *(Applied Statistics)* **41**, 337–348.

GOLUB, G. H. AND VAN, LOAN C. F. (1996). *Matrix Computations*, 3rd edition. Baltimore, MD: The Johns Hopkins University Press.

HOUGAARD, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.

INGRAM, D. D. AND KLEINMAN, J. C. (1989). Empirical comparisons of proportional hazards and logistic regression models. *Statistics in Medicine* **8**, 525–538.

KIM, H., SUN, D. AND TSUTAKAWA, R. K. (2001). A bivariate Bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. *Journal of the American Statistical Association* **96**, 1506–1521.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.

SPIEGELHALTER, D. J., BEST, N., CARLIN, B. P. AND VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit (with discussion). To appear *Journal of the Royal Statistical Society*, Series B.

SPIEGELHALTER, D. J., THOMAS, A., BEST, N. AND GILKS, W. R. (1995a). BUGS: Bayesian inference using Gibbs sampling, Version 0.50. Technical Report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University; web address: `www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`.

SPIEGELHALTER, D. J., THOMAS, A., BEST, N. AND GILKS, W. R. (1995b). BUGS examples, Version 0.50. Technical report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University; web address same as preceding reference.

STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.

VAUPEL, J. W., MANTON, K. G. AND STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 439–454.

WALL, M. M. (2000). A close look at the spatial correlation structure implied by the CAR and SAR models. Research Report 2000-022, Division of Biostatistics, University of Minnesota; web address: `ftp.biostat.umn.edu/pub/2000/rr2000-022.pdf`.

[*Received March 28, 2001*; *revised October 8, 2001*; *accepted for publication November 6, 2001*]