Far East Journal of Applied Mathematics
Volume 29, Number 1, October 2007, Pages 1-18
Published online: September 18, 2007
This paper is available online at http://www.pphmj.com

HIERARCHICAL MODELS FOR DETECTING GEOGRAPHICAL EFFECTS IN CANCER INCIDENCE AND SURVIVAL

LUYAN DAI, ZUOQIONG HE, DONGCHU SUN and MARIO SCHOOTMAN*

Department of Statistics
University of Missouri-Columbia
Columbia, Missouri 65201, U. S. A.

*Washington University School of Medicine
St. Louis, Missouri 63110, U. S. A.

Abstract

The study is motivated by the importance of assessing small-area variation for the development and implementation of medial and educational interventions to reduce disparities in breast cancer survival. The Data were collected state-wide and post-stratified to the county level by Iowa SEER program. We propose a Bayesian hierarchical model for Weibull distribution by incorporating conditional autoregressive priors for transformed rate parameters to analyze spatial-temporal effects on breast cancer survival. Gibbs-Poole-Stockmeyer algorithm on the sparse adjacency matrix enhances efficiency of Gibbs sampler in the simulation study. Results of breast cancer survivals for aged 65 or older women in Iowa are presented. Comments and further discussions are given.

1. Introduction

Breast cancer is the second leading cause of cancer death among women in the United States and estimated 41,619 death in 2003. About

2000 Mathematics Subject Classification: 62N01, 62J12.

Keywords and phrases: survival analysis, spatial correlation, Bayesian hierarchical models, conditional autoregressive priors, Weibull distribution, G.P.S. algorithm.

Received April 27, 2007

80% of the breast cancer deaths are among women age 65 years or older [15]. These women have a lower survival rate and are less likely to receive recommended treatment relative to young women [11]. Regardless of age, survival from breast cancer varies geographically in Europe [5, 10]. Among large geographical areas in the United States, 5-year survival rates in this age group varied among Surveillance, Epidemiology, and End Results (SEER) registries from 71% (Iowa) to 80% (Hawaii) [4].

Because cancer survival is believed to be influenced by both community as well as individual characteristics, yet most research focuses on either the individual or his/her environment. Some works were done by other researchers, Goldstein [8] and Reader [13]. Our study is motivated by the importance of assessing small-area variation for the development and implementation of medial and educational interventions to reduce disparities in survival. We used the Iowa SEER database and determine if geographic disparities exist at smaller areas. There are 5918 patients registered excluding dead individuals when starting from 1991 in the record. Also basic information like age, race, survival time, county, census tract and related treatment after breast cancer incidence is included. Thus, a hierarchical Bayesian modeling strategy could be used for the analysis. We adopt a linear mixed model of the logarithm of the rate parameters in a Weibull distribution. Smith [14] pointed out frequentist methods such as maximum likelihood estimates in a Weibull model are notoriously difficult with small sample sizes. Although the total sample size from the Iowa SEER data sets is quite large, the data at the county-by-year-by-age group level have many zeroes. This greatly complicates a conventional Weibull survival model. Later, Berger and Sun [1] used Bayesian analysis for Weibull distributions with competing risk models and stress-strength models. Hierarchical linear mixed models have recently been used to model response times in psychology studies by Lu et al. [12].

In this paper, we propose a Bayesian hierarchical model to capture necessary features of spatial-temporal models related to breast cancer survival. Because age plays a key role in breast cancer survival, it is also incorporated into our model. The spatial effects are modeled with conditionally autoregressive priors in [2] due to the belief of possible

correlations among adjacent geographical sites. The paper is organized as follows: Section 2 introduces the notation and description of data first and then model in detail. Section 3 summarizes simulation and results. Finally, some comments of some current studies are given in Section 4.

2. Data and Models

2.1. The data

We focus on one of SEER programs in Iowa with K counties with $K = 99$. Assume that for each county i, there are n_i subjects under study associated with jth individual in county i. We observed a survival time t_{ij} of breast cancer after diagnosed and a fixed censoring time c_{ij}. So there are total $n = n_1 + \cdots + n_K = 5918$ breast cancer female patients aged 65 or older recorded in the Iowa SEER program from 1992 to 1999. Assume that t_{ij}'s are independent with Weibull $W(\alpha, \lambda_{ij})$ distribution, whose density and survival function are given by

$$f(t \mid \alpha, \lambda_{ij}) = \alpha\lambda_{ij} t^{\alpha-1} \exp(-\lambda_{ij} t^{\alpha}),$$

and

$$S(t \mid \alpha, \lambda_{ij}) = \exp(-\lambda_{ij} t^{\alpha}),\ t > 0,$$

respectively. The exact survival time t_{ij} will be observed only if $t_{ij} < c_{ij}$. Note that the hazard function of t_{ij} is

$$h(t \mid \alpha, \lambda_{ij}) = \alpha\lambda_{ij} t^{\alpha-1},$$

which is a special case of Cox-proportional hazard model [3]. The data in this framework will be represented by the n pairs of random variables (y_{ij}, Δ_{ij}), where $y_{ij} = \min(t_{ij}, c_{ij})$, and $\Delta_{ij} = 1$ if $t_{ij} \le c_{ij}$, and 0, otherwise. Note that if $\Delta_{ij} = 1$, person j in county i died before time c_{ij}, so t_{ij} is the survival time after diagnosis. On the other hand if $\Delta_{ij} = 0$, the person is either still alive or out of investigation at the time c_{ij} after diagnosis.

Alternatively, we could use the type of nonparametric approach for estimating survival functions. Such approach is computationally feasible

in [9] with the type of transformations of the parameters. However, it is rather difficult to study the geographical variation and spatial effect. In our model, we use the parametric approach in details below. As convention, we usually take logarithm of the rate parameters as $V_{ij} = \log(\lambda_{ij})$. The density and survival function can be rewritten as

$$f(t \mid \alpha_{ij}, V_{ij}) = \alpha t^{\alpha-1} \exp[V_{ij} - \exp(V_{ij}) t^{\alpha}].$$

Then, the likelihood function of $\mathbf{V} = (V_{11}, ..., V_{Kn_K})$ and α based on $(\mathbf{y}, \mathbf{\Delta})$ is

$$\begin{aligned} & L(\alpha, \mathbf{V} \mid \mathbf{y}, \mathbf{\Delta}) \\ &= \prod_{i=1}^{K} \prod_{j=1}^{n_i} \{[f(y_{ij} \mid \alpha, V_{ij})]^{\Delta_{ij}} \times [S(y_{ij} \mid \alpha, V_{ij})]^{1-\Delta_{ij}}\} \\ &= \prod_{i=1}^{K} \prod_{j=1}^{n_i} \{\exp(\Delta_{ij}[\log(\alpha) + V_{ij} + \alpha \log(y_{ij}) - \log(y_{ij})] - e^{V_{ij}} y_{ij}^{\alpha})\} \\ &= \exp\left\{\sum_{i=1}^{K} \sum_{j=1}^{n_i} (\Delta_{ij}[\log(\alpha) + V_{ij} + \alpha \log(y_{ij}) - \log(y_{ij})] - e^{V_{ij}} y_{ij}^{\alpha})\right\}, \end{aligned} \tag{1}$$

where $\mathbf{y} = (y_{11}, ..., y_{Kn_K})$ and $\mathbf{\Delta} = (\Delta_{11}, ..., \Delta_{Kn_K 0})$.

2.2. The linear mixed model

For transformed rates $\mathbf{V}$, we propose to use a linear mixed model to the logarithm of the rate parameters,

$$V_{ij} = \mathbf{x}'_{0,ij}\beta + \mathbf{x}'_{1,ij}\mathbf{W} + \varepsilon_{ij}, \; i = 1, ..., K, \quad j = 1, ..., n_i. \tag{2}$$

Here $\mathbf{x}_{0,ij}$, including constant, is the age of person j in county i, β is the regression coefficients, $\mathbf{x}_{1,ij}$ is vector of indicators for spatial location, and $\mathbf{W}$ represents multilevel spatial effects such as county and census tract effects. ε_{ij} represents any other effects not included in the model. Sometimes, these random noises are interpreted as frailties in survival models. Note that the first row of $\mathbf{x}_{0,ij}$ could be 1 so that constant is included. If we define $\mathbf{X}_q = (\mathbf{x}_{q,11}, ..., \mathbf{x}_{q,Kn_K})'$, for $q = 0, 1$ with error

terms $\boldsymbol{\varepsilon} = (\varepsilon_{11}, ..., \varepsilon_{Kn_K})'$, model (2) can be written in a matrix form,

$$\mathbf{V} = \mathbf{X}_0\beta + \mathbf{X}_1\mathbf{W} + \varepsilon. \tag{3}$$

We assume that the error term ε_{ij}'s are i.i.d. normally distributed with variance δ_0. Then (3) is equivalent to follow a multivariate normal distribution

$$(\mathbf{V} \mid \beta, \mathbf{W}, \delta_0) \sim N_n(\mathbf{X}_0\beta + \mathbf{X}_1\mathbf{W}, \delta_0\mathbf{I}_n). \tag{4}$$

2.3. A CAR model for spatial correlation

To state the prior in the model in general form, we consider spatially correlated county effects $\mathbf{Z} = (Z_1, ..., Z_K)'$. The location of Z_i, corresponding to a county in the proposed model, is assumed to have at least one neighboring location. These locations define a so-called *adjacency matrix* $\mathbf{C}$, and $K \times K$ matrix with elements c_{ik}, where $c_{ik} = 1$ if regions i and k are adjacent, and $c_{ik} = 0$ otherwise. By definition, $c_{ii} = 0$. Let N_i be the *neighborhood* of region i, that is, the set of all regions k for which $c_{ik} = 1$. We assume that the conditional autoregressive model (CAR) defined by the conditional densities

$$f(Z_i \mid Z_k, k \neq i) = \left(\frac{1}{2\pi\delta}\exp\left\{-\frac{1}{2\pi\delta}(Z_i - \rho\overline{Z}_{-i})\right\}\right), \tag{5}$$

where $\delta > 0$, and $\overline{Z}_{-i} = \sum_{k\in N_i} Z_k$.

Correlation parameter ρ measures the dependencies of two adjacent locations. When $\rho = 0$, we know that $\mathbf{Z} = (Z_1, ..., Z_K)'$ are independent. Thus ρ serves as an index of spatial dependence. It is shown that (5) is equivalent to the multivariate normal distribution in [16]

$$(\mathbf{Z} \mid \delta, \rho) \sim N_K(\mathbf{0}, \delta(\mathbf{I} - \rho\mathbf{C})^{-1}), \tag{6}$$

where $\mathbf{I}$ is the identical matrix. In other words, $\mathbf{Z}$ is multivariate normal with mean $\mathbf{0}$ and variance matrix $(\mathbf{I} - \rho\mathbf{C})^{-1}$. To ensure the positivity of

$(\mathbf{I} - \rho\mathbf{C})^{-1}$, ρ is restricted to

$$\left(\frac{1}{\min_i \vartheta_i}, \frac{1}{\max_i \vartheta_i}\right), \tag{7}$$

where $\vartheta_1 \le \cdots \le \vartheta_K$ are eigenvalues of the adjacency matrix $\mathbf{C}$. We will use conditional autoregressive process prior to model the spatially correlated random effects such as $\mathbf{W}$ in (2).

2.4. The prior

Since the shape parameter α in the Weibull model is often quite stable, we put a gamma prior for the shape parameter α,

$$\alpha \sim \text{Gamma}(a_\alpha, b_\alpha). \tag{8}$$

In this paper, a Gamma (a, b) distribution has the density

$$f(t) = \frac{b^a}{\Gamma(a)} t^{a-1} \exp(-bt), \text{ for } t > 0.$$

The prior is chosen because of its flexibility and convenience.

Followed by the previous section, we assume that $\mathbf{W}$ follows a CAR prior,

$$(\mathbf{W} \mid \rho, \delta_1) \sim N_K(\mathbf{0}, \delta_1(\mathbf{I} - \rho\mathbf{C})^{-1}), \tag{9}$$

where K is the number of counties in the SEER program, $\mathbf{C}$ is the adjacent matrix from neighboring spatial sites, for instance, counties, census tracts in the SEER program, δ_1 is the variation parameter and ρ is a correlation parameter among locations. For regression coefficients β, we use a multivariate normal prior, i.e.,

$$\beta \sim N_p(\mathbf{0}, \tau\mathbf{I}). \tag{10}$$

If τ is large, then we have a flat normal prior for β. Note that we assume that β and $\mathbf{W}$ are independent with each other. If we write $\mathbf{X} = (\mathbf{X}_0, \mathbf{X}_1)$ and $\gamma = (\beta', \mathbf{W}')'$, model (2) becomes

$$\mathbf{V} = \mathbf{X}\gamma + \mathbf{å}. \tag{11}$$

Then the priors (9) and (10) with independence assumption are equivalent to prior on γ as

$$\gamma \sim N_{p+K}(\mathbf{0}, \boldsymbol{\delta}), \text{ where } \boldsymbol{\delta} = \begin{pmatrix} \tau\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \delta_1(\mathbf{I} - \rho\mathbf{C})^{-1} \end{pmatrix}. \tag{12}$$

We could easily see that such a model is a hierarchical model

$$(\mathbf{V} \mid \mathbf{W}, \delta_0) \sim N_n(\mathbf{X}_1\mathbf{W}, \delta_0\mathbf{I}_n),$$

$$(\mathbf{W} \mid \beta, \delta_1, \rho) \sim N_K(\mathbf{X}_0\beta, \delta_1(\mathbf{I} - \rho\mathbf{C})^{-1}). \tag{13}$$

To complete the hierarchical model, we still need a prior on the variance parameters δ_0 and δ_1. We choose the prior on δ_0 as inverse type of priors,

$$[\delta_0] \propto \frac{1}{\delta_0^{a_0+1}} \exp(-b_0/\delta_0). \tag{14}$$

For the prior of δ_1, we consider $\eta_1 = \delta_0/\delta_1$, the noise and signal ratio. We propose to use the Pareto (1, 1) prior on η_1, whose density is given by

$$[\eta_1] = \frac{1}{(\eta_1 + 1)^2}. \tag{15}$$

Note that a general case of Pareto (a, b) distribution has the density

$$f(x) = ab^a/x^{a+1}, \; x > 0.$$

(15) is equivalent to assume that

$$[\delta_1 \mid \delta_0] = \frac{\delta_0}{(\delta_1 + \delta_0)^2}.$$

The reasons of using such a prior can be found in [12]. For example, (15) is equivalent to assume that $\frac{\delta_1}{\delta_0 + \delta_1} \sim \text{uniform}\,(0, 1)$.

Because of the restriction of ρ in (7), the prior of spatial correlation parameter ρ needs to be specified. We assume that ρ to be uniformly distributed on the interval of $(1/\vartheta_1, 1/\vartheta_K)$, i.e.,

$$\rho \sim \text{Unif}\left(\frac{1}{\vartheta_1}, \frac{1}{\vartheta_K}\right). \tag{16}$$

If $\mathbf{C}$ is the adjacency matrix for counties in Iowa, we have $1/\vartheta_1 \approx -0.304$ and $1/\vartheta_K \approx 0.196$ [16]. For census tracts, $1/\vartheta_1 \approx -0.2817$ and $1/\vartheta_K \approx 0.1379$. Then the CAR prior in (9) is proper.

Therefore, the joint posterior density of $(\mathbf{V}, \alpha, \beta, \rho, \mathbf{W}, \delta_0, \eta_1)$ given $(\mathbf{y}, \mathbf{\Delta})$ is

$$[\mathbf{V}, \alpha, \beta, \rho, \mathbf{W}, \delta_0, \eta_1 \,|\, \mathbf{y}, \mathbf{\Delta}] \propto$$

$$L(\alpha, \mathbf{V} \,|\, \mathbf{y}, \mathbf{\Delta})[\mathbf{V} \,|\, \delta_0, \beta, \mathbf{W}][\beta][\mathbf{W} \,|\, \delta_0, \eta_1, \rho][\delta_0][\eta_1][\rho][\alpha], \quad (17)$$

where all the above densities are given by (1), (4), (8), (9), (15) and (16), respectively.

3. Simulation and Results

3.1. G.P.S. algorithm

The adjacency matrix $\mathbf{C}$ is a symmetric sparse matrix with few non-zero entries compared to the high matrix dimension. The computation will be burdened heavily because of the high dimensionality. Fortunately, sparse matrix with specific pattern and structure usually can be converted to a banded matrix, whose non-zero entries are confined to a diagonal band, comprising the main diagonal and zero or more diagonals on either side. Gibbs et al. [6] developed an algorithm named Gibbs-Poole-Stockmeyer algorithm to reduce the bandwidth of sparse symmetric matrices, which produces a vector storing new locations of counties in the banded matrix. The idea is to introduce a permutation matrix $\mathbf{P}$ such that $\mathbf{PWP}'$ has a small bandwidth defined by the maximum of the set $\{|i - j| : w_{ij} \neq 0\}$. Thus, the vector $\mathbf{W}$ is no longer with counties in alphabetical order. This permuted matrix leads to compact storage and reduced computation times for solving linear equations, matrix inverse or eigenvalue problems. Meantime, we have to permutate order of indices in matrix $\mathbf{X}_1$ correspondingly. This procedure greatly promote the speed of computation especially when the dimension of $\mathbf{W}$ is very large.

3.2. Simulation

From survey database, age is coded continuously with range from 66 to 104. We try to avoid grouping data because of missing observations in

categories of grouped ages by counties. Meanwhile, there is no specific clue to categorize age into groups, thus, we consider model age effect in continuous form. This would be reasonable because the main purpose of the study is to examine spatial effect by depleting other determinants.

Implementation of the Gibbs sampler is straightforward. The full conditionals (given Appendix A) for γ have close forms and those for $\mathbf{V}$, ρ, η_1, ϕ and α are log-concave. The adaptive rejection sampler of Gilks and Wild [7] is used. At the beginning of Gibbs sampling, we set both ω_1 and ω_2 equalling to 2.0 and a flat prior on age effect with precision parameter $\tau = 1500$. Hyper-parameter δ_0 is 0.5. The total number of iterations is 1,00,000 with the first 10,000 samples discarded as burn-in.

3.3. Results

The posterior mean and standard error of ρ are listed in Table 1. The posterior density of ρ is given in Figure 1(c). It seems that ρ is slightly shifted and concentration toward the positive part but not strong, showing weak spatial correlation for the breast cancer survival rate. Meanwhile, the posterior mean of individual variation δ_0 is much larger than spatial variation δ_1, indicating other possible covariates such as census tract and social-economic related factors may exist.

Table 1. Posterior Means and Standard Deviations of $(\rho, \delta_v, \alpha, \beta_v)$

	Posterior Mean	Standard error
ρ	–0.0064	0.118
δ_0	0.6324	0.2522
δ_1	0.0402	0.0272
α	0.9864	0.0364
β_0	–3.5607	0.1355
β_1	0.5757	0.0390

Figures 2 and 3 present estimated county effects W_i, i = 1, ..., 99 and hazard rates λ_i for county i in Iowa. The spatial pattern on the map is vague. One possibility leading to weak spatial correlation is that most counties have few observations, thus leading to potential inaccuracy of the estimation. Because there are around 87 percent of women alive at the end of study period, another reason arises from the large number of censoring in the Iowa SEER data. Figure 4 displays baseline survival functions for people aged 66, 76, 86 and 96, respectively. From Table 1, there is a positive age effect on hazard rates, which means older patients have lower breast cancer survival. It is consistent with conventional results. All the trace plots for sampled parameters are presented in Figure 5.

3.4. Comment

In this study, we model breast cancel survival based on Weibull distributions by incorporating spatial-temporal effects using Bayesian Hierarchical Model.

- Census tracts. Census tracts are sub-county divisions closely related to social and economic status factors. We have tried to include smaller geographic areas of census tracts in the mixed model either jointly with counties, or separately. The difficulty in modeling highly censored survival data in the current model still remains for census tract level, or even worse, in which zero observations exist in many census tracts among total 782 census tract. Simulation results show weak spatial correlation under these subdivisions, however, with relatively large variation for individuals. The computation is very expensive, taking almost 13 minutes per iteration. In this high dimensional study, G.P.S. algorithm enhances the efficiency of sampler dramatically by cutting down simulation time to only 1 minute per iteration.

- Thin-plate spline. Thin-plate spline is used as spatial smoothers for point-referenced data and can be easily implemented. White et al. [17] applied Bayesian version of thin-plate splines to success rates for turkey hunters in Missouri and performed more powerful to smooth spatial pattern data. Most important, it can solve the problem caused by lack of sufficient observations in geographical sites in the current model, especially in census tract level. We are currently implemented the thin-plate spline model.

• Objective prior. We evaluated the robustness of Bayesian estimates by changing different hyper-parameters and scales for survival time and age. The results are very sensitive towards those settings. The current study applied subjective but rather vague prior on δ_0 which is controversial due to subjective selection. One possibility is to choose invariance prior corresponding to $(a_0, b_0) = (0, 0)$ in (14). It turns out that the joint posterior is improper. Now, we are exploring objective priors for all unknown parameters in the Weibull model. We incorporate county and census tract effect alternatively.

Acknowledgment

The project is supported by the grant R01-CA100760 from National Institute of Health.

References

[1] J. O. Berger and D. Sun, Bayesian analysis for the poly-Weibull distribution, J. Amer. Statist. Assoc. 88 (1993), 1412-1418.

[2] J. Besag, J. York and A. Mollié, Bayesian image restoration with two applications in spatial statistics, Ann. Inst. Statist. Math. 43 (1991), 1-59.

[3] D. Cox, Regression models and life tables, J. R. Statist. Soc. B 34(1) (1997), 187-203.

[4] D. Farrow, J. Samet and W. Hunt, Regional variation in survival following the diagnosis of cancer, J. Clinical Epidemiology 49(1) (1996), 843-847.

[5] G. Gatta et al., Variations in the survival of adult cancer patients in Italy, Tumori 83(1) (1997), 497-504.

[6] N. Gibbs, Jr. W. G. Poole and P. Stockmeyer, A comparison of several bandwidth and profile reduction algorithms, ACM Trans. Math. Software 4(2) (1976), 322-330.

[7] W. R. Gilks and P. Wild, Adaptive rejection sampling for Gibbs sampling, Applied Statistics 41 (1992), 337-348.

[8] H. Goldstein, Multilevel Statistical Models, Institute of Education, London, 1995.

[9] C. He and J. Cao, Bayesian Estimation of Age-Specific Nest Survival Rate With Unknown Age, Submitted, 2006.

[10] S. Kaarjalainen, Geographic variation in cancer patient survival in Finland, chance, confounding or effect of treatment, J. Epidemiology and Community Health 44(1) (1995), 210-214.

[11] T. Law et al., Breast cancer in elderly women, presentation, survival and treatment options, Surgical Clinics of North America 76(1) (1996), 289-308.

[12] J. Lu, D. Sun, P. Speckman and J. Rouder, Hierarchical models for practice effects in response time, Revised for JRSS, Series A, 2005.

[13] S. Reader, Using survival analysis to study spatial point patterns in geographical epidemiology, Social Science and Medicine 50(1) (2000), 985-1000.

[14] R. Smith and J. Naylor, A comparison of maximum likelihood and Bayesian estimators for the three parameter Weibull distribution, Appl. Statist. 36(1) (1987), 358-369.

[15] S. Sturgeon, C. Schairer, M. Gail, M. McAdams, L. Brinton and R. Hoover, Geographic variation in mortality from breast cancer among white women in the United States, J. Nat. Cancer Institute 87(1) (1995), 1846-1853.

[16] D. Sun, R. K. Tsutakawa and P. L. Speckman, Posterior distribution of hierarchical models using CAR(1) distributions, Biometrika 86 (1999), 341-350.

[17] G. White, D. Sun and P. Speckman, Bayesian thin-plate splines as spatial smoothers, Revised for Biometrics, 2006.

Appendix A

The full conditional distributions of the model parameters are needed to implement a Gibbs sampler.

Lemma 1. *Under the priors* (4), (8), (12), (14), (15) *and* (16), *we have following results.*

(a) *Given* $(\alpha, \beta, \mathbf{W}, \mathbf{W}_2, \delta_0; \mathbf{\Delta}, \mathbf{y})$, *the conditional posterior of* V_{ij} *are independent with log-concave densities*,

$$[V_{ij} \mid \alpha, \beta, \mathbf{W}, \delta_0; \mathbf{\Delta}, \mathbf{y}] \propto \exp\left\{\Delta_{ij} V_{ij} - e^{V_{ij}} t_{ij}^{\alpha} - \frac{1}{2\delta_0}(V_{ij} - \mathbf{x}_{ij}'\gamma)^2\right\}, \tag{18}$$

for $i = 1, \ldots, K$, $j = 1, \ldots, n_i$.

(b) *Given* $(\mathbf{V}, \delta_0, \eta_1, \rho; \mathbf{\Delta}, \mathbf{y})$, *the conditional posterior density of* $\gamma = (\beta', \mathbf{W}')'$ *is multivariate normal with mean vector* $\boldsymbol{\mu}^*$ *and variance matrix* $\delta_0 \mathbf{\Sigma}^*$, *such that*

$$\mathbf{\Sigma}^* = (\delta_0 \mathbf{\Sigma}^{-1} + \mathbf{X}'\mathbf{X})^{-1}, \tag{19}$$

$$\boldsymbol{\mu}^* = \mathbf{\Sigma}^* \mathbf{X}'\mathbf{V}. \tag{20}$$

(c) *The conditional posterior density of* ρ *given* $(\mathbf{W}, \eta_1, \delta_0)$ *is*

$$[\rho \mid \mathbf{W}, \eta_1, \delta_0] \propto \prod_{i=1}^{K} (1 - \rho\vartheta_i)^{1/2} \exp\left\{\frac{\rho\eta_1}{2\delta_0} \mathbf{W}'\mathbf{C}\mathbf{W}\right\}. \tag{21}$$

(d) *Let* $\widetilde{\mathbf{V}} = \mathbf{V} - \mathbf{X}\gamma$. *The conditional posterior density of* δ_0 *given* $(\mathbf{V}, \gamma, \eta_1, \rho)$ *is inverse gamma* $(\widetilde{a}_0, \widetilde{b}_0)$, *where*

$$\widetilde{a}_0 = a_0 + \frac{1}{2}(N + K) \text{ and } \widetilde{b}_0 = b_0 + \frac{1}{2}[\widetilde{\mathbf{V}}'\widetilde{\mathbf{V}} + \eta_1(\mathbf{W}'\mathbf{W} - \rho\mathbf{W}'\mathbf{C}\mathbf{W})]. \quad (22)$$

(e) *The conditional posterior density of* η_1 *given* $(\mathbf{W}, \delta_0, \rho)$ *is*

$$[\eta_1 \mid \mathbf{W}, \delta_0, \rho] \propto \frac{\eta_1^{K/2}}{(\eta_1 + 1)^2} \exp\left\{-\frac{\eta_1}{2\delta_0}\mathbf{W}'(\mathbf{I} - \rho\mathbf{C})\mathbf{W}\right\}. \quad (23)$$

(f) *The conditional density of* $\varphi = \log(\eta_1)$ *given* $(\mathbf{W}, \delta_0, \rho)$ *is log-concave and given by*

$$[\varphi \mid \mathbf{W}, \delta_0, \rho] \propto \frac{e^{(K/2+1)\varphi}}{(e^{\varphi} + 1)^2} \exp\left\{-\frac{e^{\varphi}}{2\delta_0}\mathbf{W}'(\mathbf{I} - \rho\mathbf{C})\mathbf{W}\right\}. \quad (24)$$

(g) *The conditional posterior density of* α *given* $(\mathbf{V}; \boldsymbol{\Delta}, \mathbf{y})$ *is log-concave and given by*

$$[\alpha \mid \mathbf{V}; \boldsymbol{\Delta}, \mathbf{y}] \propto \alpha^{\widetilde{a}_\alpha} \exp\left(-\widetilde{b}_\alpha\alpha - \sum_{i=1}^{K}\sum_{j=1}^{n_i} t_{ij}^{\alpha} e^{V_{ij}}\right), \quad (25)$$

where

$$\widetilde{a}_\alpha = a_\alpha + \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Delta_{ij} - 1,$$

$$\widetilde{b}_\alpha = b_\alpha - \sum_{i=1}^{K}\sum_{j=1}^{n_i}\Delta_{ij}\log(t_{ij}).$$

Proof. Proving (18) is easy. We also need to show it is log-concave. Because the second derivative of the logarithm of density (18)

$$-e^{V_{ij}}t_{ij}^{\alpha} - \frac{1}{\delta_0},$$

is negative. The result holds. Parts (b), (d) and (e) can be verified easily

from (17). By (17), it is clear that the conditional posterior density of ρ is

$$[\rho \mid \mathbf{W}, \eta_1, \delta_0] \propto \mid \mathbf{I} - \rho\mathbf{C} \mid^{1/2} \exp\left\{-\frac{\eta_1}{2\delta_0}\mathbf{W}'(\mathbf{I} - \rho\mathbf{C})\mathbf{W}\right\}$$

$$\propto \mid \mathbf{O}_1\mathbf{I}\mathbf{O}_1' - \rho\mathbf{O}_1\boldsymbol{\Upsilon}_1\mathbf{O}_1' \mid^{1/2} \exp\left\{\frac{\rho\eta_1}{2\delta_0}\mathbf{W}'\mathbf{C}\mathbf{W}\right\}$$

$$\propto \mid \mathbf{I} - \rho\boldsymbol{\Upsilon}_1 \mid^{1/2} \exp\left\{\frac{\rho\eta_1}{2\delta_0}\mathbf{W}'\mathbf{C}\mathbf{W}\right\},$$

where $\boldsymbol{\Upsilon}_1 = \mathrm{diag}(\vartheta_1, ..., \vartheta_K)$ and $\mathbf{O}_1$ is the eigenvector matrix of $\mathbf{C}$. Because $\mathbf{I} - \rho\boldsymbol{\Upsilon}_1$ is a diagonal matrix, (c) hold.

The density (24) is log-concave by verifying the second derivative

$$-\frac{2e^{\varphi}}{(e^{\varphi}+1)^2} - \frac{e^{\varphi}\mathbf{W}'(\mathbf{I} - \rho\mathbf{C})\mathbf{W}}{2\delta_0},$$

which is always negative. Then (f) holds.

Finally, the distribution functions (25) can be verified easily from (17). We take the second derivative of the logarithm of (25) and have

$$-\frac{\tilde{a}_\alpha}{\alpha^2} - \sum_{i=1}^{K}\sum_{j=1}^{n_i} t_{ij}^{\alpha} e^{V_{ij}} \log^2(t_{ij}),$$

which is negative. The conclusion (g) follows.

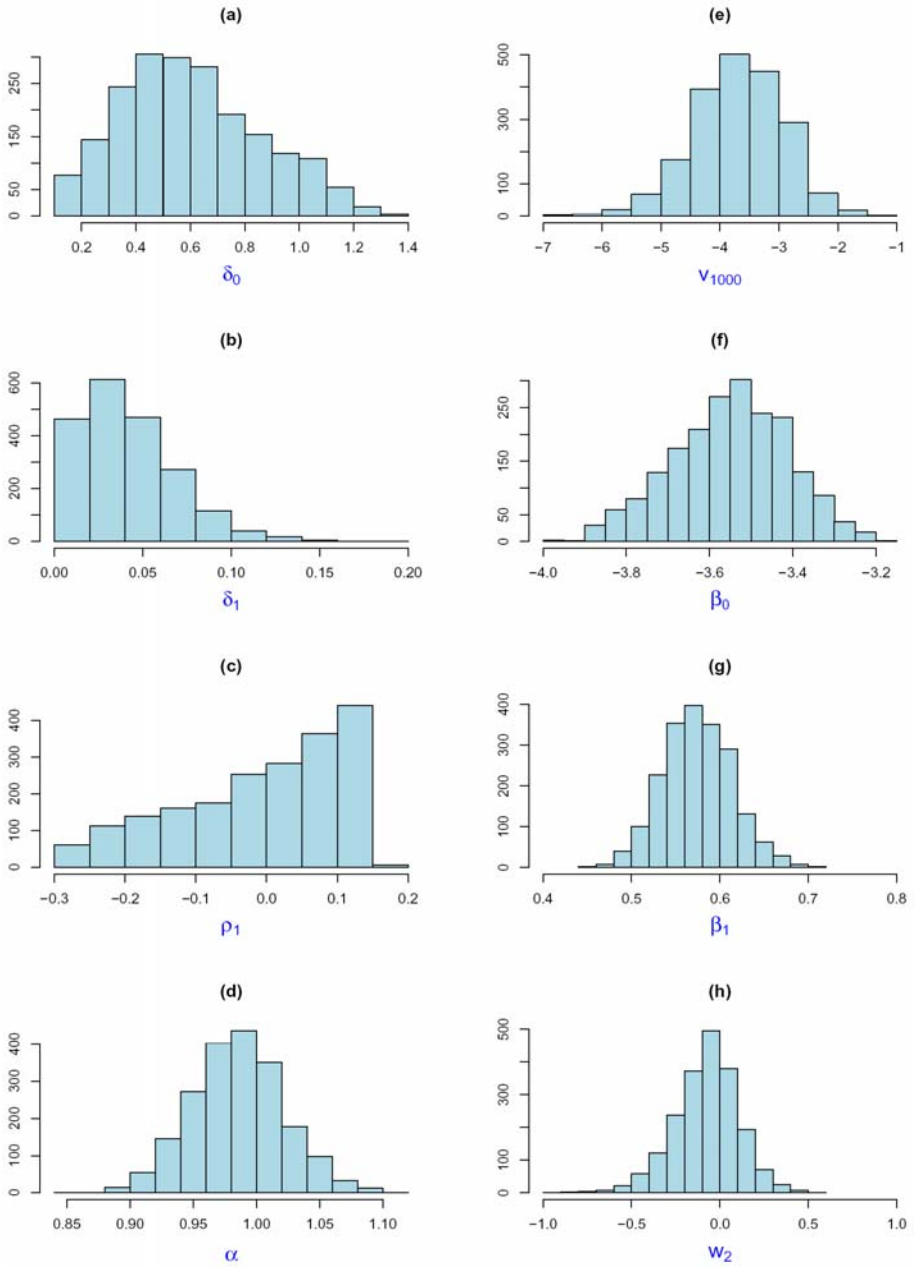

Figure 1. Histogram of posterior densities of selected parameters $(\delta_0, \delta_1, \rho, \alpha, V_{1000}, \beta_0, \beta_1, W_2)$ for Iowa breast cancer data.

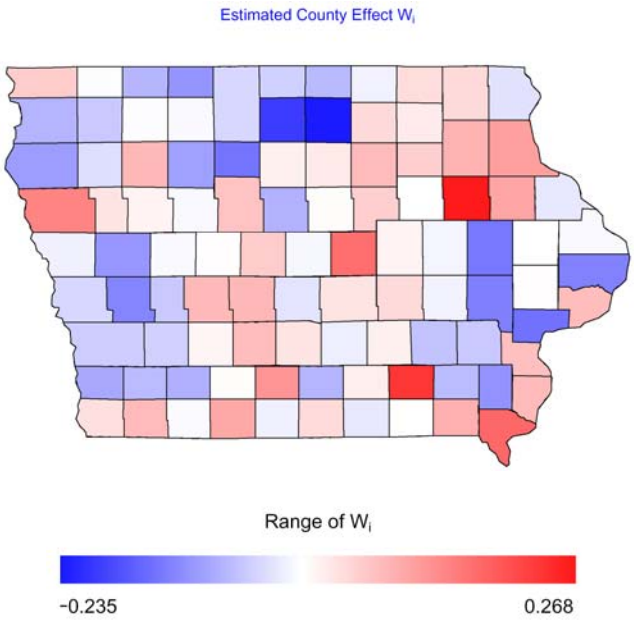

Figure 2. The map of posterior means of the county effects W_i, $i = 1, ..., 99$ for Iowa breast cancer data.

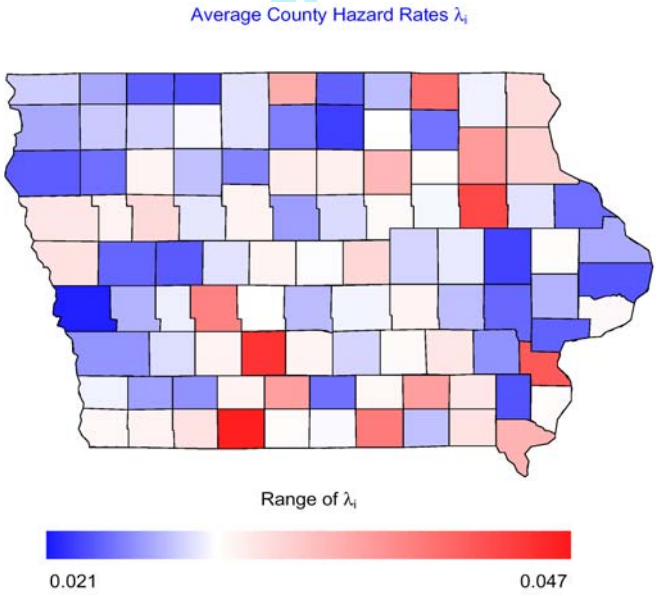

Figure 3. The map of average hazard rates λ_i, $i = 1, ..., 99$ of county for Iowa breast cancer data.

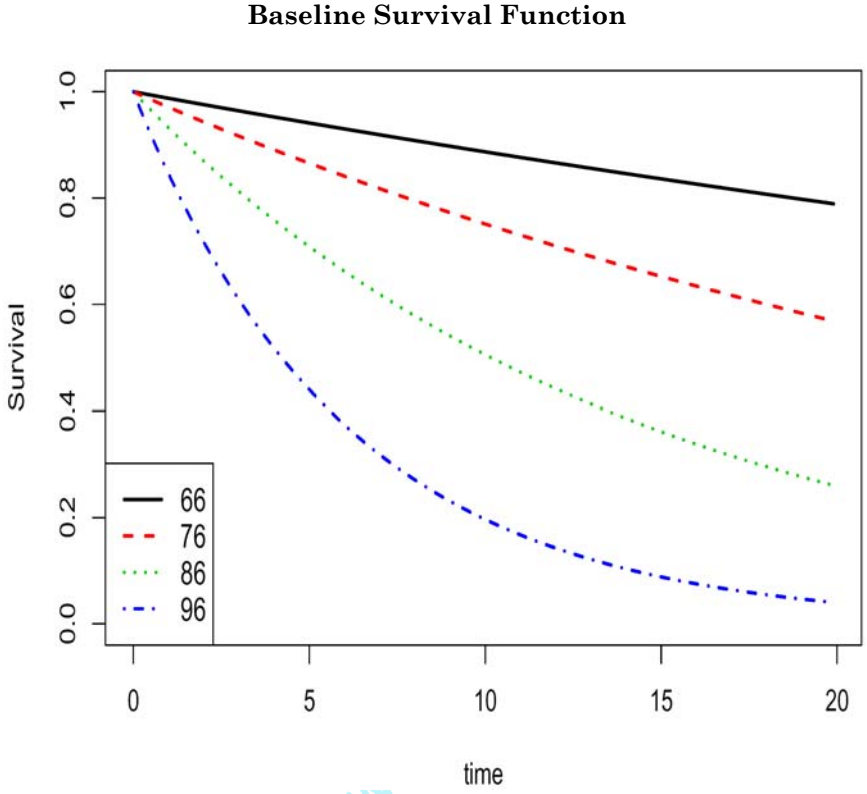

Figure 4. The estimated baseline survival functions for patients aged 66, 76, 86, 96 over 20 years period for Iowa breast cancer data.

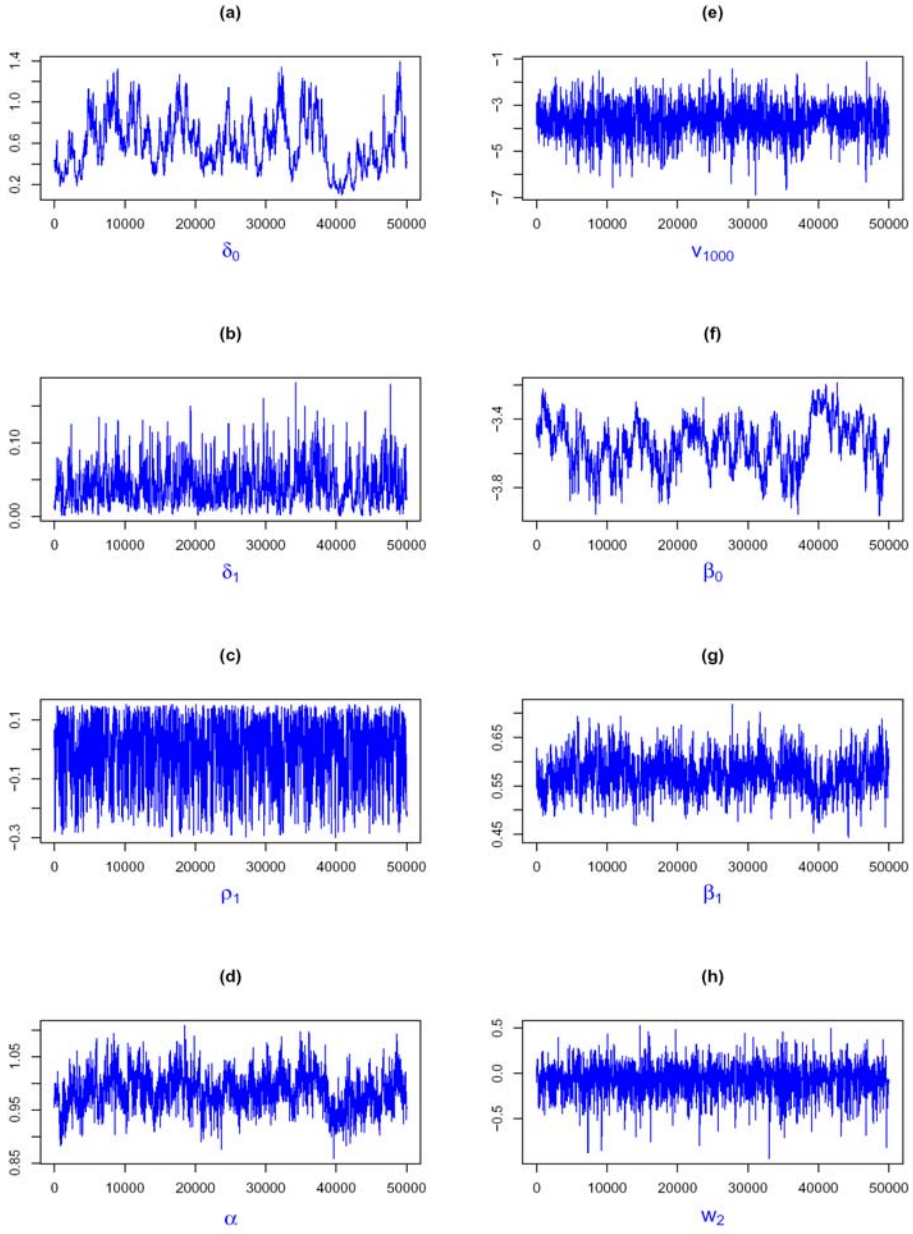

Figure 5. The trace plots for selected parameters $(\delta_0, \delta_1, \rho, \alpha, V_{1000}, \beta_0, \beta_1, W_2)$ for Iowa breast cancer data.

■