Statistics and Its Interface Volume 12 (2019) 457–465

Bayesian approach for clustered interval-censored data with time-varying covariate effects

Yue Zhang, Xia Wang, and Bin Zhang*

Interval-censored data arise when the failure time cannot be observed exactly but can only be determined to lie within an interval. Interval-censored data are very common in clinical trials and epidemiological studies. In this study, we consider a Bayesian approach for clustered interval-censored data under a dynamic Cox regression model. Some methods that incorporate right censoring have been developed for clustered data with temporal covariate effects. However, interval-censored data analysis under the same circumstance is much less developed. In this paper, we estimate piecewise constant coefficients based on a dynamic Cox regression model under the Bayesian framework. The dimensions of coefficients are automatically determined by the reversible jump Markov chain Monte Carlo algorithm. Meanwhile, we use a shared frailty factor for unobserved heterogeneity or for statistical dependence between observations. Simulation studies are conducted to evaluate the performance of the proposed method. The methodology is exemplified with a pediatric study on children's dental health data.

Keywords and phrases: Cox model, Frailty, Interval censoring, Reversible jump Markov chain Monte Carlo, Time-varying coefficient, Children's dental health data.

1. INTRODUCTION

Interval censoring is commonly referred to as a type of sampling scheme or incomplete data. It often arises in longitudinal studies in which subjects are assessed only periodically at some specific times. By interval-censored data, the failure time is known to lie within certain time intervals, instead of being observed exactly. For example, HIV infection time is only known to fall between the last visit time with a negative result and the first visit time with a positive result. Examples of interval-censored data in AIDS studies can be found in [1, 2, 3, 4, 5]. Besides AIDS, other studies in demographic, epidemiology and medical science also target interval-censored data. See for example, [6, 7, 8, 9, 10, 11].

The most popular semiparametric regression model in survival literature is the proportional hazards model, which is also referred to as the Cox model [12]. It specifies that covariates have a multiplicative effect on the hazard function of the failure time of interest. Many approaches have been developed for interval-censored data under the Cox model. However, there are limitations in existing models. One example is that the relative risk of two subjects may change over time. Hence, it is important to detect the temporal effects on the failure time. The time-varying coefficients in right-censored data can be estimated by several ways, such as the partial likelihood approach [13], histogram sieve procedures [14], and the one-step estimation procedure for the cumulative parameter function [15, 16]. However, Cox models with time-varying coefficients for interval-censored data are much less developed. Kneib [17] developed an extended geoadditive Cox model that estimates the nonlinear effects of covariates based on penalized splines. Sinha *et al.* [18] treated the unobserved exact time as latent variables and sampled from the full conditional posterior distribution via Gibbs sampling. The estimated curves in these approaches depend on a fixed number of knots, and the smoothness of the curves is controlled by the prior distribution or penalizing the difference between adjacent regression coefficients.

In some recent studies, the reversible jump Markov chain Monte Carlo (MCMC) algorithm [19] was used for automatic knot selection. Kim *et al.* [20] used such an algorithm for a dynamic baseline hazard function. Wang *et al.* [21] proposed a Bayesian extension of the Cox model by applying an efficient reversible jump MCMC and putting dynamics on all coefficients as well as the baseline hazard, which were specified as piecewise constants. However, their methods only considered the case in which the subjects are independent, which may not be realistic in some applications. For instance, a lot of clinical trials are multi-center studies, especially for rare diseases. Thus, the correlation of subjects within each cluster becomes crucial and needs to be addressed in the analysis. The motivating example is a longitudinal oral health study conducted in Flanders (Belgium) — the Signal Tandmobiel [22]. The aim of this project was to assess the oral health condition of Flemish school children and to determine the benefit of the intervention. The outcome of interest here is the time to the emergence of permanent tooth 24, which is interval-censored due to the annual examination scheme. Children involved in this study were from five provinces. The correlation of the subjects within the same province is nonignorable.

In this paper, we propose a frailty Cox regression model when the subjects are correlated. The model also allows the

*Corresponding author.

existence of time-varying coefficients. A Bayesian approach is discussed with efficient implementation. The remainder of the paper is organized as follows. Section 2 discusses the general data structure, the model and the associated likelihood function. Section 3 describes prior specifications and posterior computation details. Section 4 shows simulation results of the proposed method and compares with Wang's model [21]. Children's dental health data is analyzed in Section 5. Conclusions and discussions are enclosed in Section 6.

2. MODEL AND THE LIKELIHOOD

Assume that there are n clusters in a study and m_i subjects in each cluster, $i = 1, 2, ..., n$. Hence there are a total of $N = \sum_{i=1}^{n} m_i$ subjects in the study. Let $T_{i,j}$ denote the survival time for the j^{th} subject in the i^{th} cluster, $j = 1, \ldots, m_i$. The p-dimensional vector $\boldsymbol{x}_{i,j}$ represents the p covariates and ω_i denotes the unobserved frailty random variable for the i^{th} cluster. For interval-censored data, the unobserved event time $T_{i,j}$ is located in the censoring interval $(L_{i,j}, R_{i,j}]$. The contribution of the j^{th} subject in the i^{th} cluster to the observed data likelihood is

$$\begin{aligned}&\Pr(T_{i,j} \in (L_{i,j}, R_{i,j}]|\omega_i, \boldsymbol{x}_{i,j})\\ &\quad = \Pr(T_{i,j} > L_{i,j}|\omega_i, \boldsymbol{x}_{i,j}) - \Pr(T_{i,j} > R_{i,j}|\omega_i, \boldsymbol{x}_{i,j}).\end{aligned}$$

Under the Cox model with time-varying regression coefficients, the hazard function for the j^{th} subject in the i^{th} cluster can be written as

$$\lambda(t|\omega_i, \boldsymbol{x}_{i,j}) = \lambda_0(t)\omega_i \exp\left\{\boldsymbol{x}_{i,j}^T \boldsymbol{\beta}(t)\right\},$$

where $\lambda_0(t)$ is an unknown baseline hazard function common to all the subjects and $\boldsymbol{\beta}(t)$ is the p-dimensional regression coefficient function of main interest. This is a shared frailty model, which is a common type of the frailty model used for within-cluster dependence. Note that the "shared frailty" implies that individuals in the same cluster share the common frailty. The frailty ω_i is assumed to follow a parametric distribution, which can either take the form of finite mean frailty distributions including but not limited to the gamma or the log-normal distribution; or take the form of infinite mean distributions such as the positive stable distribution [25].

In the above model, both $\lambda_0(t)$ and $\boldsymbol{\beta}(t)$ are assumed to be left continuous step functions, where both the number of jumps and the locations of the jumps are random and are estimated. A fine time grid is specified as $\mathcal{T} = \{0 = \tau_0 < \tau_1 < \tau_2 < ...\tau_k < \ldots < \tau_K < \infty\}$. It contains all the potential jump points of the functions. The length of each time interval may be taken to be sufficiently small to approximate any hazard and coefficient function. Let $\lambda_k = \lambda_0(\tau_k)$, and $\boldsymbol{\beta}_k = \boldsymbol{\beta}(\tau_k)$ denote the baseline hazard function and the coefficient function evaluated at each grid point k, $k = 1, \ldots, K$; $dN_{i,j,k}$ indicates whether or not the event time $T_{i,j}$ falls within the k^{th} interval of the grid, i.e., $dN_{i,j,k} = I(T_{i,j} \in (\tau_{k-1}, \tau_k])$; $Y_{i,j,k}$ denotes the at-risk variable. If $dN_{i,j,k} = 1$ for some value k, $Y_{i,j,l} = 1$ for $l < k$ and $Y_{i,j,l} = 0$ for $l > k$, while $Y_{i,j,k} = (T_{i,j} - \tau_{k-1})/\Delta_k$ for $l = k$, where $\Delta_k = \tau_k - \tau_{k-1}$ is the width of the k^{th} interval. The augmented likelihood function for the j^{th} subject of the i^{th} cluster is

$$\begin{aligned}&\ell_{i,j}(\Theta|\{dN_{i,j,k}, Y_{i,j,k}\}_{k=1}^{K}, \omega_i, \boldsymbol{x}_{i,j})\\ &\quad = \prod_{k=1}^{K} \{\lambda_k \omega_i \exp(\boldsymbol{x}_{i,j}^T \boldsymbol{\beta}_k)\}^{dN_{i,j,k}}\\ &\qquad \times \exp\{-\Delta_k \lambda_k \omega_i \exp(\boldsymbol{x}_{i,j}^T \boldsymbol{\beta}_k) Y_{i,j,k}\},\end{aligned}$$

where $i = 1, 2, ..., n$, $j = 1, 2, ..., m_i$ and $\Theta = \{\lambda_k, \boldsymbol{\beta}_k; k = 1, 2, ..., K\}$.

3. PRIOR AND POSTERIOR

3.1 Prior

In the following description, we use $\theta(t)$ to denote either $\log \lambda_0(t)$ or one element in the p-dimensional vector $\boldsymbol{\beta}(t)$. Assume that the number of jumps P in $\theta(t)$ follows a discrete uniform distribution ranging from 1 to K. For a fixed P, the jump times $0 < \tau_1 < \tau_2 < ...\tau_p < \ldots < \tau_P = \tau_K$ are randomly selected from all time grids except the last one. Given P and the jump times, a hierarchical Markov type process prior for $\theta(t)$ proposed by Wang *et al.* [21] is specified as follows

$$\begin{aligned}\theta(\tau_1) &\sim \mathcal{N}(0, a_0\nu),\\ \theta(\tau_p)|\theta(\tau_{p-1}) &\sim \mathcal{N}(\theta(\tau_{p-1}), \nu), \quad p = 2, 3, .., P,\\ \nu &\sim \mathcal{IG}(\alpha_0, \eta_0),\end{aligned}$$

where $a_0 > 0$ is a hyper-parameter which can be chosen as a large number to reflect higher uncertainty in the prior input, and $\mathcal{IG}(\alpha_0, \eta_0)$ denotes an inverse-gamma distribution with a shape parameter α_0 and a scale parameter η_0 such that the mean is $\eta_0/(\alpha_0 - 1)$. Similar priors have been used in generalized additive models [24, 23]. In order to compare simulation results, we set $a_0 = 100, \alpha_0 = 2, \eta_0 = 1$, which are the same as in [21]. The gamma distribution, the most commonly used finite mean distribution, is used to model the frailty term $\omega_i, i = 1, \ldots, n$. For finite mean frailty distributions, we need the mean of the frailty distribution to be unity in order for the parameters of the model to be identifiable [25]. Thus, we assume

$$\omega_i \overset{\text{i.i.d.}}{\sim} \mathcal{G}(\kappa^{-1}, \kappa^{-1}), \quad i = 1, 2, .., n,$$

where κ is the variance of the ω_i's and larger values of κ imply greater heterogeneity among clusters. Let $\eta = \kappa^{-1}$ for notational convenience. Vague hyper-priors for η are commonly used, such as the uniform distribution $\mathcal{U}(0, a)$ with a large a or the gamma distribution $\mathcal{G}(b, b)$ with b close to zero. In this paper, a vague gamma prior $\mathcal{G}(0.001, 0.001)$ [25]

is used, which is denoted as $\pi_\eta(.)$. The joint prior density is proportional to

$$\pi_\eta(\eta)\prod_{i=1}^{n}\{\omega_i^{\eta-1}\exp(-\eta\omega_i)\}\frac{\eta_0^{\alpha_0}}{\Gamma(\alpha_0)}\nu^{-\alpha_0-1}\exp\big(-\frac{\eta_0}{\nu}\big)\nu^{-\frac{P}{2}}$$
$$\times\exp\left\{-\frac{\theta(\tau_1)^2}{2a_0\nu}\right\}\prod_{p\geq 2}\exp\left[-\frac{\{\theta(\tau_p)-\theta(\tau_{p-1})\}^2}{2\nu}\right].$$

3.2 Posterior computation

The posterior samples are obtained under a Gibbs sampling framework based on the j^{th} subject of the i^{th} group observed in the k^{th} time interval, where $i=1,2,...,n, j=1,2,...,m_i, k=1,2,...,K$, and K is the total number of time grids. The parameters of interest include $\theta(t)$ and the frailty term ω_i's. The event indicators $dN_{i,j,k}$'s, the event time $T_{i,j}$'s, and the at-risk variables $Y_{i,j,k}$'s also need to be sampled. Let $D=\{dN_{i,j,k},Y_{i,j,k}\}$, $\Theta=\{\theta(t)\}$, $W=\{\omega_i\}$. The Gibbs sampling algorithm draws D, Θ, ν, W and η iteratively, where ν and η are hyper-parameters.

The first step is to sample the event time $T_{i,j}$, event indicators $dN_{i,j,k}$'s and at-risk variables $Y_{i,j,k}$'s for augmented data given Θ and W. This can be decomposed into two steps.

(I) Locate the grid interval for each event time. For a finite interval-censored subject, the event indicators $dN_{i,j,k}$'s follow a multinomial distribution with a size 1 and a probability vector $(e_{i,j,1},e_{i,j,2},...,e_{i,j,k})$, where for $k=1,2,...,K$,

$$e_{i,j,k}=\frac{p_{i,j,k}I(\tau_k\in(L_{i,j},R_{i,j}])}{\sum_{\tau_l\in(L_{i,j},R_{i,j}]}p_{i,j,l}},$$
$$p_{i,j,k}=\begin{cases}\exp\left\{-\sum_{l=1}^{k-1}\Delta_l\lambda_l\omega_i\exp(\boldsymbol{x}_{i,j}^T\boldsymbol{\beta}_k)\right\}\\ \quad-\exp\left\{-\sum_{l=1}^{k}\Delta_l\lambda_l\omega_i\exp(\boldsymbol{x}_{i,j}^T\boldsymbol{\beta}_k)\right\}\\ \quad\text{if } k>1,\\ 1-\exp\left\{-\Delta_1\lambda_1\omega_i\exp(\boldsymbol{x}_{i,j}^T\boldsymbol{\beta}_1)\right\}\quad\text{if } k=1.\end{cases}$$

Thus, if the observed interval $(L_{i,j},R_{i,j}]$ only covers one time grid τ_k, then $dN_{i,j,k}=1$ and all the other event indicators equal 0. Otherwise, if $(L_{i,j},R_{i,j}]$ covers multiple time grids τ_k's, $dN_{i,j,k}$ is sampled from the multinomial distribution with the probability vector calculated based on these covered time grids τ_k's. In other words, the time interval $(\tau_{k-1},\tau_k]$ with $dN_{i,j,k}=1$ is sampled in this step.

(II) Within the selected time grids, the exact time $T_{i,j}$ follows a doubly truncated exponential distribution on $(\tau_{k-1},\tau_k]$ with a distribution function

$$F(u)=\frac{1-\exp\{-\lambda_k(u-\tau_{k-1})\omega_i\exp(\boldsymbol{x}_{i,j}^T\boldsymbol{\beta}_k)\}}{1-\exp\{-\lambda_k\Delta_k\omega_i\exp(\boldsymbol{x}_{i,j}^T\boldsymbol{\beta}_k)\}}.$$

Then $T_{i,j}$ is sampled by the inverse distribution method, and the at-risk variables $Y_{i,j,k}$'s are calculated as defined in Section 2.

The next step is to sample each component of the baseline hazard $\log\lambda_0(t)$ and the regression coefficients $\boldsymbol{\beta}(t)$ given D and W. The reversible jump MCMC algorithm [19] is applied here because the number of jumps P is random and the dimension of the posterior distribution could vary from iteration to iteration. The probabilities of taking a birth, death and update move are set as 0.3, 0.3, and 0.4 [21], respectively.

(I) Update move. In this step, both P and the jump times are fixed. The conditional posterior distribution of $\theta(\tau_p)$ given D, W and all the other components in Θ is

(1)
$$\pi(\theta(\tau_p)|\Theta/\{\theta(\tau_p)\},D,W)\propto\exp\left[-\frac{\{\theta(\tau_p)-\mu_p\}^2}{2\sigma_p^2}\right]$$
$$\times\exp\left\{-\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}I(\tau_k\in(\tau_{p-1},\tau_p])\Delta_k\lambda_k\omega_i\right.$$
$$\left.\times\exp(\boldsymbol{x}_{i,j}^T\boldsymbol{\beta}_k)Y_{i,j,k}\right\},$$

where $\theta(\tau_p)$ is either $\log\lambda(\tau_p)$ or one component in $\boldsymbol{\beta}(\tau_p)$. The steps of computing μ_p and σ_p^2 are listed in Appendix. Since it can be shown that (1) is log-concave, the adaptive rejection algorithm [26] is applied to sample $\theta(\tau_p)$.

(II) Birth move. A new jump time τ' is "born" in this move. This new τ' is randomly selected from the non-jump time grids.

(III) Death move. One of current jump time τ_p is removed, where the index p is uniformly selected from the current jump point set $\{1,2,...,P-1\}$. Details of birth move and death move are listed in Appendix.

The hyper-parameter ν has a conjugate inverse-gamma prior and the posterior distribution is also an inverse-gamma specified as

$$\nu|\Theta,D\sim\mathcal{IG}\left[\alpha_0+\frac{P}{2},\eta_0+\frac{\theta(\tau_1)^2}{2a_0}\right.$$
$$\left.+\sum_{p\geq 2}\frac{\{\theta(\tau_p)-\theta(\tau_{p-1})\}^2}{2}\right].$$

The conditional posterior distribution of η given W is

$$\eta|W\propto\left(\prod_{i=1}^{n}\omega_i\right)^{\eta-1}\eta^{n\eta}\frac{\exp(-\eta\sum_{i=1}^{n}\omega_i)}{\Gamma(\eta)^n}\pi_\eta(\eta),$$

i.e., the conditional posterior distribution of η depends on the data only through W. The Metropolis-Hastings algorithm is implemented to evaluate the posterior distribution of η, where the acceptance rate is tuned to be around 25%.

Table 1. Model specifications in the simulation study

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
β_1	1	1	1	$0.5+\sin(t\pi/9)$	$0.5+\sin(t\pi/9)$	$0.5+\sin(t\pi/9)$
x_1	$\mathcal{B}(0.5)^*$	$\mathcal{N}(0,1)^{**}$	$\mathcal{B}(0.5)$	$\mathcal{B}(0.5)$	$\mathcal{N}(0,1)$	$\mathcal{B}(0.5)$
β_2			1			1
x_2			$\mathcal{N}(0,1)$			$\mathcal{N}(0,1)$

* $\mathcal{B}(0.5)$ denotes a Bernoulli distribution with the success probability as 0.5.
** $\mathcal{N}(0,1)$ represents a standard normal distribution.

As mentioned before, a gamma distribution $\mathcal{G}(0.001, 0.001)$ is used as the prior $\pi_\eta(\cdot)$ for η in the following analysis.

The frailty ω_i is sampled as follows

$$\omega_i|\Theta, D, \eta \sim \mathcal{G}\left(\eta + \sum_{j=1}^{m_i}\sum_{k=1}^{K} dN_{i,j,k}, \eta + \sum_{j=1}^{m_i}\sum_{k=1}^{K} \Delta_k \lambda_k \exp(\boldsymbol{x}_{i,j}^T \boldsymbol{\beta}_k) Y_{i,j,k}\right).$$

4. SIMULATION

In this section, we present simulation study results to assess the properties of the methods introduced in previous sections. Both constant and time-varying coefficients are considered in the simulation study, where the constant coefficients are set to 1 and time-varying coefficient function is $\beta(t) = 0.5 + \sin(t\pi/9)$. The time interval of interest is set to be $(0, 9)$. In addition, we consider two types of covariates (binary or continuous). Then we simulate the survival time t under six models with various combinations of coefficients and covariates. Table 1 lists the six different models, where Models 1, 2, 4, 5 contain one covariate and Models 3, 6 have two covariates.

Assume that the baseline hazard function is $\lambda_0(t) = 0.1\sqrt{t}$, and let $u = S(t)$ follow a uniform distribution $\mathcal{U}(0, 1)$. The shared frailty ω_i is sampled from a gamma distribution $\mathcal{G}(1, 1)$. The survival time is sampled with the following steps. Firstly a random variable u is generated from the uniform distribution $\mathcal{U}(0, 1)$. If the coefficients are constant, the survival time is computed from the inverse survival function $t = S^{-1}(u) = \left[-15\log(u)/\left\{\omega_i \exp(\boldsymbol{x}_{i,j}^T\boldsymbol{\beta})\right\}\right]^{\frac{2}{3}}$. If at least one coefficient is time-varying, we have $u = S(t) = S_0(t)^{\omega_i \exp\left\{\boldsymbol{x}_{i,j}^T \boldsymbol{\beta}(t)\right\}}$, where $S_0(t) = \exp\left\{-\int_0^t \lambda_0(z)dz\right\}$ and t is calculated with the numerical method after taking natural logarithm on both sides of the equation. To generate the censoring intervals, the log-normal density function $\mathcal{LN}(x; 0, 0.4)$ is used to simulate gap times. If the exact event time occurs between two consecutive visits, the time interval of the visits is taken as the censoring interval. If the exact event time does not occur before the maximum follow up time, the subject would be considered as right-censored with the last visit time as the censoring point.

For each model, 300 datasets were simulated. Each simulated dataset contains three clusters and each cluster contains 100 subjects. The computation is implemented via R 3.2.2 and C++. The code is avaialble upon request from the corresponding author. We ran 12,000 MCMC iterations with the first 2, 000 iterations as the burn-in period. We checked the convergence based on trace plots, autocorrelation plots and Geweke statistics. LPML (log pseudo marginal likelihood) [25] is used for model comparison. LPML is the summation over all CPO(i) (conditional predictive ordinate), where CPO(i) is the posterior predictive probability of the ith observation given all the other observed data under the assumption that the current model is true. LPML is preferred in this study because the dimension of the parameters changes from one iteration to another. Models with higher LPML are preferred to models with lower LPML. Note that when all frailties are equal to one, our model reduces to Wang's model [21].

Figure 1 displays estimated coefficients from all six models. The upper two rows of Figure 1 contains the models with one covariate and the lower two rows is for the models with two covariates. Panels in the first row consist of plots of results from the proposed model with frailty and those in the second row present Wang's model without frailty. Each panel includes the median of the posterior means and the median of 95% credible intervals of the regression coefficients from our model or Wang's model. By comparing each pair of the plots, it is obvious that the proposed model outperforms Wang's method by capturing the true values for both constant and time-varying coefficients. Besides, the proposed model produces credible intervals for the regression coefficients that are similar or narrower in width compared to those from Wang's model.

Table 2 shows the estimates of frailties and model comparison results. Each cell contains an entry $m(\ell, u)$ to summarize the posterior estimates by the 100 data replicates, where m is the median of posterior means, ℓ is the median of the posterior 2.5% percentiles and u is the median of the posterior 97.5% percentiles for the frailties $(\omega_1, \omega_2, \omega_3)$ and the inverse of the frailty variance η. Thus the interval (ℓ, u) is the median of the 95% credible intervals. The results show that all these intervals contain the true values of the parameters. In addition, for LPML, m, ℓ, u are simply the posterior mean, the posterior 2.5% and 97.5% percentiles of "LPML Diff", where "LPML Diff" is calculated as LPML

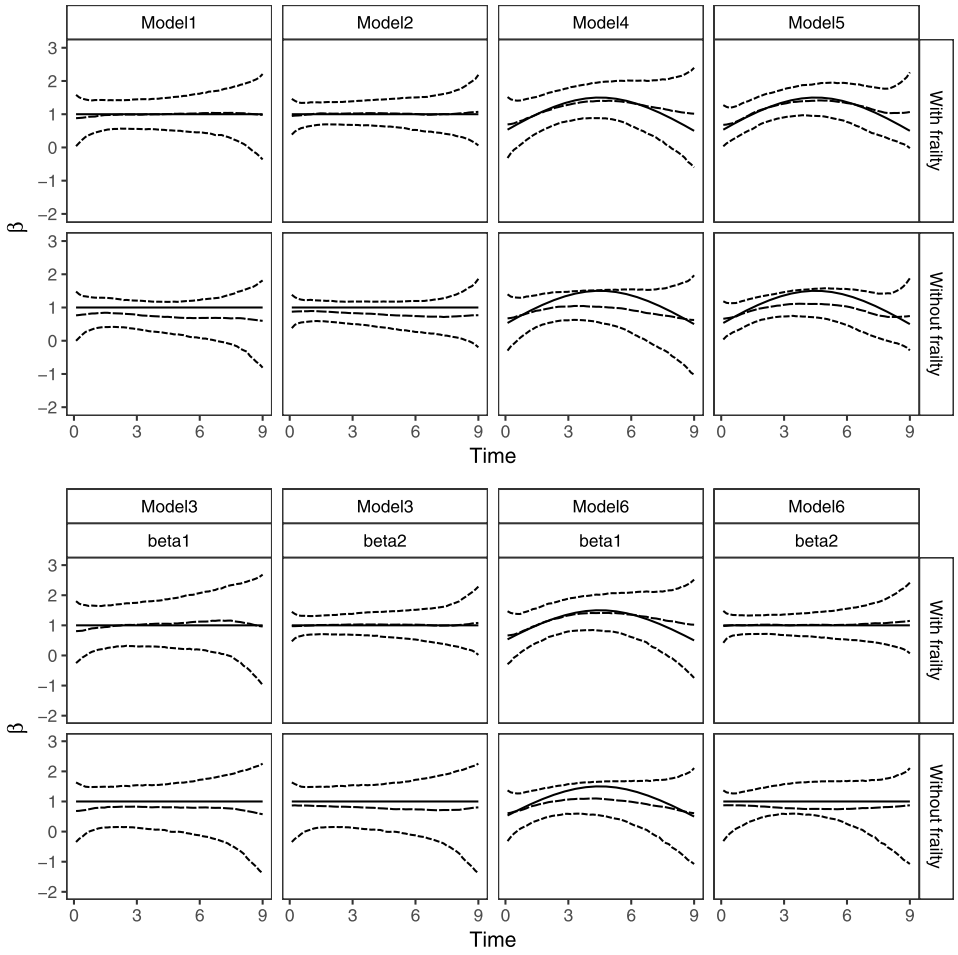

Figure 1. Estimate of coefficients for models with three clusters. The solid lines in the middle are the true coefficients. The long dashed lines are medians of posterior means. The top and bottom short dashed lines are medians of 95% *credible intervals.*

Table 2. Estimates of frailties and model comparison results

	Model 1	Model 2	Model 3
$\omega_1(0.419^*)$	0.281 (0.019, 0.623)	0.177 (0.001, 0.506)	0.266 (0.012, 0.605)
$\omega_2(0.936^*)$	0.648 (0.047, 1.363)	0.387 (0.002, 1.099)	0.588 (0.029, 1.322)
$\omega_3(2.171^*)$	1.436 (0.106, 3.046)	0.886 (0.005, 2.532)	1.345 (0.070, 2.897)
$\eta(1^*)$	1.729 (0.031, 4.951)	1.217 (0.011, 3.905)	1.574 (0.026, 4.694)
LPML Diff	39.651 (-13.787, 82.223)	30.810 (-31.919, 154.180)	36.743 (-14.974, 97.956)
% Diff > 0	95.7%	87.0%	93.3%
	Model 4	Model 5	Model 6
$\omega_1(0.419^*)$	0.303 (0.028, 0.644)	0.185 (0.001, 0.507)	0.272 (0.016, 0.591)
$\omega_2(0.936^*)$	0.685 (0.066, 1.431)	0.414 (0.003, 0.414)	0.595(0.036, 1.308)
$\omega_3(2.171^*)$	1.519 (0.147, 3.200)	0.915 (0.007, 2.392)	1.399(0.087, 2.984)
$\eta(1^*)$	1.834 (0.035, 5.098)	1.253 (0.012, 4.163)	1.622(0.028, 4.735)
LPML Diff	36.743 (-14.974, 97.956)	32.784 (-27.650, 141.947)	35.798 (-46.414, 116.700)
% Diff > 0	96.3%	87.0%	91.3%

* True value.

The entry $m(\ell, u)$ in each cell summarizes the posterior estimates by the 100 data replicates, where m is the median of posterior means, ℓ is the median of the posterior 2.5% percentiles and u is the median of the posterior 97.5% percentiles for the frailties $(\omega_1, \omega_2, \omega_3)$ and the inverse of the frailty variance η. Thus the interval (ℓ, u) is the median of the 95% credible interval (CI). "LPML Diff" is calculated as LPML of the proposed model minus that of Wang's model [21]. For "LPML Diff", m, ℓ, u are simply the posterior mean, the 2.5% and 97.5% percentiles of "LMPL Diff".

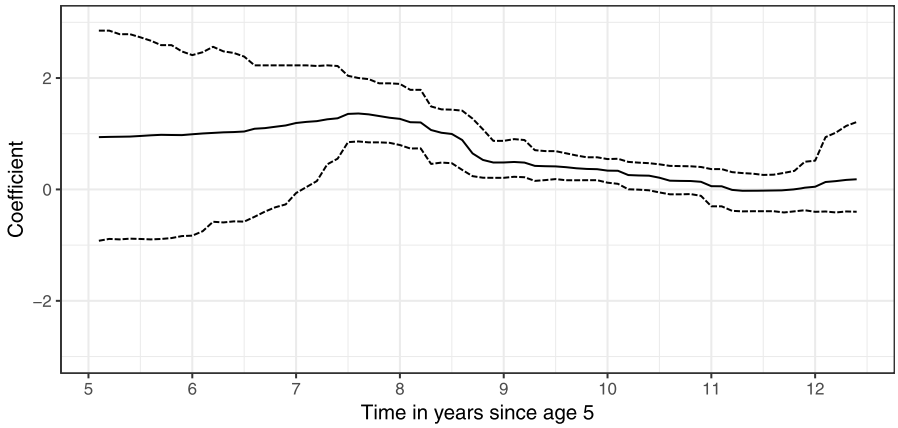

Figure 2. Estimate of coefficients in children's dental health data. The black solid line is the posterior mean. The black dashed lines are 95% credible intervals.

Table 3. Estimate of frailties with 95% credible intervals in the children's dental health data

	Antwerpen	Limburg	Oost Vlaanderen	Vlaams Brabant	West Vlaanderen
Girls	1.18 (1.02, 1.35)	1.08 (0.92, 1.25)	1.19 (1.02, 1.36)	1.16 (0.98, 1.35)	1.06 (0.89, 1.23)
Boys	0.84 (0.72, 0.96)	0.81 (0.68, 0.93)	0.91 (0.78, 1.05)	0.93 (0.79, 1.08)	0.87 (0.73, 1.00)

of the proposed model minus the LPML of Wang's model. It also contains the percentages of "LPML Diff" greater than zero. Table 2 shows that the posterior mean of LPML difference is at least 30.810 and at least 87% of "LPML Diff" is greater than zero, which further indicates that the proposed model performs consistently better and consideration of within-cluster correlation is necessary in some cases.

In order to see the performance of the proposed method under various scenarios, we also tried the frailty model with 10 clusters, equal or unequal sample size assignment on different clusters as well as different magnitudes of frailties. Although not shown here, the results all indicate that the proposed model performs consistently better than Wang's model [21].

5. REAL DATA ANALYSIS

In this section, we apply the proposed model to the dental health data from the Signal Tandmobiel project that was conducted in Flanders (Belgium) to examine the oral health condition of Flemish primary schoolchildren. The children were divided into 15 strata, a combination of 3 educational systems (public, municipal or private) and 5 provinces. Over 6,000 children were recruited, which represented approximately 7% of the total target population in Flanders [22]. A total of 4,468 of them were randomized and examined annually by 16 trained dentists using the standardized and widely accepted criteria recommended by the WHO. We focus on the time to the emergence of permanent tooth 24 (central incisor) here.

For the analysis, we considered the covariate *dmf* as a dichotomized variable which denotes the status of the primary predecessor of this tooth (0=sound, 1=decayed, missing or filled). A random effect of province-by-gender was considered. Frailty is assumed to follow a gamma distribution with an equal shape and scale parameter η, which has a gamma hyper-prior $\mathcal{G}(0.001, 0.001)$. A total of 12,000 Gibbs samples were generated with a burn-in period of 2,000 samples. The convergence of MCMC chains was checked by trace plots, autocorrelation plots and Geweke's statistics.

Figure 2 presents the analysis results by applying the proposed method. As tooth 24 does not emerge before age 5, the time scale in Figure 2 is the time in years since age 5. The results include the point estimates and the corresponding 95% credible intervals. The positive estimate indicates that the children with a decayed primary predecessor have higher risks than those without, which is consistent with the results based on the iterative convex minorant algorithm [27]. However, there is an obvious trend of the coefficient estimate, and the effect becomes weaker over time. In our study, if the 95% credible interval includes 0, then one may conclude that the parameter is not significantly different from 0, and thus the effect is said to be statistically insignificant [28, 29]. As shown in Figure 2, the credible interval of the *dmf* effect after "Year 11" contains 0, which indicates that the *dmf* effect becomes insignificant. All these findings were not captured in [27]. Table 3 shows the frailty estimates. In general, girls have higher risk than boys. Specifically, in Antwerpen and Limburg, the two provinces in the north and adjacent to the Netherlands, the difference between girls and boys is larger than that in Vlaams Brabant and West Vlaanderen, the two provinces in the middle of Belgium. Moreover, boys in Vlaams Brabant, where the capital city, Brussels, is located,

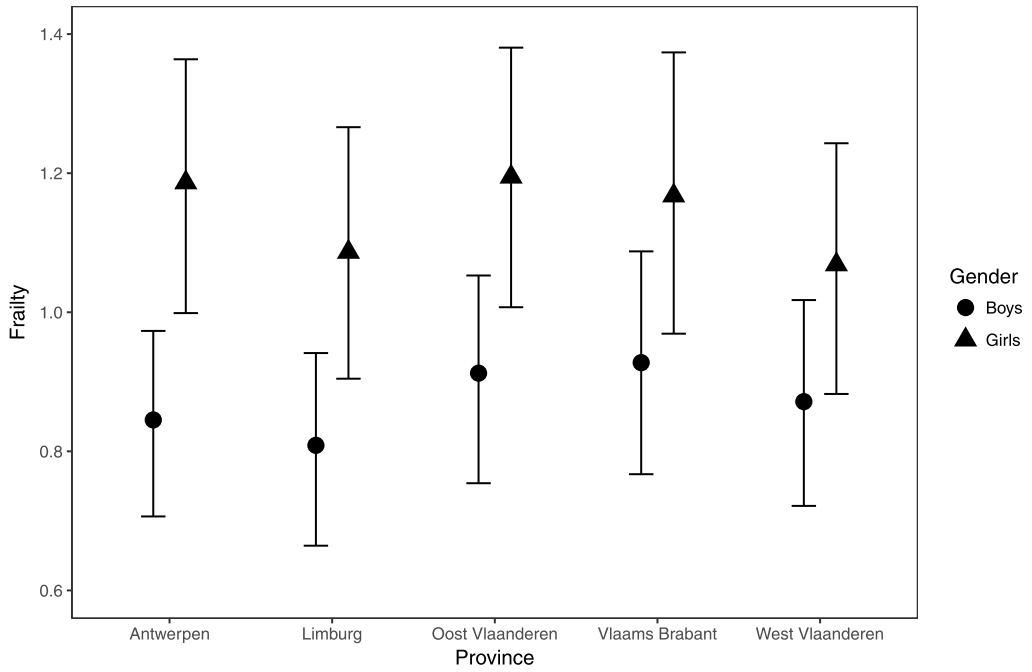

Figure 3. Estimate of frailties with 95% *credible intervals in the children's dental health data.*

have the highest risk compared to boys in other provinces (Figure 3). We also fitted Wang's model [21] to the data with the covariate *dmf*. The LPML value for the proposed model and Wang's model are -5480.8 and -5511.0, respectively. In summary, the proposed model is preferred to the models in [27] and [21].

6. DISCUSSION

In the present study, we proposed a Bayesian approach for the clustered interval-censored Cox regression model with time-varying covariate effects. In order to capture the temporal nature of covariate effects more precisely, we have shown that it is important to consider the dependence structure for clustered outcomes. In our study, we used a gamma frailty model for unobserved heterogeneity. A non-informative hyper-prior of gamma parameter is used for heterogeneity among clusters. Other finite mean frailty models, such as the log-normal frailty model, can also be applied in the proposed model.

A key step in this procedure is to determine the number of jumps, which will affect the estimated smoothness of the estimated curves as well as effectiveness assessment. Previous studies usually specify a large number of jumps and then select jumps after model fitting. This may not be appropriate when prior information of jumps is not available. In our approach, we used the reversible jump MCMC to automatically select jumps during model fitting. The regression coefficients and the baseline hazard are piecewise constant and can be estimated given the number of pieces and jump locations. A dynamic prior was specified in this study to capture the time-varying coefficients. A recent study by Rue *et al.* [30] has shown that the prior should probably be adapted to the interval length of the time grids and the number of the time points. It can be of our major interest to improve the dynamic prior in future studies.

Although our discussion focuses on the Cox model with shared gamma frailty, much wider extensions and applications are feasible with the proposed Bayesian approach. In some cases, independent frailty assumption may not hold and the method can be extended to more general frailty distributions, such as "correlated frailty models" [31], where the within cluster frailty factors are correlated and the restriction of unobserved factors acting similarly within clusters is elevated. For example, our proposed procedure can be extended to incorporate spatial correlation in the data.

APPENDIX: DETAILS OF BIRTH MOVE AND DEATH MOVE

Birth move

A new jump time τ_p' is "born" in this move. This new τ_p' is randomly selected from non-jump time grids. Let $\{\tau_p', p = 1, 2, ..., P+1\}$ and $\{\tau_p, p = 1, 2, ..., P\}$ denote new and current jump times, respectively. Assume $\tau_p' \in (\tau_{p-1}, \tau_p)$, then $\theta(\tau_p')$ and $\theta(\tau_{p+1}')$ need to be sampled.

$$\theta(\tau_p') = \pi_1\theta(\tau_{p-1}) + \pi_2\{\theta(\tau_p) + u\},$$
$$\theta(\tau_{p+1}') = \pi_1\{\theta(\tau_p) - u\} + \pi_2\theta(\tau_{p+1}),$$

where weights are defined as

$$\pi_1 = (\tau'_p - \tau'_{p-1})/(\tau'_{p+1} - \tau'_{p-1}),$$
$$\pi_2 = (\tau'_{p+1} - \tau'_p)/(\tau'_{p+1} - \tau'_{p-1}),$$

where u is generated from a uniform distribution $\mathcal{U}(-\epsilon_0, \epsilon_0)$ and ϵ_0 is set to 1 in this study. Variable u here is an auxiliary variable to the old model, which helps balance out the one dimension increase of the proposed new model. When τ' is near the boundaries, set $\theta(\tau_0) = \theta(\tau_1)$ and $\theta(\tau_{P+1}) = \theta(\tau_P)$. Let $\theta = \{\theta(\tau_1), \theta(\tau_2), ..., \theta(\tau_P)\}$ and $\theta' = \{\theta(\tau'_1), \theta(\tau'_2), ..., \theta(\tau'_{P+1})\}$. The acceptance ratio can be computed with the posterior distribution $\pi(\theta'|\Theta/\{\theta'\}, \omega, \nu, D)$, the uniform density function π_u and the Jacobian of the transformation,

$$R_{\text{birth}} = \frac{\pi(\theta'|\Theta/\{\theta'\}, \omega, \nu, D)}{\pi(\theta|\Theta/\{\theta\}, \omega, \nu, D)\pi(u)} \left| \frac{\partial \theta'}{\partial(\theta, u)} \right|.$$

The acceptance probability is defined as $\min\{1, R_{\text{birth}}\}$.

Death move

One of the current jump times τ_p is removed, where the index p is uniformly selected from the current jump point set $\{1, 2, ..., P-1\}$. Then this can be treated as an inverse step of birth move. By using the same transformation function of the birth move, the expression of $\theta(\tau'_p)$ can be computed as follows

$$\theta(\tau'_p) = \frac{1}{2}\Big\{ -\frac{\pi_1}{\pi_2}\theta(\tau_{p-1}) + \frac{1}{\pi_2}\theta(\tau_p) + \frac{1}{\pi_1}\theta(\tau_{p+1}) - \frac{\pi_2}{\pi_1}\theta(\tau_{p+2})\Big\},$$

where the weights are defined as

$$\pi_1 = (\tau_p - \tau_{p-1})/(\tau_{p+1} - \tau_{p-1}),$$
$$\pi_2 = (\tau_{p+1} - \tau_p)/(\tau_{p+1} - \tau_{p-1}),$$

and the acceptance probability is $\min\{1, R^{-1}_{\text{birth}}\}$.

Calculation of μ_p and σ_p^2

For $p = 1$,

$$\mu_1 = \sigma_1^2 \left[\sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{K} I\{s_k \in (0, \tau_1]\} \boldsymbol{x}_{i,j} dN_{i,j,k} \right] + \frac{a_0\theta(\tau_2)}{1 + a_0},$$
$$\sigma_1^2 = a_0\omega/(1 + a_0);$$

for $p = 2, ..., P-1$,

$$\mu_p = \sigma_p^2 \left[\sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{K} I\{s_k \in (\tau_{p-1}, \tau_p)\} \boldsymbol{x}_{i,j} dN_{i,j,k} \right] + \frac{\theta(\tau_{p-1})}{2} + \frac{\theta(\tau_{p+1})}{2},$$
$$\sigma_p^2 = \omega/2;$$

for $p = P$,

$$\mu_p = \sigma_P^2 \left[\sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{K} I\{s_k \in (\tau_{P-1}, \tau_P]\} \boldsymbol{x}_{i,j} dN_{i,j,k} \right] + \theta(\tau_{P-1}),$$
$$\sigma_P^2 = \omega.$$

ACKNOWLEDGMENTS

We thank the Associate Editor and the three referees for their thoughtful and constructive comments which greatly improved the paper. Dr. Yue Zhang was supported by the Shanghai Philosophy and Social Sciences Planning Project (2018EJB006) and the Fundamental Research Funds for the Central Universities (17X100040066).

Received 22 June 2017

REFERENCES

[1] De Gruttola, V. G. and Lagakos, S. W. (1989). Analysis of doubly-censored survival data with application to AIDS. *Biometrics* **45**, 1–11. MR0999438

[2] Jewell, N. P., Malani, H. M. and Vittingho, E. (1994). Nonparametric estimation for a form of doubly censored data, with application to two problems in AIDS. *Journal of the American Statistical Association* **89**, 7–18.

[3] Kim, M. Y., De Gruttola, V. G. and Lagakos, S. W. (1993). Analyzing doubly censored data with covariates, with application to AIDS. *Biometrics* **49**, 13–22. MR0999438

[4] Sun, J. (1996). A non-parametric test for interval-censored failure time data with application to AIDS studies. *Statistics in medicine* **15**, 1387–1395.

[5] Shiboski, S. and Jewell, N. P. (1992). Statistical analysis of the time dependence of HIV infectivity based on partner study data. *Journal of the American Statistical Association* **87**, 360–372.

[6] Diamond, I., Mcdonald, J. W. and Shah, I. H. (1986). Proportional hazards models for current status data: application to the study of differentials in age at weaning in Pakistan. *Demography* **23**, 607–620.

[7] Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42**, 845–854. MR0872963

[8] Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41**, 933–945. MR0833140

[9] Hoel, D. G. and Walburg, H. E. (1972). Statistical analysis of survival experiments. *Journal of the National Cancer Institute* **49**, 361–372.

[10] Self, S. G. and Grossman, E. (1986). Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics* **42**, 521–530.

[11] Sun, J. and Kalbfleisch, J. D. (1996). Nonparametric tests of tumor prevalence data. *Biometrics* **52**, 726–731.

[12] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)* **34**, 187–220. MR0341758

[13] Zucker, D. M. and Karr, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *The Annals of Statistics* **18**, 329–353. MR1041396
[14] Higle, J. L. and Sen, S. (1991). Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research* **16**, 650–669. MR1120475
[15] Martinussen, T. and Scheike, T. H. (2002). A flexible additive multiplicative hazard model. *Biometrika* **89**, 28–298. MR1913959
[16] Martinussen, T., Scheike, T. H. and Skovgaard, I. M. (2002). Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. *Scandinavian Journal of Statistics* **29**, 57–74. MR1894381
[17] Kneib, T. (2006). Mixed model-based inference in geoadditive hazard regression for interval-censored survival times. *Computational Statistics & Data Analysis* **51**, 777–792. MR2297486
[18] Sinha, D., Chen, M. and Ghosh, S. K. (1999). Bayesian analysis and model selection for interval-censored survival data. *Biometrics* **55**, 585–590. MR1705161
[19] Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732. MR1380810
[20] Kim, S., Chen, M., Dey, D. K. and Gamerman, D. (2007). Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Analysis* **13**, 17–35. MR2355294
[21] Wang, X., Chen, M. and Yan, J. (2013). Bayesian dynamic regression models for interval censored survival data with application to children dental health. *Lifetime Data Analysis* **19**, 297–316. MR3084025
[22] Vanobbergen, J., Martens, L., Lesaffre, E. and Declerck, D. (2000). The Signal-Tandmobiel project a longitudinal intervention health promotion study in Flanders (Belgium): baseline and first year results. *European Journal of Paediatric Dentistry* **2**, 87–96.
[23] Brezger, A. and Lang, S. (2006). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics & Data Analysis* **50**, 967–991. MR2210741
[24] Fahrmeir, L. and Lang, S. (2001). Bayesian semiparametric regression analysis of multicategorical time-space data. *Annals of the Institute of Statistical Mathematics* **53**, 11–30. MR1820949
[25] Ibrahim, J. G., Chen, M. and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer, New York. MR1876598
[26] Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **37**, 337–348.
[27] Gómez, G., Calle, M. L., Oller, R. and Langohr, K. (2009). Tutorial on methods for interval-censored data and their implementation in R. *Statistical Modelling* **9**, 259–297. MR2751307
[28] Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Part 2, Inference, Oxford University Press. MR0168084
[29] Lee, P. M. (2012). *Bayesian Statistics: An Introduction*, 4th Edition, Wiley. MR3237439
[30] Simpson, D., Rue, H., Riebler, A., Martins, T. G. and Sorbye, S. H. (2017). Penalising model component complexity: a principled, practical approach to constructing priors. *Statistical Science* **32**, 1–28. MR3634300
[31] Hens, N., Wienke, A., Aerts, M. and Molenberghs, G. (2009). The correlated and shared gamma frailty model for bivariate current status data: An illustration for cross-sectional serological data. *Statistics in Medicine* **28**, 2785–2800. MR2750165

Yue Zhang
Department of Bioinformatics and Biostatistics
School of Life Sciences and Biotechnology
Shanghai Jiao Tong University
Shanghai 200240
PR China
E-mail address: yue.zhang@sjtu.edu.cn

Xia Wang
Department of Mathematical Sciences
University of Cincinnati
Cincinnati, OH 45221
USA
E-mail address: xia.wang@uc.edu

Bin Zhang
Division of Biostatistics and Epidemiology
Cincinnati Children's Hospital Medical Center
Department of Pediatrics
University of Cincinnati College of Medicine
Cincinnati, OH 45229
USA
E-mail address: Bin.Zhang@cchmc.org