Advances in Data Science and Adaptive Analysis
Vol. 11, Nos. 3 & 4 (2019) 1950005 (32 pages)

DOI: 10.1142/S2424922X19500050

Semi-Parametric Cure Rate Proportional Odds Models with Spatial Frailties for Interval-Censored Data

Bao Yiqi*, Vicente G. Cancho†, Francisco Louzada‡ and Adriano K. Suzuki§

Departamento de Matemática Aplicada e Estatística, Universidade de São Paulo, Brazil
*baoyiqi@gmail.com
†garibay@icmc.usp.br
‡louzada@icmc.usp.br
§suzuki@icmc.usp.br

Received 24 June 2019
Accepted 25 June 2019
Published 19 July 2019

In this work, we proposed the semi-parametric cure rate models with independent and dependent spatial frailties. These models extend the proportional odds cure models and allow for spatial correlations by including spatial frailty for the interval censored data setting. Moreover, since these cure models are obtained by considering the occurrence of an event of interest is caused by the presence of any nonobserved risks, we also study the complementary cure model, that is, the cure models are obtained by assuming the occurrence of an event of interest is caused when all of the nonobserved risks are activated. The MCMC method is used in a Bayesian approach for inferential purposes. We conduct an influence diagnostic through the diagnostic measures in order to detect possible influential or extreme observations that can cause distortions on the results of the analysis. Finally, the proposed models are applied to the analysis of a real data set.

Keywords: Bayesian inference; cure rate; influence diagnostics; semi-parametric models; spatial frailty; survival analysis.

1. Introduction

With the development of medical and health sciences, the data sets collected from clinical studies pose some new challenges to statisticians. New statistical models which can incorporate these changes should be investigated. The most prevalent change noted in many clinical studies is that more patients respond favorably to a treatment or are not susceptible to the event of interest in the study, so they are considered cured or have prolonged disease-free survival. This proportion of patients is called the cure fraction or cure rate. Incorporating the cure fraction

*Corresponding author.

in survival models leads to cure rate models or long-term survival models. These models have been widely developed in the biostatistics literature. One of the most famous cure rate models is the mixture cure model introduced by Berkson and Gage [1952]. This model has been extensively discussed by several authors, including Farewell [1982], Maller and Zhou [1996], Ewell and Ibrahim [1997] and Stangl and Greenhouse [1998]. Later, Yakovlev and Tsodikov [1996] and Chen *et al.* [1999] proposed the promotion time cure model or bounded cumulative hazard model in cancer relapse settings, assuming that a latent biological process of propagation of latent carcinogenic tumor cells is generating the observed failure (relapse). Cooner *et al.* [2007] generalized this framework to a flexible class of cure models under latent activation schemes, Rodrigues *et al.* [2009] extend the promotion time cure model proposed by Chen *et al.* [1999] through the generating function of a real sequence introduced by Feller [1968]. Cancho *et al.* [2011] proposed a flexible cure rate model that encompasses as special cases the mixture model [Berkson and Gage (1952)], the promotion time cure model [Chen *et al.* (1999)] and the cure rate proportional odds model proposed by Gu *et al.* [2011].

The second challenge is the existence of incomplete (censoring) data sets. In many clinical trials, the patients are examined periodically for disease occurrence or progression. In this situation, the exact failure time of each patient cannot be observed. Rather, it can only be determined to lie in an interval obtained from a sequence of examination times. This time to event is known as the interval censoring [Peto (1973)]. The estimation methods available for right censored data, such as the Kaplan–Meier estimator, are not adequate for application to interval-censored data because they can lead to biased estimation and invalid inferences. The interval censorship information should be taken into account in modeling [Rücker and Messerer (1988); Lindsey and Ryan (1998); Sun and Chen (2010)].

Another challenge arises with the development of geographic information systems (GIS) and computing technology. Data sets increasingly incorporate geographical information about the subjects under study. Adopting a traditional cure rate model by including random effects for each region fails to consider the correlations of the regions. Therefore, several researchers have developed survival models that account for spatial clustering and variation. Banerjee *et al.* [2003] investigated spatially correlated frailties in traditional parametric survival models. Later, Banerjee and Carlin [2004] introduced spatially correlated frailties in the parametric cure model. They developed a Bayesian approach to the mixture cure model [Berkson and Gage (1952)] with spatial random effects in the survival function for subjects at risk and spatial frailties using a multivariate conditionally autoregressive (MCAR) prior. Recently, Pan *et al.* [2014] proposed a Bayesian approach under a proportional hazards frailty model to analyze interval-censored survival data with spatial correlation. Dan Li and Dey [2016] proposed flexible cure rate models in analyzing univariate right-censored data based on the assumption that the logarithm of

survival time follows a generalized extreme value distribution with spatial and nonlinear covariate effects. Recently, Bao *et al.* [2017] proposed a proportional odds cure rate models to allow spatial correlations by including spatial frailty in the interval censored data set.

Considering these challenges, the main goal of this paper is to propose a new semi-parametric cure rate models with independent and dependent spatial frailties for the interval-censored data. Following Bao *et al.* [2017], we assume two most natural activations schemes, the first and last activations schemes. The first activation scheme presents the situation where the presence of any of latent risk will ultimately lead to the occurrence of the event, while the last activation scheme presents a situation where the occurrence of the event will happen when all latent risks are activated. To investigate the correlation between the hazard function and cure fraction, the covariates and frailties are incorporated into both of them, assuming the spatial frailties can be independent or dependent. The inference procedures are developed under a Bayesian perspective.

The proposed models are also compared with the models introduced by Pan *et al.* [2014] and Bao *et al.* [2017] through the deviance information criterion (DIC) proposed by Spiegelhalter *et al.* [2002]. Furthermore, we also conduct influence diagnostics in order to check the model assumptions and conduct sensitivity analysis to detect possible influential or extreme observations that can cause distortions in the results. A Bayesian case deletion influence diagnostics is developed for the joint posterior distribution based on the ψ-divergence [Peng and Dey (1995); Weiss (1996)].

The rest of the paper is organized as follows. In Sec. 2, we present the semi-parametric cure rate proportional odds models with spatial frailties for interval-censored data. Bayesian inference is investigated in Sec. 3. In Sec. 4, we conduct some simulation studies. An example application to a real data set (smoking cessation study data) is addressed in Sec. 5. Finally, in Sec. 6, we offer some concluding remarks.

2. Model

In the cure rate proportional odds model (CRPO model) proposed by Gu *et al.* [2011], the ratio of hazards for the CRPO model for two covariate values do not change over time, unlike the Chen *et al.* [1999] model.

As reported by Bao *et al.* [2017], this model can be characterized by the latent factors model of Cooner *et al.* [2007] with a geometric distribution for the number of latent factors. Suppose that there are I regions and n_i individuals in ith region. We denote by T_{ij} the random variable for the observed time to event of the jth individual in the ith region, where $j = 1, \ldots, n_i$ and $i = 1, \ldots, I$. Suppose that the (i, j)th individual is potentially exposed to M_{ij} latent risk, where M_{ij} denote, the initial number of competing causes related to the occurrence of an event and assuming M_{ij} has a Geometric distribution with parameter $1/(1+\theta_{ij})$ having probability

mass function (pmf) given by

$$P(M_{ij} = m) = \frac{\theta_{ij}^m}{(\theta_{ij}+1)^{m+1}}, \quad m = 0, 1, 2 \ldots, \tag{1}$$

where $\theta_{ij} > 0$, $E(M_{ij}) = \theta_{ij}$ and $\mathrm{Var}(M_{ij}) = \theta_{ij}(1+\theta_{ij})$. Let Y_{cij} for $c = 1, \ldots, M_{ij}$ denote the failure time of jth individual in the ith region due to the cth latent risk. Also, suppose that given M_{ij}, the random variables Y_{cij}'s are mutually independent with distribution function $F(\cdot) = 1 - S(\cdot)$. Thus, if we assume the situation that the presence of any of latent risk will ultimately lead to the occurrence of the event, the time to event of interest can be defined by random variable $T_{ij} = \min\{Y_{cij}, c = 1, \ldots, M_{ij}\}$ for $M_{ij} \geq 1$ and $T_{ij} = \infty$ if $M_{ij} = 0$ with $P(T_{ij} = \infty \,|\, M_{ij} = 0) = 1$. Using the terminology borrowed from Cooner *et al.* [2007], this scenario is regarded as first activation scheme because in this case, we assume that the event of interest occurs when the first possible cause is activated.

In this setup, the survival function for the population is given by

$$S_{\mathrm{pop}}(t_{ij}) = \left[1 + \theta_{ij} F(t_{ij})\right]^{-1}. \tag{2}$$

We can note that this survival function has proportional odds structure when covariates $\boldsymbol{x}_{ij}$ are modeled via $\theta_{ij}(\boldsymbol{x}_{ij})$ and the latent survival $F(t_{ij})$ is free of $\boldsymbol{x}_{ij}$ because

$$\frac{1 - S_{\mathrm{pop}}(t_{ij} \,|\, \boldsymbol{x}_{ij})}{S_{\mathrm{pop}}(t_{ij} \,|\, \boldsymbol{x}_{ij})} = \theta_{ij}(\boldsymbol{x}_{ij}) F(t_{ij}).$$

From (2), the probability density function and hazard function are given, respectively, by

$$f_{\mathrm{pop}}(t_{ij}) = \theta_{ij} f(t_{ij}) \left[1 + \theta_{ij} F(t_{ij})\right]^{-2} \quad \text{and} \quad h_{\mathrm{pop}}(t_{ij}) = \theta_{ij} f(t_{ij}) \left[1 + \theta_{ij} F(t_{ij})\right]^{-1},$$

where $f(t_{ij}) = \frac{\partial}{\partial t_{ij}} F(t_{ij})$.

We note that model (2) can be rewritten as a mixture model given by

$$S_{\mathrm{pop}}(t_{ij}) = (1+\theta_{ij})^{-1} + \left(1 - (1+\theta_{ij})^{-1}\right) \left\{ \frac{\left[1 + \theta_{ij} F(t_{ij})\right]^{-1} - (1+\theta_{ij})^{-1}}{1 - (1+\theta_{ij})^{-1}} \right\}.$$

The survival function of uncured (susceptible) individuals has the expression given by

$$S_{\mathrm{sus}}(t_{ij}) = \frac{\left[1 + \theta_{ij} F(t_{ij})\right]^{-1} - (1+\theta_{ij})^{-1}}{1 - (1+\theta_{ij})^{-1}}.$$

Another situation occurs if we assume that the presence of all latent risks will ultimately lead to the occurrence of the event. In this case, the time to event of interest is defined by random variable $T_{ij} = \max\{Y_{cij}, c = 1, \ldots, M_{ij}\}$ for $M_{ij} \geq 1$ and $T_{ij} = \infty$ if $M_{ij} = 0$ with $P(T_{ij} = \infty \,|\, M_{ij} = 0) = 1$. This second situation is known as last activation scheme because the event of interest only takes place after all the latent causes have been occurred.

The survival function for the population is given by

$$S_{\text{pop}}(t_{ij}) = 1 + (1+\theta_{ij})^{-1} - [1+\theta_{ij}S(t_{ij})]^{-1}. \tag{3}$$

From (3), the corresponding probability density function and hazard function are given, respectively, by

$$f_{\text{pop}}(t_{ij}) = \theta_{ij} f(t_{ij})(1+\theta_{ij}S(t_{ij}))^{-2}$$

and

$$h_{\text{pop}}(t_{ij}) = \frac{\theta f(t_{ij})\,[1+\theta_{ij}S(t_{ij})]^{-2}}{1+(1+\theta_{ij})^{-1} - (1+\theta_{ij}S(t_{ij}))^{-1}}.$$

The survival function (3) can also be rewritten as a mixture model given by

$$S_{\text{pop}}(t_{ij}) = (1+\theta_{ij})^{-1} + \left(1-(1+\theta_{ij})^{-1}\right)\left\{\frac{1-(1+\theta_{ij}S(t_{ij}))^{-1}}{1-(1+\theta_{ij})^{-1}}\right\}.$$

The survival function of susceptible individuals is given by

$$S_{\text{sus}}(t_{ij}) = \frac{1-(1+\theta_{ij}S(t_{ij}))^{-1}}{1-(1+\theta_{ij})^{-1}}.$$

In this work, we denoted the survival functions (2) and (3) by $S^F_{\text{pop}}(t_{ij})$ and $S^L_{\text{pop}}(t_{ij})$, respectively. There is another kind of situation that the event of interest occurs when some of the possible causes are activated, and given the number of latent causes M_{ij}, the number of activated causes is a random variable with the discrete uniform distribution on $\{1,\ldots,M\}$. This situation is known as random activation scheme. In this setup, the survival function for the population has the expression given by

$$S_{\text{pop}}(t_{ij}) = (1+\theta_{ij})^{-1} + (1-(1+\theta_{ij})^{-1})S(t_{ij}), \tag{4}$$

and it is denoted by $S^R_{\text{pop}}(t_{ij})$. The improper survival function (4) can be written as $S^R_{\text{pop}}(t_{ij}) = p_{0ij} + (1-p_{0ij})S(t_{ij})$, which is the same as the mixture rate model (Berkson and Gage [1952]). Under the different activation schemes, the models differ by its surviving, density and hazard functions. However, in both setups, the cured fraction is given by $p_{0ij} = \lim_{t_{ij}\to\infty} S_{\text{pop}}(t_{ij}) = (1+\theta_{ij})^{-1}$. Moreover, under the survival functions in (2)–(4) for any distribution function $F(\cdot)$, we have $S^F_{\text{pop}}(t_{ij}) \leq S^R_{\text{pop}}(t_{ij}) \leq S^L_{\text{pop}}(t_{ij})$ for all $t_{ij} > 0$.

As well known, the cure fraction plays a key role in the survival models with a cure fraction. So, we consider the parametrization of the model in the cure fraction in expressions. Since $p_{0ij} = (1+\theta_{ij})^{-1}$, we have $\theta_{ij} = p_{0ij}^{-1} - 1$. Moreover, we propose that the cured probability of an individual (i,j) is associated with covariates $\boldsymbol{x}_{ij}$ and it can be modeled by a logistic regression

$$p_{0ij} = \frac{\exp(\xi_{ij})}{1+\exp(\xi_{ij})},$$

where ξ_{ij} is a linear form of covariates, $\xi_{ij} = \boldsymbol{x}_{ij}^\top \boldsymbol{b}$ and $\boldsymbol{b}$ is a p_1-dimensional vector representing the effects of covariates on the cured probability.

Using p_{0ij} as a parameter, the improper survival functions (2) and (3) can be written as

$$S^F_{\text{pop}}(t_{ij}) = \left[1 + (p_{0ij}^{-1} - 1)F(t_{ij})\right]^{-1} \tag{5}$$

and

$$S^L_{\text{pop}}(t_{ij}) = 1 + p_{0ij} - \left[1 + (p_{0ij}^{-1} - 1)S(t_{ij})\right]^{-1}. \tag{6}$$

Henceforth, we assume that the hazard function of latent random variables Y_{cij} is a proportional hazard (PH) model with the baseline hazard function $h_0(t \,|\, \boldsymbol{\alpha})$, the conditional hazard function and corresponding survival function are given, respectively, by

$$h(t \,|\, \boldsymbol{\phi}) = h_0(t \,|\, \boldsymbol{\alpha}) \exp(\lambda_{ij}) \;\text{ and }\; S(t \,|\, \boldsymbol{\phi}) = S_0(t \,|\, \boldsymbol{\alpha})^{\exp(\lambda_{ij})}, \tag{7}$$

where $\boldsymbol{\phi} = (\boldsymbol{\alpha}, \lambda_{ij})$, $\lambda_{ij} = \boldsymbol{z}'_{ij}\boldsymbol{\beta}$, $\boldsymbol{z}_{ij}$ and $\boldsymbol{\beta}$ are defined as above, $S_0(t \,|\, \boldsymbol{\alpha})$ is the baseline survival function corresponding to $h_0(t \,|\, \boldsymbol{\alpha})$ and $\boldsymbol{\alpha}$ is the parameter vector of the baseline functions. Here, we assume that the baseline functions have the piecewise exponential distribution. Let the vector $\boldsymbol{a} = (a_0, a_1, \ldots, a_{Q-1})$ with $0 = a_0 < a_1 < \ldots < a_{Q-1} < \infty$ be a finite partition of time axis and α_q be the hazard rate of qth interval of intervals $(0, a_1], \ldots, (a_{Q-1}, \infty]$, for $q = 1, \ldots, Q$, so the baseline survival function can be written as

$$S_0(t \,|\, \boldsymbol{\alpha}) = \exp\left\{-\sum_{q=1}^{Q} \alpha_q \Delta_q(t)\right\}, \quad t > 0,$$

where

$$\Delta_q(t) = \begin{cases} 0, & \text{if } t < a_{q-1}, \\ t - a_{q-1}, & \text{if } a_{q-1} \le t < a_q, \\ a_q - a_{q-1}, & \text{if } t \ge a_q. \end{cases} \quad q = 1 \ldots, Q,$$

Note that if $Q = 1$, we have an exponential distribution with a parameter α as the particular case. Moreover, if we partition the time axis $0 = a_0 < a_1 < \cdots < a_{Q-1} < \infty$ so that they denoted the ordered distinct time points of all observed interval end points, then we have $t_q = a_q$ for $q = 0, \ldots, Q$. The survival function can now be written as

$$S_0(t_q \,|\, \boldsymbol{\alpha}) = \begin{cases} 1, & q = 0, \\ \exp\left\{-\sum_{k=1}^{q} \alpha_k (a_k - a_{k-1})\right\}, & q = 1, \ldots, Q-1, \\ 0, & q = Q, \end{cases} \tag{8}$$

where $\alpha_q(a_q - a_{q-1}) \ge 0$, $q = 1, \ldots, Q$. Here, we called the function (2) by PH cure rate model under the first activation, denoted by PHCRM-FA and also the function (3) by PH cure rate model under the last activation, denoted by PHCRM-LA.

In this work, we introduce the frailties U_i and V_i to better explain the effect of survival time of susceptible individuals and on the cured probability through linear predictor expression

$$\lambda_{ij} = \boldsymbol{z}'_{ij}\boldsymbol{\beta} + U_i,$$
$$\xi_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{b} + V_i, \quad \text{for } j = 1\ldots,\ n_i, i = 1,\ldots,I.$$

The frailties U_i and V_i are spatially correlated across the regions. We propose two approaches. In the first approach, we employ separate independent conditionally auto-regressive (CAR) prior distribution on $(\boldsymbol{U}, \boldsymbol{V})$.

The CAR model (Besag [1974]) have become very popular in the Bayesian analysis of areal data, especially in disease mapping. Here, we denote the random frailties vectors by $\boldsymbol{\psi}_1 = \boldsymbol{U}$ and $\boldsymbol{\psi}_2 = \boldsymbol{V}$, $\boldsymbol{W}$ is the adjacent matrix of the map so that $\boldsymbol{W}_{ii'} = 1$ if region i and i' are adjacent, and 0 otherwise, and $w_{i+} = \sum_j w_{ij}$ is the number of regions adjacent to region i. The zero-centered CAR prior for $\boldsymbol{\psi}_l$, $l = 1, 2$, is an I-dimensional Gaussian distribution with mean $\mathbf{0}$ and precision matrix $\theta_l^{-1}(\boldsymbol{D_W} - \boldsymbol{W})$, where $\theta_l > 0$ and $\boldsymbol{D_W}$ is diagonal with $(\boldsymbol{D_W})_{ii} = w_{i+}$ for $i = 1,\ldots,I$. The CAR model is denoted by CAR(θ_l) for l fixed. We can note that the precision matrix $\theta_l^{-1}(\boldsymbol{D_W} - \boldsymbol{W})$ is rank, deficient, so it is a nonpositive definite matrix and it leads to improper distribution. This singularity, while theoretically awkward, creates a little problem in a Bayesian implementation, since the identifying sum-to-zero constraint $\sum_{i=1}^{I} \psi_{il} = 0$ is easily imposed in a Gibbs sampler simply by recentering the ϕ_i draws around zero after every iteration.

In the second approach, we assume that the spatial priors on $(\boldsymbol{U}, \boldsymbol{V})$ are dependent, and they have multivariate conditionally auto-regressive MCAR prior distribution. Let $\boldsymbol{\psi} = (\boldsymbol{U}^\top, \boldsymbol{V}^\top)^\top$, we employ a multivariate MCAR distribution with a common smoothness parameter a, that is, $\boldsymbol{\psi}$ has a normal distribution with mean $\mathbf{0}$ and precision matrix $\boldsymbol{\Lambda} \otimes (\boldsymbol{D_W} - a\boldsymbol{W})$, where $\otimes$ denotes the Kronecker product, $\boldsymbol{\Lambda}$ is a 2×2 symmetric and positive definite matrix, $a \in (0, 1)$ and $\boldsymbol{W}$ is standardized so that each of its rows sum to 1. This prior is a proper distribution, the parameter a has a spatial smoothness interpretation. Value of a closer to 1 implies greater weight on the adjacency matrix $\boldsymbol{W}$, while a close to 0 implies that the adjacency structure has few roles to play in the precision matrix. This prior is denoted by MCAR$(a, \boldsymbol{\Lambda})$.

Moreover, we also employ the parameter $\boldsymbol{\psi}$ with an extended MCAR distribution which was proposed by Gelfand and Vounatsou [2003] and Carlin and Banerjee [2003], assuming different smoothness parameters for the parameters $\boldsymbol{U}$ and $\boldsymbol{V}$, say a_1 and a_2. Let $(\boldsymbol{D_W} - a_i\boldsymbol{W})$ denotes the corresponding positive definite matrix and $\boldsymbol{R_i}^\top \boldsymbol{R_i}$ denotes its Cholesky factorization, where $\boldsymbol{R_i}$ is a upper triangular matrix with real and positive diagonal entries with $I \times I$ dimension for $i = 1, 2$. The precision matrix can be written as

$$\boldsymbol{\Lambda} \otimes (D_W - \boldsymbol{aW}) = \begin{bmatrix} \lambda_{11}\boldsymbol{R_1}^\top \boldsymbol{R_1} & \lambda_{12}\boldsymbol{R_1}^\top \boldsymbol{R_2} \\ \lambda_{21}\boldsymbol{R_2}^\top \boldsymbol{R_1} & \lambda_{22}\boldsymbol{R_2}^\top \boldsymbol{R_2} \end{bmatrix} = \boldsymbol{R}'(\Lambda \otimes I_I)\boldsymbol{R},$$

where λ_{ij}'s are the elements of $\boldsymbol{\Lambda}$ and $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R_1} & 0 \\ 0 & \boldsymbol{R_2} \end{bmatrix}$ has dimension $2I \times 2I$. The $\boldsymbol{\Lambda} \otimes (\boldsymbol{D_W} - \boldsymbol{aW})$ is positive definite since $\boldsymbol{\Lambda}$ is positive definite. We denoted this prior by MCAR$(a_1, a_2, \boldsymbol{\Lambda})$.

3. Bayesian Inference

In this section, we briefly discuss the inference from a Bayesian viewpoint. Let $\boldsymbol{D}_{\text{obs}} = \{(A_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \delta_{ij});\ j = 1, \ldots, n_i, i = 1, \ldots M\}$ denote the observed data, where $A_{ij} = (t_{ijL}, t_{ijR}]$ is the interval during which individual j in cluster i occur at the event of interest, $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are the p_1-dimensional and p_2-dimensional vectors of covariates, and δ_{ij} is following interval censoring indicator: $\delta_{ij} = I(t_{ijR} < \infty)$. For the special case in which the survival time is right-(left-)censored, $R_{ij} = +\infty (L_{ij} = 0)$, whereas for exact observations, $t_{ijL} = t_{ijR}$. The likelihood function for the general interval-censored cure rate model [Finkelstein (1986)] is given by

$$L\{\boldsymbol{\varphi} \,|\, \boldsymbol{D}_{\text{obs}}\} \propto \prod_{i=1}^{I} \prod_{j=1}^{n_i} \left(S_{\text{pop}}(t_{ijL} \,|\, \boldsymbol{\varphi}) - S_{\text{pop}}(t_{ijR} \,|\, \boldsymbol{\varphi})\right)^{\delta_{ij}} S_{\text{pop}}(t_{ijL} \,|\, \boldsymbol{\varphi})^{1-\delta_{ij}}$$

$$\propto \prod_{i=1}^{I} \prod_{j=1}^{n_i} S_{\text{pop}}(t_{ijL} \,|\, \boldsymbol{\varphi}) \left(1 - \frac{S_{\text{pop}}(t_{ijR} \,|\, \boldsymbol{\varphi},)}{S_{\text{pop}}(t_{ijL} \,|\, \boldsymbol{\varphi})}\right)^{\delta_{ij}}, \tag{9}$$

where $\boldsymbol{\varphi} = (\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{U}, \boldsymbol{V})$, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_Q)$. We assume the follow prior densities for parameters $\boldsymbol{\alpha}$, $\boldsymbol{b}^\top$ and $\boldsymbol{\beta}^\top$:

- $\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)\boldsymbol{I}_{(0,\infty)}$, with μ_{α_i} and σ_{α_i} known, $i = 1, \ldots, Q$;
- $b_j \sim N(\mu_b, \sigma_b^2)$, $j = 0, \ldots, (p1 - 1)$, with μ_b and σ_b known;
- $\beta_j \sim N(\mu_\beta, \sigma_\beta^2)$, $j = 1, \ldots, p2$, with μ_β and σ_β known,

where $N(\mu, \sigma^2)\boldsymbol{I}_{(a,b)}$ denote the truncated Normal distribution which is the probability distribution of a normally distributed random variable whose value lies within the interval $-\infty \leq a < b \leq \infty$. As pointed out by Bao *et al.* [2017], the use of a truncated normal distribution as prior facilitates the insertion of information in certain regions of the parameter space, since the hyper-parameters no longer represent the mean and variance but still control the region of higher probability mass.

3.1. *Independent assumption*

We employ separate independent CAR prior on the random frailties $\boldsymbol{U} = (U_1, \ldots, U_I)^\top$ and $\boldsymbol{V} = (V_1, \ldots, V_I)^\top$, that is,

- $U_1, \ldots, U_I \sim \text{CAR}(\theta_1)$;
- $V_1, \ldots, V_I \sim \text{CAR}(\theta_2)$,

where θ_1 and θ_2 are positive unknown hyper-parameters, and we assume they have Inverse-Gamma prior with the known shape parameter $a_0 > 0$ and scale parameter $b_0 > 0$.

We assume $\boldsymbol{\varphi}$, θ_1 and θ_2 are *a priori* independent, that is,

$$\pi(\boldsymbol{\varphi}, \theta_1, \theta_2) = \pi(\boldsymbol{b})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\alpha})\pi(\boldsymbol{U} \,|\, \theta_1)\pi(\boldsymbol{V} \,|\, \theta_2)\pi(\theta_1)\pi(\theta_2). \tag{10}$$

For expressing prior noninformation, we consider $\mu_b = \mu_\beta = \mu_\eta = \mu_\alpha = 0$ with large values of σ_b^2, σ_β^2 and σ_α^2. Combining the likelihood function (9) and the prior distribution in (10), the joint posterior distribution for the parameters is given by

$$\pi(\boldsymbol{\varphi}, \theta_1, \theta_2 \,|\, \boldsymbol{D}_{\text{obs}}) \propto L(\boldsymbol{\varphi} \,|\, \boldsymbol{D}_{\text{obs}})\pi(\boldsymbol{\varphi}, \theta_1, \theta_2).$$

This joint posterior density is analytically intractable. So, we based our inference on the Markov chain Monte Carlo (MCMC) simulation methods. We can observe that the full conditional distributions for parameters $\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{U}$ and $\boldsymbol{V}$ have not closed forms, thus we will use the Metropolis–Hastings algorithm to generate a posterior samples for these parameters. To avoid range restrictions on the parameters α_i's, we define $\zeta_i = \log(\alpha_i)$ for $i = 1, 2, \ldots, Q$, to transform all parameters space to real space (necessary for the work with Gaussian proposal densities). Let $\boldsymbol{\vartheta} = (\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{U}, \boldsymbol{V}, \theta_1, \theta_2)$, regarding the Jacobian of this transformation, the joint prior density $\pi(\boldsymbol{\vartheta})$ has expression

$$\pi(\boldsymbol{\vartheta}) = \pi(\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\zeta}^{-1}, \boldsymbol{U}, \boldsymbol{V}, \theta_1, \theta_2) \times \exp\left(\sum_{i=1}^{Q} \zeta_i\right), \tag{11}$$

where $\boldsymbol{\zeta}^{-1}$ denote inverse function of $\boldsymbol{\zeta}$, that is $\boldsymbol{\zeta}^{-1} = \{\zeta_i^{-1} = \exp(\zeta_i), i = 1, \ldots, Q\}$.

On the other hand, the full conditional distributions for parameters θ_i's are given by

$$\begin{aligned}
\pi(\theta_i \,|\, \boldsymbol{\vartheta}_{-\theta_i}, \boldsymbol{D}_{\text{obs}}) &\propto \pi(\boldsymbol{\psi}_i \,|\, \theta_i)\pi(\theta_i) \\
&\propto (\theta_i)^{-k/2} \exp\left(-\frac{1}{2\theta_i}\boldsymbol{\psi}_i^\top(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{\psi}_i\right) \theta_i^{-a_0-1} \exp(-b_0\theta_i^{-1}) \\
&\propto \theta_i^{-\left(a_0+\frac{k}{2}\right)-1} \exp\left\{-\left(\frac{\boldsymbol{\psi}_i^\top(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{\psi}_i}{2} + b_0\right)\theta_i^{-1}\right\}, \quad i = 1, 2,
\end{aligned}$$

where $\boldsymbol{\psi}_1 = \boldsymbol{U}$, $\boldsymbol{\psi}_2 = \boldsymbol{V}$ and k is the rank of the matrix $\boldsymbol{D_W} - \boldsymbol{W}$. Thus, the full conditional distributions of the parameter θ_i show an Inverse-Gamma distribution with parameters $a_0 + \frac{k}{2}$ e $b_0 + \frac{1}{2}(\boldsymbol{\psi}_i'(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{\psi}_i)$. To generate a posterior sample we use the Gibbs sampler algorithm (see [Gamerman and Lopes (2006)]).

The joint posterior density of $\pi(\boldsymbol{\vartheta} \,|\, \boldsymbol{D}_{\text{obs}})$ is proportional to

$$\begin{aligned}
L(\boldsymbol{\varphi}^{-1} \,|\, \boldsymbol{D}_{\text{obs}}) \exp\Bigg\{-\frac{1}{2}\Bigg[&\sigma_b^{-2}\sum_{i=0}^{p1-1} b_i^2 + \sigma_\beta^{-2}\sum_{i=1}^{p2}\beta_i^2 + \sum_{i=1}^{Q}\frac{\exp(2\zeta_i)}{\sigma_\alpha^2} + \frac{\boldsymbol{U}^\top(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{U}}{\theta_1} \\
&+ \frac{\boldsymbol{V}^\top(\boldsymbol{D_W} - \boldsymbol{W})\boldsymbol{V}}{\theta_2}\Bigg] - (a_0 + 1)(\log(\theta_1) + \log(\theta_2)) - \left(\frac{b_0}{\theta_1} + \frac{b_0}{\theta_2}\right) + \sum_{i=1}^{Q}\zeta_i\Bigg\},
\end{aligned}$$

where $\boldsymbol{\varphi}^{-1} = (\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\zeta}^{-1}, \boldsymbol{U}, \boldsymbol{V})$.

3.2. *Dependent assumption*

Here, we assume that the spatial priors on the parameters $(\boldsymbol{U}, \boldsymbol{V})$ are dependent on each other. Let $\boldsymbol{\psi} = (\boldsymbol{U}^\top, \boldsymbol{V}^\top)^\top$, we first employ the parameter $\boldsymbol{\psi}$ has a MCAR distribution with a common smoothness parameter a, i.e.

$$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda}).$$

Further, we employ the parameter $\boldsymbol{\psi}$ with an extended MCAR distribution, assuming the different smoothness parameters for the parameters $\boldsymbol{U}$ and $\boldsymbol{V}$, say a_1 and a_2, that is,

$$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda}).$$

Similar to Carlin and Banerjee [2003], we consider the prior distributions for parameters $\boldsymbol{a}$ and $\boldsymbol{\Lambda}$ are given by

- $a_i \sim \text{Uniform}(0, 1)$ or $a_i \sim \text{Beta}(18, 2)$,
- $\boldsymbol{\Lambda} \sim \text{Wishart}(n_0, \Lambda_0)$, with n_0 and Λ_0 known,

where $i = 1$ for $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$ and $i = 1, 2$ for $\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$. To avoid range restrictions on the parameters a_i, considering the transformations $\rho_i = \log(a_i/(1 - a_i)) \in \mathbb{R}$, the joint prior density for the parameters $\boldsymbol{\vartheta} = (\boldsymbol{b}, \boldsymbol{\beta}, \kappa, \boldsymbol{\zeta}, \boldsymbol{\psi}, \boldsymbol{\Lambda}, \boldsymbol{\rho})$ can be written as

$$\pi(\boldsymbol{\vartheta}) = \pi(\boldsymbol{b})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\zeta})\pi(\boldsymbol{\psi} \mid \boldsymbol{\Lambda}, \boldsymbol{\rho})\pi(\boldsymbol{\Lambda})\pi(\boldsymbol{\rho}) \times \exp\left\{\sum_{i=1}^{Q} \zeta_i\right\} \prod_{i=1}^{2} \frac{\exp(-\rho_i)}{(1 + \exp(-\rho_i))^2}.$$

Then, the joint posterior density is given by

$$\begin{aligned}\pi(\boldsymbol{\vartheta} \mid \boldsymbol{D}_{\text{obs}}) \propto L(\boldsymbol{\varphi}^{-1} \mid \boldsymbol{D}_{\text{obs}}) \exp\Bigg\{ &-\frac{1}{2}\left[\sigma_b^{-2}\sum_{i=0}^{p_1} b_i^2 + \sigma_\beta^{-2}\sum_{i=1}^{p_2}\beta_i^2 + \sum_{i=1}^{Q}\frac{\exp(2\zeta_i)}{\sigma_\alpha^2}\right] \\ &+ \boldsymbol{\psi}^\top\left[\boldsymbol{\Lambda} \otimes (\boldsymbol{D}_{\boldsymbol{W}} - \boldsymbol{a}\boldsymbol{W})\right]\boldsymbol{\psi} + \log|\boldsymbol{\Lambda} \otimes \boldsymbol{a}\boldsymbol{W}| + \frac{n_0 - 4}{2}\log|\boldsymbol{\Lambda}| \\ &- \frac{1}{2}\text{tr}(\Lambda_0^{-1}\Lambda) + \sum_{i=1}^{Q}\zeta_i\Bigg\}\pi(\boldsymbol{\rho}),\end{aligned}$$

where $\boldsymbol{\varphi}^{-1} = (\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\zeta}^{-1}, \boldsymbol{U}, \boldsymbol{V})$ and $\pi(\rho_i) = 1$ if $a_i \sim \text{Uniform}(0, 1)$ and $\pi(\rho_i) = \frac{1}{B(18,2)}\frac{\exp(17\rho_i)}{(1+\exp(\rho_i))^{18}}$ if $a_i \sim \text{Beta}(18, 2)$, where $B(18, 2) = \frac{17!}{18!} = \frac{1}{18}$.

This joint posterior density is analytically intractable. So, we have based our inference on the MCMC simulation methods. No closedform is available for the conditional distributions for parameters $\boldsymbol{b}$, $\boldsymbol{\beta}$, $\boldsymbol{\zeta}$, $\boldsymbol{\psi}$ and $\boldsymbol{\rho}$. Thus, we have resorted to the Metropolis–Hastings algorithm to generate posterior samples for these parameters.

The full conditional distribution $\pi(\boldsymbol{\Lambda} \mid \boldsymbol{\vartheta}_{(-\boldsymbol{\Lambda})}, \boldsymbol{D}_{\text{obs}})$ is proportional to

$$\begin{aligned}&\pi(\boldsymbol{\psi} \mid \boldsymbol{\Lambda}, \boldsymbol{a})\pi(\boldsymbol{\Lambda}) \\ &\quad\propto |\boldsymbol{\Lambda} \otimes \boldsymbol{D}_{\boldsymbol{W}} - \boldsymbol{a}\boldsymbol{W}|^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{\psi}^\top(\boldsymbol{D}_{\boldsymbol{W}} - \boldsymbol{a}\boldsymbol{W})\boldsymbol{\psi}\right)|\boldsymbol{\Lambda}|^{(n_0-4)/2}\end{aligned}$$

$$\times \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Lambda})\right)$$

$$\propto |\boldsymbol{\Lambda}|^{(I+n_0-4)/2} \exp\left(-\frac{1}{2}\mathrm{tr}((\boldsymbol{\Lambda}_0^{-1}+\boldsymbol{B})\boldsymbol{\Lambda})\right), \tag{12}$$

where

$$\boldsymbol{B} = \begin{bmatrix} \mathrm{tr}(\boldsymbol{R_1 U}(\boldsymbol{R_1 U})^\top) & \mathrm{tr}(\boldsymbol{R_1 U}(\boldsymbol{R_2 V})^\top) \\ \mathrm{tr}(\boldsymbol{R_2 V}(\boldsymbol{R_1 U})^\top) & \mathrm{tr}(\boldsymbol{R_2 V}(\boldsymbol{R_2 V})^\top) \end{bmatrix}.$$

The full conditional distribution for $\boldsymbol{\Lambda}$ can be taken as the Wishart distribution with scalar matrix $(\Lambda_0^{-1}+B)^{-1}$ and degrees of freedom $I+n_0$. The Gibbs sampler algorithm is used to generate a posterior sample for the parameter $\boldsymbol{\Lambda}$.

3.3. *Model comparison criteria*

A variety of methodologies can be applied for the comparison of several competing models for a given data set and selecting the best one to fit the data. The DIC criterion is the one of the most used in applied works. The statistic DIC is defined as

$$\mathrm{DIC} = \bar{d} + pd,$$

where $\bar{d} = E[D_M(\boldsymbol{\varphi})]$, $pd = E[D_M(\boldsymbol{\varphi})] - D_M[E(\boldsymbol{\varphi})]$ and $D_M(\boldsymbol{\varphi})$ is the deviance function of model M defined as $-2\log L_M(\boldsymbol{\varphi})$, L_M is the likelihood function of model M. Spiegelhalter *et al.* [2002] provide evidence that pd is a suitable measure of model complexity even in hierarchical settings. Thus, DIC is a sensible generalization of the expected Akaike information criterion to hierarchical settings. Comparing alternative models, the preferred model for describing the given data set is the one with the smallest DIC value.

3.4. *Bayesian case influence diagnostics*

Since regression models are sensitive to the underlying model assumptions, generally performing a sensitivity analysis is strongly advisable.

The best known perturbation schemes are based on case deletion [Cook and Weisberg (1982)] whose effects are studied by completely removing cases from the analysis. This reasoning will form the basis for our Bayesian global influence methodology for the determination of the influential observation in the analysis.

In this work, the Bayesian case deletion influenced diagnostic measures for the joint posterior distribution based on the ψ-divergence [Peng and Dey (1995); Weiss (1996)] will be introduced in the following.

Let $D_\psi(P, P_{(-i)})$ denote the ψ-divergence between P and $P_{(-i)}$, where P denotes the posterior distribution of $\boldsymbol{\vartheta}$ for full data, and $P_{(-i)}$ denotes the posterior distribution of $\boldsymbol{\vartheta}$ without the ith case. Specifically,

$$D_\psi(P, P_{(-i)}) = \int \psi\left(\frac{\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}}^{(-i)})}{\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})}\right) \pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})\, d\boldsymbol{\vartheta}, \tag{13}$$

where ψ is a convex function with $\psi(1) = 0$. Several choices of ψ are given in Dey and Birmiwal [1994]. For example, $\psi(z) = -\log(z)$ defines Kullback–Leibler (K–L) divergence, $\psi(z) = (z-1)\log(z)$ gives J-distance (or the symmetric version of K–L divergence), $\psi(z) = 0.5|z-1|$ defines the variational distance or L_1 norm and $\psi(z) = (z-1)^2$ defines the χ^2-square divergence.

Let $\boldsymbol{\vartheta}^{(1)}, \ldots, \boldsymbol{\vartheta}^{(Q)}$ be a size Q sample of $\pi(\boldsymbol{\vartheta}|\boldsymbol{\mathcal{D}})$, $D_\psi(P, P_{(-i)})$ can be calculated numerically by

$$\widehat{D_\psi}(P, P_{(-i)}) = \frac{1}{Q}\sum_{q=1}^{Q} \psi\left(\frac{\widehat{CPO_i}}{L(y_i \mid \boldsymbol{\vartheta}^{(q)})}\right), \tag{14}$$

where $\widehat{CPO_i} = \{\frac{1}{Q}\sum_{q=1}^{Q} \frac{1}{L(y_i \mid \boldsymbol{\vartheta}^{(q)})}\}^{-1}$ is numerical approximation of the conditional predictive ordinate statistic of ith observation [Ibrahim *et al.* (2001)].

The $D_\psi(P, P_{(-i)})$ can be interpreted as the ψ-divergence of the effect of deletion of ith case from the full data on the joint posterior distribution of $\boldsymbol{\vartheta}$. As pointed out by Peng and Dey [1995] and Weiss [1996], it may be difficult for a practitioner to judge the cutoff point of the divergence measure so as to determine whether a small subset of observations is influential or not. In this work, we will use the proposal given by Peng and Dey [1995] and Weiss [1996] by considering a biased coin, which has success probability p. Then the ψ-divergence between the biased and an unbiased coin is

$$D_\psi(f_0, f_1) = \int \psi\left(\frac{f_0(x)}{f_1(x)}\right) f_1(x)dx, \tag{15}$$

where $f_0(x) = p^x(1-p)^{1-x}$ and $f_1(x) = 0.5$, $x = 0, 1$. Now, if $D_\psi(f_0, f_1) = d_\psi(p)$ then it can be easily checked that d_ψ satisfies the following equation:

$$d_\psi(p) = \frac{\psi(2p) + \psi(2(1-p))}{2}. \tag{16}$$

It is easy to note that for the divergence measures considered, d_ψ increases as p moves away from 0.5. In addition, $d_\psi(p)$ is symmetric about $p = 0.5$ and d_ψ achieves its minimum at $p = 0.5$. In this point, $d_\psi(0.5) = 0$, and $f_0 = f_1$. Therefore, if we consider $p > 0.90$ (or $p \leq 0.10$) as a strong bias in a coin, then $d_{\text{K–L}}(0.90) = 0.51$, $d_{\text{J}}(0.90) = 0.88$, $d_{L_1}(0.90) = 0.4$ and $d_{\chi^2}(0.90) = 0.64$. This equation implies that ith case is considered influential when $d_{L_1} > 0.4$ or $d_{\chi^2} > 0.64$. Thus, if we use the K–L divergence, we can consider an influential observation when $d_{\text{K–L}} > 0.51$. Analogously, using the K–L divergence and the J-distance, we can consider an influential observation when $d_{\text{K–L}} > 0.51$ and $d_{\text{J}} > 0.88$, respectively.

4. Simulation Study

In this section, we present some simulation studies for PH cure rate model ($Q = 1$) under first and last activations with the dependent assumption in order to examine their performances. The interval-censored survival times $(t_{ijL}, t_{ijR}, \delta_{ij})$ with the

cure fraction under the first and last activations are generated in a manner similar to that employed by Yau and Ng [2001] with some modifications.

More specifically, for the jth individual in the ith region, $j = 1, \ldots, n_i$, $i = 1, \ldots, I$, the initial number of competing causes related to the event, M_{ij} is generated from a geometric distribution with parameter $p_{0ij} = [1 + \exp(-(b_0 + b_1)x_{ij} + v_i)]^{-1}$, where covariate x_{ij} follows Bernoulli (0.5) distribution. Interval-censored data $(t_{ijL}, t_{ijR}, \delta_{ij})$ are then generated as follows:

(i) If $M_{ij} = 0$, then let $t_{ij} = t_{ijL}$ from the exponential distribution with hazard rate 10 and let censoring indicator $\delta_{ij} = 0$.

(ii) If $M_{ij} > 0$, then we generate Y_{cij} $(i = 1, \ldots, I,\ j = 1, \ldots, n_i$ and $c = 1, \ldots, M_{ij})$, from exponential distribution with hazard rate $\alpha\lambda_{ij} = \exp(\beta x_{ij} + u_i)$.

Let t_{ij} take the lowest generated variable in the case of variables of model generated under first activation and t_{ij} take the largest generated variable in the case of variables of model generated under last activation. The censoring variable c_{ij} is generated from $U(0, cc)$, $cc > 0$ is fixed to control the percentage of censored data. Let $\delta_{ij} = 1$ if $t_{ij} \leq c_{ij}$ and $\delta_{ij} = 0$ otherwise.

(iii) For $\delta_{ij} = 0$, let $0 < t_{ijL} < t_{ijR} = \infty$.

(iv) For $\delta_{ij} = 1$, we generate len_{ij} from distribution $U(0.2, 0.7)$ and l_{ij} from $U(0, 0.01)$. Then, from $(0, l_{ij}], (l_{ij}, l_{ij} + \text{len}_{ij}], \ldots, (l_{ij} + k\text{len}_{ij}, \infty]$, $k = 1, 2, \ldots, (t_{ijL, t_{ijR}}]$ is chosen as that satisfying $t_{ijL} < t_{ij} \leq t_{ijR}$.

We considered $I = 5$ regions (Zip) whose corresponding adjacent matrix is given by $\begin{bmatrix} 0&0&1&0&0 \\ 0&0&0&0&1 \\ 1&0&0&1&0 \\ 0&0&1&0&0 \\ 0&1&0&0&0 \end{bmatrix}$. The random effects U_i and U_i are generated from normal distribution with mean $\mathbf{0}$ and precision matrix $\boldsymbol{\Lambda} \bigotimes (\boldsymbol{D_W} - a\boldsymbol{W})$, where $\boldsymbol{W}$ is standardized adjacent matrix so that each of its rows sum to one, $D_W = \text{Diag}(1, 1, 2, 1, 1)$ is a diagonal matrix and we fixed $a = 0.9$ and $\Lambda = \text{Diag}(4, 4)$, i.e. we fixed $\Lambda_{11} = 4$, $\Lambda_{22} = 4$ and $\Lambda_{12} = \Lambda_{21} = 0$. We consider 100 individuals and the number of Zip was distributed for each individual using sample with replacement. So, the number of individuals in each region n_i, $i = 1, \ldots, 5$ is varied, that is, the five regions could have different numbers of individuals with $\sum_{i=1}^{5} n_i = 100$. Thus, we consider the sample size $n = 100$ and we fix parameters $b_0 = -1.50$, $b_1 = -0.50$, $\beta = -0.15$, $\alpha = 0.30$. For each model, we consider 500 repeated samples each one with around 40% censored data. The priors for the parameters b_0, b_1, β_1 and α used in the study are $b_0 \sim N(0, 100)$, $b_1 \sim N(0, 100)$, $\beta_1 \sim N(0, 100)$ and $\alpha \sim N(0, 100)\boldsymbol{I}_{(0,\infty)}$.

For each generated data set, we consider 1,000 sample burn-in and use every third sample from the 10,000 MCMC posterior samples to reduce the autocorrelations and yield better convergence results, thus obtaining an effective sample of size

3,000 upon which the posterior is based on. The average bias (Bias), standard deviation (SD) of the estimate, average standard deviation (SD mean) and mean square error (MSE) were obtained to evaluate the performance of the parameter estimates for PH cure rate model under the first and last activations. The summaries are presented in Table 1.

We can note that the bias and MSE of the parameter Λ_{12} are larger than others in all fitting models. The estimator of Λ_{12} present negative biases for the cure model under the first activation and it presents positive biases for the cure model under the last activation, however, its biases and MSEs are always near zero. Moreover, for both mechanism activations, the simulation results for the cure models considering prior 1 are very close to those obtained considering prior 2.

4.1. *Influence of outliers*

One of our main goals in this study is to show the need for robust models to deal with the presence of outliers in the data. Considering the same parameter values and setup as above and two cases for perturbation, eight data sets of size 100 were generated from the Weibull and PH cure models under first and last activations with dependent spatial frailties.

We selected cases 18 and 80 for perturbation. To create an influential observation in the data set, we choose one or two of these selected cases and perturbed the response variable as follows: $t_{kL} = t_{kL} + 10S_L$ and $t_{kR} = t_{kR} + 10S_L$, for $k = 18$ and 80, where S_L is the standard deviations of the t_{ijL}'s. Note that using this kind of perturbation, the interval of observed interval time of perturbation candidate observation is not charged. Here, we consider four setups in the study. Setup A: original data set, without outliers; Setup B: data with outlier 18; Setup C: data with outlier 80; and Setup D: data with outliers 18 and 80. The MCMC computations are made similar to those in Sec. 5 and further to monitoring the convergence of the Gibbs samples, we use the Geweke's convergence diagnostic [Geweke (1992)].

Tables 2 and 3 report posterior mean, standard deviation (SD), bias and mean square error (MSE) of the parameters of PHCRM-FA and PHCRM-LA, respectively. For PH cure rate model under the first activation, Table 2 shows that parameter Λ_{11} is litter sensitive to perturbations. The posterior mean of Λ_{11} decreases in the perturbation cases when considering prior 1 or prior 2 for the parameters and it is more sensitive using prior 1 than prior 2. For PH cure rate model under the last activation, considering prior 1, Λ_{11} is litter sensitive in cases C and D and Λ_{12} is sensitive in case B; considering prior 2, Λ_{11} is litter sensitive in cases B and C and Λ_{12} is sensitive in cases B and D. This results can be observed in Table 3.

Now, we consider the sample from the posterior distributions of the parameters of PHC-FA and PHCM-LA for each simulated data set to calculate the ψ-divergence measures (d_{KL}, d_J, d_{L_1}, d_{χ^2}) described in Sec. 3.4 are computed. The results in Table 4 show, before perturbation (data set A), that all the selected cases are not influential according to all ψ-divergence measures. However, after perturbation

Table 1. Simulation results for PH cure models under first and last activations with dependent spatial frailties.

Parameter	True value	Estimate mean	SD of the estimate	Bias	MSE	SDs mean
First activation						
Prior 1. $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.9, 1))$						
b_0	-1.50	-1.6441	0.0515	-0.1441	0.0234	0.2710
b_1	-0.50	-0.5215	0.1179	-0.0215	0.0143	0.2482
β	-0.15	-0.1538	0.0479	-0.0038	0.0023	0.1877
α	1.00	1.1920	0.0382	0.1920	0.0370	0.1396
Λ_{11}	4.00	4.2224	0.1479	0.2224	0.0713	2.5349
Λ_{22}	4.00	3.9272	0.1801	-0.0728	0.0377	2.5894
Λ_{12}	0.00	-0.4142	0.1420	-0.4142	0.1917	1.9325
a	0.90	0.8999	0.0015	-0.0001	0.0000	0.0655
Prior 2. $\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.9, 1))$						
b_0	-1.50	-1.6418	0.0464	-0.1418	0.0222	0.2710
b_1	-0.50	-0.5146	0.1231	-0.0146	0.0153	0.2491
β	-0.15	-0.1552	0.0502	-0.0052	0.0025	0.1875
α	1.00	0.8980	0.0612	-0.1020	0.1900	0.0100
Λ_{11}	4.00	4.2411	0.1471	0.2411	0.0797	2.5337
Λ_{22}	4.00	3.9259	0.1915	-0.0741	0.0421	2.5853
Λ_{12}	0.00	-0.4153	0.1437	-0.4153	0.1931	1.9307
a_1	0.90	0.9001	0.0016	0.0001	0.0000	0.0655
a_2	0.90	0.9000	0.0016	0.0000	0.0000	0.0653
Last activation						
Prior 1. $\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.75, 1))$						
b_0	-1.50	-1.6533	0.0852	-0.1533	0.0308	0.2635
b_1	-0.50	-0.5056	0.0998	-0.0056	0.0100	0.2652
β	-0.15	-0.1298	0.0933	0.0202	0.0091	0.1323
α	1.00	0.9090	0.0408	-0.0910	0.1770	0.0850
Λ_{11}	4.00	4.2564	0.2096	0.2564	0.1096	2.3576
Λ_{22}	4.00	3.7852	0.3098	-0.2148	0.1420	2.5456
Λ_{12}	0.00	0.3803	0.1762	0.3803	0.1756	1.7807
a	0.90	0.9001	0.0016	0.0001	0.0000	0.0653
Prior 2. $\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda}_0 \sim \text{Wishart}(2, \text{Diag}(0.75, 1))$						
b_0	-1.50	-1.6518	0.0867	-0.1518	0.0306	0.2657
b_1	-0.50	-0.5190	0.1084	-0.0190	0.0121	0.2656
β	-0.15	-0.1373	0.1006	0.0127	0.0103	0.1325
α	1.00	0.9200	0.0407	-0.0800	0.1800	0.0600
Λ_{11}	4.00	4.0064	0.2029	0.0064	0.0411	2.2193
Λ_{22}	4.00	3.7107	0.3010	-0.2893	0.1741	2.4853
Λ_{12}	0.00	0.3619	0.1646	0.3619	0.1580	1.7052
a_1	0.90	0.9002	0.0015	0.0002	0.0000	0.0654
a_2	0.90	0.8996	0.0017	-0.0004	0.0000	0.0655

(data set, A–D), the measures increase, indicating that the perturbed cases are influential.

In Figs. 1 and 2, we have depicted the ψ-divergence measure (J-distance) with prior 1 for the cases (A)–(D) for PHCRM-FA and PHCRM-FA, respectively. Clearly,

Table 2. Simulation results of the perturbed cases for PH cure models under first activation.

PH cure rate model-first activation											
		Prior 1					Prior 2				
Setup	Perturbed case	Parameters	Mean	SD	Bias	MSE	Parameters	Mean	SD	Bias	MSE
A	None	b_0	−1.456	0.270	0.044	0.002	b_0	−1.670	0.267	−0.170	0.029
		b_1	−0.431	0.249	0.069	0.005	b_1	−0.611	0.241	−0.111	0.012
		β	−0.176	0.182	−0.026	0.001	β	−0.156	0.177	−0.006	0.000
		α	1.192	0.137	0.192	0.037	α	0.898	0.190	−0.102	0.010
		Λ_{11}	4.275	2.500	0.275	0.075	Λ_{11}	3.840	2.348	−0.160	0.026
		Λ_{22}	3.927	2.568	−0.073	0.005	Λ_{22}	4.058	2.688	0.058	0.003
		Λ_{12}	−0.363	1.951	−0.363	0.132	Λ_{12}	−0.637	1.917	−0.637	0.405
		a	0.898	0.070	−0.002	0.000	a_1	0.901	0.064	0.001	0.000
							a_2	0.902	0.064	0.002	0.000
B	{18}	b_0	−1.348	0.278	0.152	0.023	b_0	−1.387	0.250	0.113	0.013
		b_1	−0.436	0.233	0.064	0.004	b_1	−0.337	0.251	0.163	0.026
		β	−0.207	0.184	−0.057	0.003	β	−0.193	0.191	−0.043	0.002
		α	0.912	0.095	−0.088	0.008	α	0.797	0.205	−0.203	0.041
		Λ_{11}	3.137	2.021	−0.863	0.745	Λ_{11}	3.240	2.181	−0.760	0.577
		Λ_{22}	4.050	2.661	0.050	0.003	Λ_{22}	4.302	2.828	0.302	0.091
		Λ_{12}	−0.300	1.776	−0.300	0.090	Λ_{12}	−0.161	1.867	−0.161	0.026
		a	0.903	0.065	0.003	0.000	a_1	0.899	0.066	−0.001	0.000
							a_2	0.899	0.065	−0.001	0.000

Table 2. (*Continued*)

PH cure rate model-first activation										
		Prior 1					Prior 2			
Setup	Perturbed case	Parameters	Mean	SD	Bias	MSE	Parameters	Mean	SD	Bias	MSE
		b_0	−1.494	0.253	0.006	0.000	b_0	−1.336	0.273	0.164	0.027
		b_1	−0.643	0.248	−0.143	0.020	b_1	−0.193	0.247	0.307	0.094
		β	−0.183	0.186	−0.033	0.001	β	−0.148	0.193	0.002	0.000
		α	1.112	0.112	0.112	0.012	α	0.866	0.198	−0.134	0.018
C	{80}	Λ_{11}	3.101	2.049	−0.899	0.809	Λ_{11}	3.419	2.174	−0.581	0.337
		Λ_{22}	4.221	2.637	0.221	0.049	Λ_{22}	3.600	2.468	−0.400	0.160
		Λ_{12}	−0.532	1.831	−0.532	0.283	Λ_{12}	−0.263	1.756	−0.263	0.069
		a	0.900	0.065	0.000	0.000	a_1	0.899	0.067	−0.001	0.000
							a_2	0.900	0.065	0.000	0.000
		b_0	−1.498	0.276	0.002	0.000	b_0	−1.526	0.277	−0.026	0.001
		b_1	−0.551	0.249	−0.051	0.003	b_1	−0.343	0.243	0.157	0.025
		β	−0.136	0.188	0.014	0.000	β	−0.302	0.185	−0.152	0.023
		α	1.029	0.099	0.029	0.001	α	0.761	0.216	−0.239	0.057
D	{18, 80}	Λ_{11}	2.821	1.858	−1.179	1.389	Λ_{11}	3.190	2.045	−0.810	0.656
		Λ_{22}	4.276	2.734	0.276	0.076	Λ_{22}	4.160	2.698	0.160	0.026
		Λ_{12}	−0.428	1.791	−0.428	0.183	Λ_{12}	−0.661	1.799	−0.661	0.437
		a	0.905	0.067	0.005	0.000	a_1	0.900	0.064	0.000	0.000
							a_2	0.898	0.067	−0.002	0.000

Table 3. Simulation results of the perturbed cases for PH cure rate models under last activation.

PH cure rate model-last activation											
		Prior 1					Prior 2				
Setup	Perturbed case	Parameters	Mean	SD	Bias	MSE	Parameters	Mean	SD	Bias	MSE
A	None	b_0	−1.679	0.268	−0.179	0.032	b_0	−1.840	0.290	−0.340	0.116
		b_1	−0.492	0.271	0.008	0.000	b_1	−0.600	0.263	−0.100	0.010
		β	−0.226	0.132	−0.076	0.006	β	−0.177	0.119	−0.027	0.001
		α	0.909	0.177	−0.091	0.008	α	0.920	0.180	−0.080	0.006
		Λ_{11}	4.258	2.370	0.258	0.067	Λ_{11}	4.351	2.449	0.351	0.123
		Λ_{22}	3.880	2.560	−0.120	0.014	Λ_{22}	2.980	2.239	−1.020	1.040
		Λ_{12}	0.002	1.774	0.002	0.000	Λ_{12}	−0.054	1.594	−0.054	0.003
		a	0.902	0.066	0.002	0.000	a_1	0.901	0.066	0.001	0.000
							a_2	0.902	0.065	0.002	0.000
B	{18}	b_0	−1.678	0.267	−0.178	0.032	b_0	−1.646	0.263	−0.146	0.021
		b_1	−0.599	0.252	−0.099	0.010	b_1	−0.369	0.266	0.131	0.017
		β	−0.384	0.124	−0.234	0.055	β	−0.230	0.138	−0.080	0.006
		α	0.918	0.179	−0.082	0.007	α	0.789	0.194	−0.211	0.045
		Λ_{11}	4.124	2.242	0.124	0.015	Λ_{11}	3.180	1.937	−0.820	0.672
		Λ_{22}	3.743	2.440	−0.257	0.066	Λ_{22}	3.798	2.552	−0.202	0.041
		Λ_{12}	0.259	1.742	0.259	0.067	Λ_{12}	0.420	1.745	0.420	0.176
		a	0.901	0.064	0.001	0.000	a_1	0.899	0.068	−0.001	0.000
							a_2	0.900	0.066	0.000	0.000

Table 3. (*Continued*)

PH cure rate model-last activation											
		Prior 1					Prior 2				
Setup	Perturbed case	Parameters	Mean	SD	Bias	MSE	Parameters	Mean	SD	Bias	MSE
		b_0	−1.634	0.262	−0.134	0.018	b_0	−1.716	0.271	−0.216	0.047
		b_1	−0.521	0.260	−0.021	0.000	b_1	−0.401	0.263	0.099	0.010
		β	−0.231	0.123	−0.081	0.007	β	−0.040	0.132	0.110	0.012
		α	0.903	0.180	−0.097	0.009	α	0.850	0.184	−0.150	0.023
C	{80}	Λ_{11}	3.972	2.205	−0.028	0.001	Λ_{11}	3.464	1.970	−0.536	0.287
		Λ_{22}	3.928	2.642	−0.072	0.005	Λ_{22}	3.779	2.531	−0.221	0.049
		Λ_{12}	0.130	1.793	0.130	0.017	Λ_{12}	−0.040	1.623	−0.040	0.002
		a	0.901	0.066	0.001	0.000	a_1	0.900	0.066	0.000	0.000
							a_2	0.902	0.063	0.002	0.000
		b_0	−1.643	0.275	−0.143	0.021	b_0	−1.515	0.254	−0.015	0.000
		b_1	−0.444	0.270	0.056	0.003	b_1	−0.305	0.246	0.195	0.038
		β	−0.041	0.134	0.109	0.012	β	−0.161	0.137	−0.011	0.000
		α	0.770	0.216	−0.230	0.053	α	0.806	0.192	−0.194	0.037
D	{18, 80}	Λ_{11}	2.547	1.553	−1.453	2.111	Λ_{11}	3.708	2.101	−0.292	0.085
		Λ_{22}	3.958	2.529	−0.042	0.002	Λ_{22}	4.303	2.689	0.303	0.092
		Λ_{12}	−0.038	1.707	−0.038	0.001	Λ_{12}	0.350	1.774	0.350	0.122
		a	0.901	0.065	0.001	0.000	a_1	0.899	0.067	−0.001	0.000
							a_2	0.901	0.064	0.001	0.000

Table 4. Divergence measures of the perturbed cases and DIC values for the simulated data sets.

Activation	Prior	Setup	Case number	$d_{\mathrm{K-L}}$	d_J	d_{L1}	d_{χ^2}	DIC
First	1	A	18	0.016	0.032	0.071	0.034	228.007
			80	0.036	0.073	0.106	0.081	
		B	18	1.644	4.288	0.693	168.408	263.195
		C	80	1.265	4.761	0.649	584.824	241.465
		D	18	1.750	4.472	0.718	84.141	240.257
			80	0.479	1.248	0.402	8.804	
	2	A	18	0.002	0.004	0.026	0.004	211.243
			80	0.013	0.027	0.065	0.029	
		B	18	1.455	3.501	0.662	35.776	278.356
		C	80	1.438	3.271	0.660	22.489	284.756
		D	18	0.462	1.097	0.390	3.959	290.520
			80	0.460	1.162	0.390	5.777	
Last	1	A	18	0.018	0.037	0.076	0.040	387.037
			80	0.025	0.050	0.086	0.057	
		B	18	2.727	5.539	0.788	73.127	427.090
		C	80	3.659	7.131	0.845	110.592	452.871
		D	18	2.087	4.366	0.726	34.092	483.157
			80	4.358	9.412	0.900	648.576	
	2	A	18	0.011	0.023	0.059	0.023	423.219
			80	0.025	0.052	0.090	0.057	
		B	18	3.790	7.791	0.870	199.173	446.544
		C	80	3.347	6.929	0.829	164.743	470.571
		D	18	3.880	8.238	0.870	365.590	477.539
			80	3.829	8.759	0.883	595.102	

we can see that the measure performed well in identifying influential case(s), providing larger ψ-divergence measure when compared to the other cases.

5. Application

We illustrate the proposed method to the interval-censored smoking cessation data. In this data set, all of the patients (smokers) in the study were randomized into either a smoking intervention (SI) group or a usual care (UC) group which received no special anti-smoking intervention. The treatments of smoking intervention program were realized in Rochester city localizing in the center of the maps. More details of the program can be found in Murray *et al.* [1998]. Here, each patient was observed once a year over the five-year follow-up. Our event of interest is whether they relapse (resume smoking) or not. If a smoker resumed smoking after an initial attempt to quit, then only an approximate one-year time interval was observed from the previous observation to the current observation. Thus, the relapse times are interval-censored. In this analysis, we limit our attention to those patients who are known to have quit smoking at least once during the study period and who have an identifiable Minnesota zip code of residence. The data include 223 patients who reside in 51 zip codes in the southeastern corner of Minnesota, among which

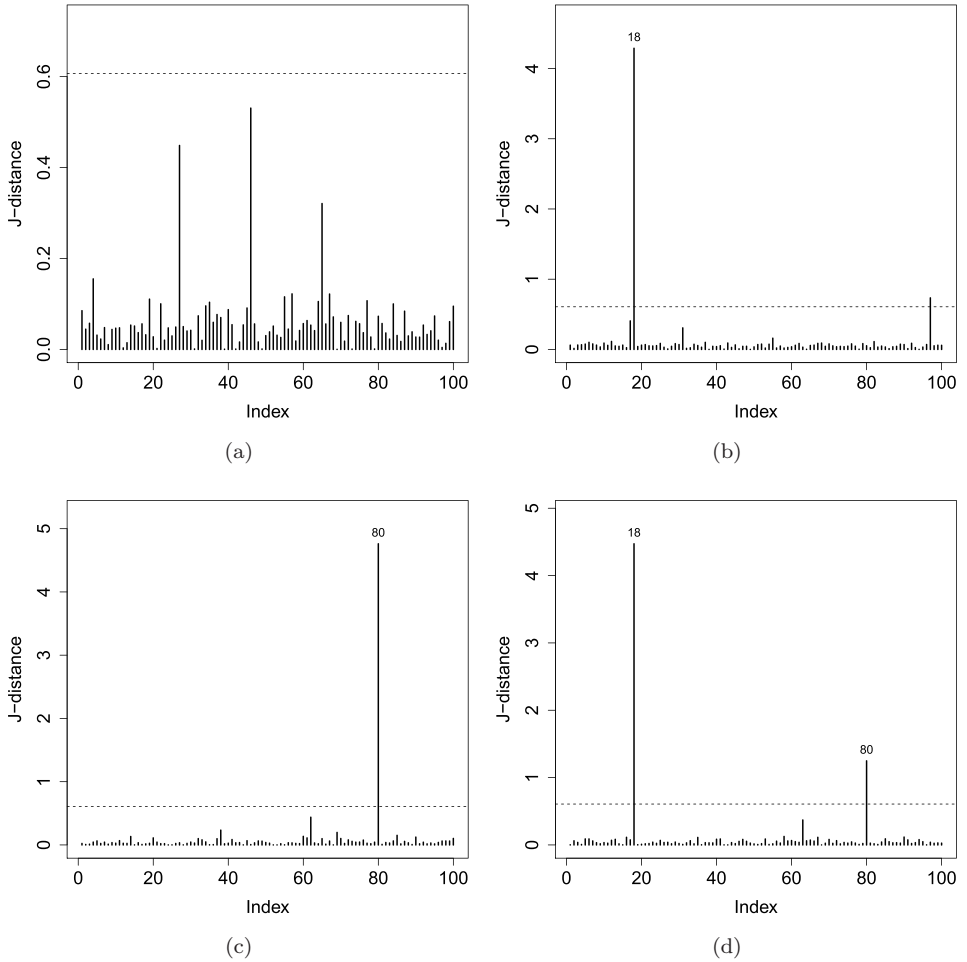

Fig. 1. Index plots of ψ-divergence measure from the fitted PHCRM-FA for cases (A)–(D).

65 patients to have undergone relapse, which implies the empirical cure rate is approximately 71%.

We fitted the PH cure model under the first and last activations, considering the different spatial frailties in the models to the data set. As we know, the piecewise exponential distribution has a better approximation to any unknown function when the length of each interval becomes smaller. Therefore, we partition the time axis so that they denote the ordered distinct time points of all observed interval endpoints. Thus, there are 178 parameters to be estimated. The prior distributions for the parameters $\boldsymbol{b}$, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{178})$ are as follows:

- $b_j \sim N(0, 100)$, $j = 0, \ldots, 4$;
- $\beta_j \sim N(0, 100)$, $j = 1, \ldots, 4$;
- $\alpha_i \sim N(0, 100) I_{(0,\infty)}$, $i = 1, \ldots, 178$.

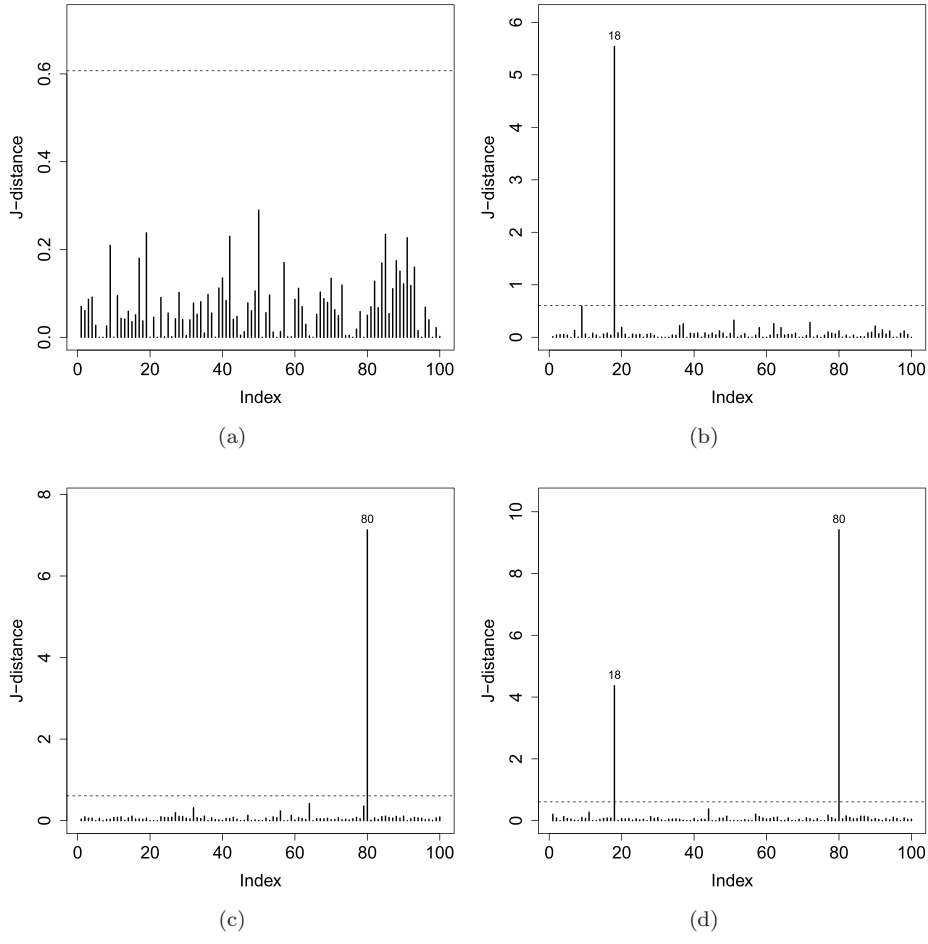

Fig. 2. Index plots of ψ-divergence measure from the fitted PHCRM-LA for cases (A)–(D).

Because of the high computational cost, we implement the MCMC in the algorithms with C language and the results were analyzed in R language [R Development Core Team (2010)] through the "coda" package [Plummer *et al.* (2005)]. For the models with the vague priors, MCMC algorithms ran a total of 60,000 iterations, discarding the first 20,000 realizations as burn-in and thinning to every fifth iteration, and for the models with the informative priors, MCMC algorithms ran a total of 30,000 iterations discarding the first 20,000 realizations as burn-in and thinning to every second iteration. Posterior results are then based on 5,000 realizations of the Markov chain. Our Metropolis acceptance rate for these parameters ranged from 25% to 50%. The convergence was checked using the Geweke diagnostic which did not indicate lack of convergence.

Table 5 provides the DIC scores for a variety of effects of PH cure model under the first activation and last activation. The DIC scores of the models under first and

Table 5. Real data set. Bayesian criteria for the PH cure model.

Model	Priors	Criteria DIC	Criteria *pd*
First activation			
1	$\boldsymbol{U} \sim \text{CAR}(\theta_1)$, $\boldsymbol{V} \sim \text{CAR}(\theta_2)$, $\theta_1, \theta_2 \sim \text{InvGamma}(0.01, 0.01)$	395.7928	9.5501
2	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	396.2012	10.6375
3	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	395.7993	12.1205
4	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	394.7907	11.5739
5	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	397.7197	11.6590
Last activation			
6	$\boldsymbol{U} \sim \text{CAR}(\theta_1)$, $\boldsymbol{V} \sim \text{CAR}(\theta_2)$, $\theta_1, \theta_2 \sim \text{InvGamma}(0.01, 0.01)$	395.9287	10.2954
7	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	395.3409	12.2265
8	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Uniform}(0, 1)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	395.3662	12.1625
9	$\boldsymbol{\psi} \sim \text{MCAR}(a, \boldsymbol{\Lambda})$, $a \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	394.5312	11.7031
10	$\boldsymbol{\psi} \sim \text{MCAR}(a_1, a_2, \boldsymbol{\Lambda})$, $a_1, a_2 \sim \text{Beta}(18, 2)$, $\boldsymbol{\Lambda} \sim \text{Wishart}(2, \text{Diag}(0.1, 0.1))$	394.8002	11.7053

last activations are closely related, this indicated these models are almost equivalent. In what follows, we present results for model 9 with low DIC scores, but emphasize that virtually any of the models in Table 5 could be used with equal confidence.

We compared the results with those obtained by the Bayesian semi-parametric model with the spatial frailty proposed by Pan *et al.* [2014] and with the parametric cure rate model under first and last activations proposed by Bao *et al.* [2017] which assume that baseline hazard function in (7), $h_0(t|\alpha) = \alpha t^{\alpha-2}$, in this case hazard function is given by $h(t|\alpha, \lambda_{ij}) = \alpha t^{\alpha-2} \exp(\lambda_{ij})$, which is the hazard function of two-parameter Weibull distribution with shape parameter $\alpha > 0$ and scale parameter $\exp(\lambda_{ij})$. Comparing the obtained DIC scores with the DIC values presented in the papers of Pan *et al.* [2014] and Bao *et al.* [2017], it is shown that both PHCMs have smaller DIC values. Here, we select model 9 which has the smallest DIC as our working model.

Table 6 presents posterior means, standard deviations and 95% highest posterior density (HPD) intervals of the parameter of model 9. We can note that all covariates are not significant when we consider the (95%) credibility interval, but the covariate of the "intervention type SI/UC" in the cure rate and in the survival model will become significant since we consider the lower credibility interval. In cure rate, the sign of the parameters b_2 and b_3 are the same as the results above, which means that the individuals with a higher level of cigarette consumption have lower

Table 6. Posterior summaries of the parameter of model 9 for the smoking cessation data.

		Survival model					Cure rate			
Parameter		Mean	SD	2.5%	97.5%		Mean	SD	2.5%	97.5%
Intercept						b_0	0.0072	0.8279	−1.6658	1.5794
Sex (male = 0)	β_1	0.1748	0.2694	−0.3421	0.7128	b_1	−0.3696	0.4800	−1.2913	0.5993
SI/UC (UC = 0)	β_2	−0.1392	0.3107	−0.7249	0.4876	b_2	0.8275	0.5500	−0.2217	1.9833
Cigarettes per day	β_3	−0.0168	0.0234	−0.0602	0.0267	b_3	−0.0499	0.0470	−0.1241	0.0606
Duration as smoker	β_4	−0.0292	0.0259	−0.0787	0.0178	b_4	0.0306	0.0600	−0.1055	0.1202
a		0.8981	0.0668	0.7349	0.9879					
Λ_{11}		2.6680	0.6501	1.5594	4.0684					
Λ_{22}							2.5805	0.6427	1.4840	4.0155
Λ_{12}		−0.0085	0.4670	−0.9244	0.9317					
Σ_{11}		0.4130	0.1084	0.2511	0.6668					
Σ_{22}							0.4289	0.1192	0.2537	0.7187
$\Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$		0.0028	0.1820	−0.3645	0.3524					

Note: Λ_{ij} is the element of precision matrix $\boldsymbol{\Lambda}$ in position (i, j), and Σ_{ij} is the element of matrix $\boldsymbol{\Sigma} = \Lambda^{-1}$ in position (i, j), this Σ_{11} is the spatial variance component of $\boldsymbol{U}$ and Σ_{22} is the spatial variance component of $\boldsymbol{V}$, $\Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$ denote their correlation.

probability of quiting smoking and the individuals with special intervention have higher probability of quiting smoking than those with usual care. In the survival function, the negative value of β_2 implies that individuals with special intervention have a lower hazard rate of the relapse time than those with usual care.

The estimates of the spatial variance component of $\boldsymbol{U}$ in the survival model (Σ_{11}) is 0.4130, and the spatial variance component of $\boldsymbol{V}$ in the cure rate (Σ_{22}) is 0.4289, which indicate that there is considerable heterogeneity among the clusters. Moreover, it is observed that there are no correlations between the spatial effects $\boldsymbol{U}$ and $\boldsymbol{V}$.

The posterior means and 95% HPD intervals of α_i's are presented in Fig. 3, it is shown that there are some values of α's, which are indicated as having different values of others. Based on this figure, we can repartition the time axis so that we consider just risk parameters (α_1, α_2, α_{34}, α_{35}, α_{64}, α_{65}, α_{84}, α_{85}, α_{88}, α_{98}, α_{99}), thus just 11 parameters need to be estimated, it will lower computational time cost.

Figure 4 maps the posterior means of frailties $\boldsymbol{U}$ and $\boldsymbol{V}$ in PHCRM-LA. For the frailties $\boldsymbol{U}$, the result is closer to that obtained in model 15, but the frailties $\boldsymbol{V}$ show that almost all of the regions have closed cure probabilities. The corresponding posterior standard deviations presented in Fig. 5 have results similar to model 15 as well.

Considering the samples of the posterior distribution of the parameters of model 9, the ψ-divergence measures are computed to detect possible influential

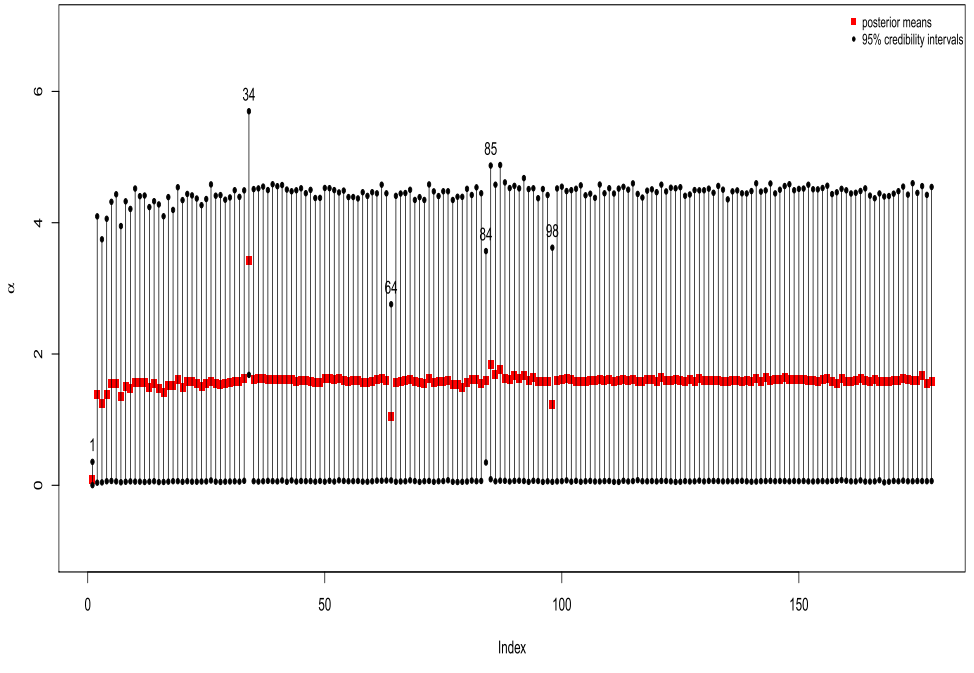

Fig. 3. Posterior means and credibility intervals of α_i's.

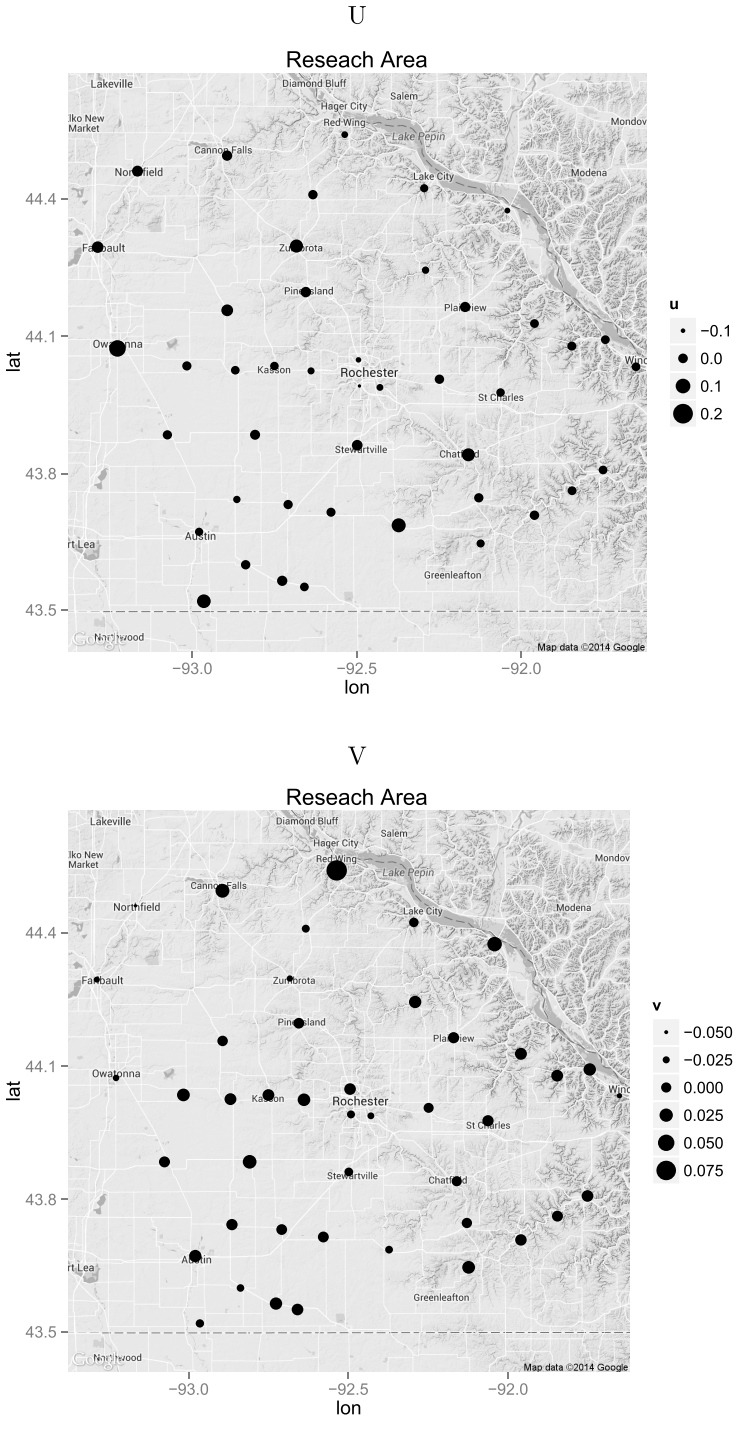

Fig. 4. Maps of posterior means for frailties $\boldsymbol{U}$ (top row) and $\boldsymbol{V}$ (bottom row).

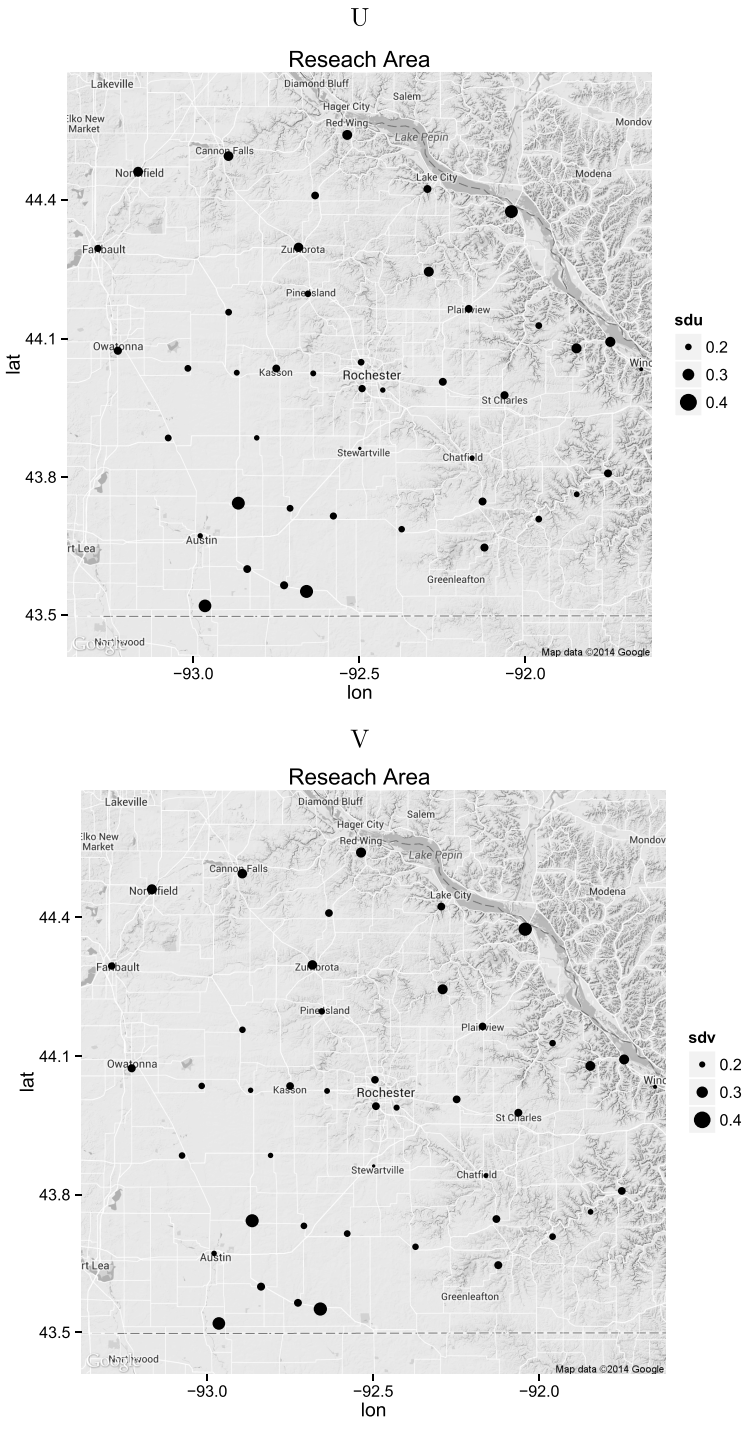

Fig. 5. Maps of posterior standard derivations for frailties $\boldsymbol{U}$ (top row) and $\boldsymbol{V}$ (bottom row).

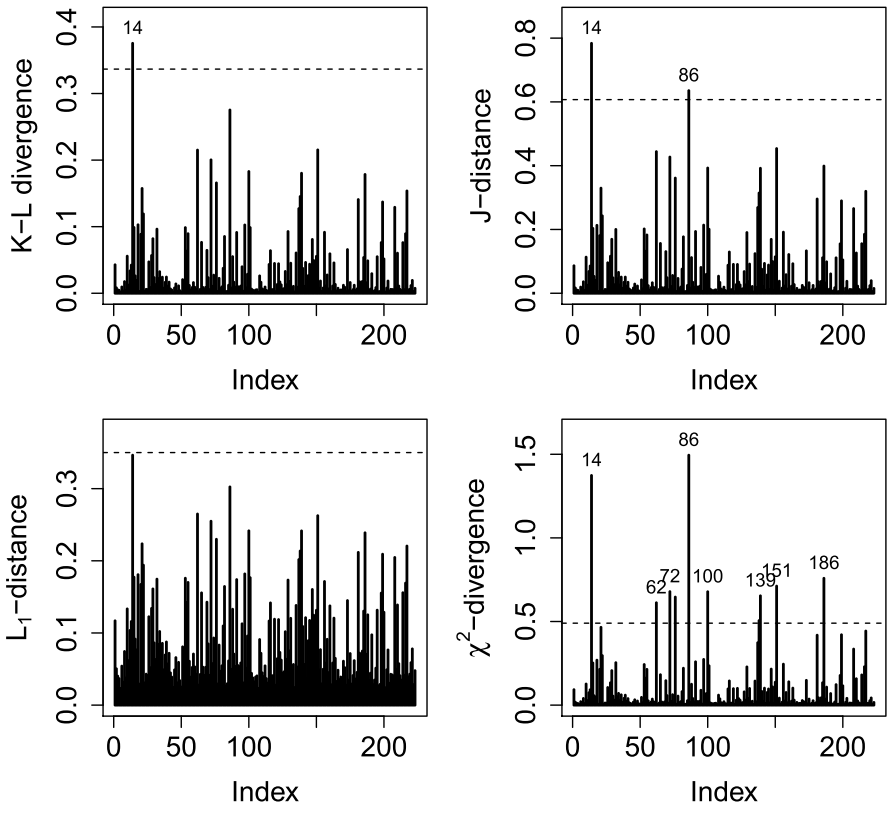

Fig. 6. Estimates of ψ-divergence measures for model 9.

observations in the posterior distribution of the parameters of model 9 and presented in Fig. 6, which shows that there are some possible influential observations were detected by divergence measures, but they are different from the observations which were detected previously. Here, we will just analyze individuals 14 and 86 which were detected by both J-distance and χ^2-divergence measure. In Table 7, we can note that both individuals had special interventions but relapse occurred. In order to reveal the impact of this possible influential observation on the parameter estimates and inference, we removed this observation and readjusted the model. Note that, in the piecewise exponential model, the time axis is partitioned by the ordered distinct time points of all observed interval endpoints, thus we have different and less risk parameter α's after removing the observations.

Table 7. Possible influential observations detected by the divergence measures.

Obs.	Sex	Duration	Intervention	Num. cigarettes	Relapse	Time interval	Zip
14	0	32	1	60	1	(1.035, 4.211)	55904
86	1	24	1	40	1	(3.885, 5.073)	55987

Table 8. Posterior summaries of the parameter of model 9 and RV adjusted for the smoking cessation data without detected individuals 14 and 86.

		Survival model					Cure rate			
Parameter		Mean	SD	2.5%	97.5%		Mean	SD	2.5%	97.5%
Intercept						b_0	−0.0358 (−5.9465)	0.8369	−1.6263	1.5895
Sex (male = 0)	β_1	0.1934 (0.1064)	0.2708	−0.3352	0.7389	b_1	−0.3168 (−0.1428)	0.4536	−1.1974	0.5985
SI/UC (UC = 0)	β_2	−0.2241 (0.6097)	0.3002	−0.7793	0.4069	b_2	0.6060 (−0.2677)	0.4968	−0.3488	1.5578
Cigarettes per day	β_3	−0.0334 (0.9819)	0.0193	−0.0709	0.0024	b_3	−0.0862 (0.7276)	0.0298	−0.1425	−0.0232
Duration as smoker	β_4	−0.0109 (−0.6254)	0.0206	−0.0479	0.0332	b_4	0.0708 (1.3100)	0.0350	0.0000	0.1313
a		0.9038 (0.0063)	0.0619	0.7568	0.9864					
Λ_{11}		2.6738 (0.0021)	0.6573	1.5785	4.1271					
Λ_{22}							2.5637 (−0.0065)	0.6365	1.4845	3.9550
Λ_{12}		−0.0063 (−0.2618)	0.4685	−0.9134	0.9258					
Σ_{11}		0.4126 (−0.0011)	0.1085	0.2483	0.6684					
Σ_{22}							0.4320 (0.0073)	0.1221	0.2569	0.7200
Σ_{12}		0.0029 (0.0349)	0.1830	−0.3552	0.3599					

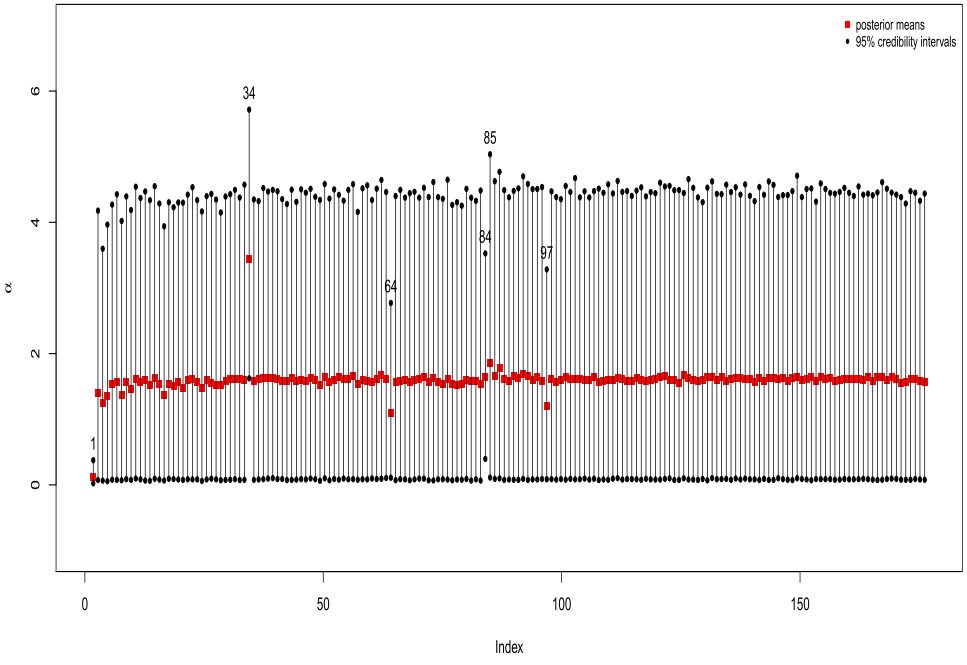

Fig. 7. Posterior means and credibility intervals of α_i's for the data set without individuals 14 and 86.

The posterior summaries of the parameters for the readjusted model 9 and RV for the posterior mean of the parameters are presented in Table 8. We can note that just the posterior means of the parameters b_0 and β_4 have relative higher variations, but both have values close to zero. All parameters had not changed except b_0. The posterior means and 95% credibility intervals of α_i's are presented in Fig. 7, it is shown that the values of α's are similar to the estimates in Fig. 3. In this case, we do not have inferential changes after removing the observations. So, this model is not sensitive to influential observations. The value of DIC for fitted models is 385.0749, which is lower than model 9 for the data without removing the detected observations.

6. Conclusions

In this work, we proposed a new semi-parametric cure rate proportional odds models which allow spatial correlations by including spatial frailty for the interval-censored data setting. The proposed cure rate models are very flexible because they encompass several known cure rate models as the particular cases. A Bayesian approach was described for estimating the model parameters. We provide some simulation results to assess the performance of the proposed models. An application to real data is presented in order to illustrate the flexibility of the proposed models. Moreover, the proposed models are not sensitive to influential observations, which can be

observed through the influence diagnostic in the simulation study as well as in the application. The interpretation of the covariates is easy due to the parametrization of the models considered in the cure rate. Furthermore, the MCAR prior can be used even if frailties effects are low or they are not correlated.

References

Banerjee, S. and Carlin, B. P. (2004). Parametric spatial cure rate models for interval-censored time-to-relapse data. *Biometrics*, **60**: 268–275.

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

Bao, Y., Cancho, V., Louzada, F. and Suzuki, A. (2017). Cure rate proportional odds models with spatial frailties for interval-censored data. *Commun. Statistic. Appl. Meth.*, **24**: 605–625.

Berkson, J. and Gage, R. P. (1952). Survival cure for cancer patients following treatment. *J. Am. Statistic. Assoc.*, **47**: 501–515.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Royal Statistic. Soc. Ser. B (Methodol.)*, **36**: 192–236.

Cancho, V. G., Louzada, F. and Barriga, D. C. G. (2011). The poisson-exponential lifetime distribution. *Computat. Statistic Data Anal.*, **55**: 677–686.

Carlin, B. and Banerjee, S. (2003). *Hierarchical Multivariate CAR Models for Spatio-Temporally Correlated Survival Data (with discussion)*, Oxford University Press.

Chen, M. H., Ibrahim, J. G. and Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *J. Am. Statistic. Assoc.*, **94**: 909–919.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman & Hall/CRC, Boca Raton, FL.

Cooner, F., Banerjee, S., Carlin, B. and Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *J. Am. Statistic. Assoc.*, **102**: 560–572.

Dan Li, X. W. and Dey, D. K. (2016). A flexible cure rate model for spatially correlated survival data based on generalized extreme value distribution and gaussian process priors. *Biometric. J.*, **58**: 1178–1197.

Dey, D. and Birmiwal, L. (1994). Robust bayesian analysis using divergence measures. *Statistics Probability Lett.*, **20**: 287–294.

Ewell, M. and Ibrahim, J. G. (1997). The large sample distribution of the weighted log rank statistic under general local alternatives. *Lifetime Data Anal.*, **3**: 5–12.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, **38**: 1041–1046.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, 3rd edn., Wiley, New York.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, **42**: 845–854.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd edn., Chapman & Hall/CRC, Boca Raton.

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

Geweke, J. (1992). *Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments*, Clarendon Press, Oxford, UK.

Gu, Y., Sinha, D. and Banerjee, S. (2011). Analysis of cure rate survival data under proportional odds model. *Lifetime Data Analysis*, **17**: 123–134.

Ibrahim, J., Chen, M. and Sinha, D. (2001). *Bayesian survival analysis*, Springer Verlag.
Lindsey, J. C. and Ryan, L. M. (1998). Methods for interval-censored data. *Statistics in Medicine*, **17**: 219–238.
Maller, R. A. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*, Wiley, New York.
Murray, R. P., Anthonisen, N. R., Connett, J. E., Wise, R. A., Lindgren, P. G., Greene, P. G., Nides, M. A. *et al.* (1998). Effects of multiple attempts to quit smoking and relapses to smoking on pulmonary function. *J. Clinic. Epidemiol.*, **51**: 1317–1326.
Pan, C., Cai, B., Wang, L. and Lin, X. (2014). Bayesian semiparametric model for spatially correlated interval-censored survival data. *Comput. Statistic. Data Anal.*, **74**: 198–208.
Peng, F. and Dey, D. (1995). Bayesian analysis of outlier problems using divergence measures. *Canadian J. Statistic.*, **23**: 199–213.
Peto, R. (1973). Experimental survival curves for interval-censored data. *Appl. Statistic.*, **22**: 86–91.
Plummer, M., Best, N., Cowles, K. and Vines, K. (2005). Output analysis and diagnostics for mcmc. *R package version 0.10-3*, http://cran.rproject.org.
R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
Rodrigues, J., Cancho, V., Castro, M. and Louzada-Neto, F. (2009). On the unification of long-term survival models. *Statistic. Probabil. Lett.*, **79**: 753–759.
Rücker, G. and Messerer, D. (1988). Remission duration: An example of interval-censored observations. *Statistic. Med.*, **7**: 1139–1145.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. Roy. Statistic. Soc. B*, **64**: 583–639.
Stangl, D. K. and Greenhouse, J. B. (1998). Assessing placebo response using bayesian hierarchical survival models. *Lifetime Data Anal.*, **4**: 5–28.
Sun, X. and Chen, C. (2010). Comparison of finkelstein's method with the conventional approach for interval-censored data analysis. *Statistic. Biopharm. Res.*, **2**: 97–108.
Weiss, R. (1996). An approach to bayesian sensitivity analysis. *J. Roy. Statistic. Soc. Ser. B (Methodol.)*, **58**(4): 739–750.
Yakovlev, A. Y. and Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*, World Scientific, New Jersey.
Yau, K. K. and Ng, A. S. (2001). Long-term survivor mixture model with random effects: Application to a multi-centre clinical trial of carcinoma. *Statistic. Med.*, **20**: 1591–1607.