JIRSS (2013)
Vol. 12, No. 1, pp 127-151

Dynamic Frailty and Change Point Models for Recurrent Events Data

Changhong Song[1], Lynn Kuo[2]

[1]Division of Biostatistics, Center for Devices and Radiological Health, Food and Drug Administration, USA.
[2]Department of Statistics, University of Connecticut, USA.

Abstract. We present a Bayesian analysis for recurrent events data using a nonhomogeneous mixed Poisson point process with a dynamic subject-specific frailty function and a dynamic baseline intensity function. The dynamic subject-specific frailty employs a dynamic piecewise constant function with a known pre-specified grid and the baseline intensity uses an unknown grid for the piecewise constant function. Implementation of Bayesian inference using a reversible jump Markov Chain Monte Carlo (RJMCMC) algorithm is developed to handle the change of the dimension in the parameter space for models with a random number of change points. A data set provided by Grubbs et al. (1991) with recurrent times to mammary tumors for 59 rats is used to illustrate the application of the new models. We compare several models including constant or piecewise constant subject-specific frailty and a fixed number or a random number for the change points in the baseline using the pseudo-marginal likelihood criterion. We show that models with a random number of change points in the baseline improve upon that of a fixed number.

Keywords. Dynamic frailty models, intensity, nonhomogeneous Pois-

Changhong Song(✉)(Changhong.Song@fda.hhs.gov),Lynn Kuo(lynn.kuo@uconn.edu)
Received: June 18, 2011; Accepted: October 3, 2012

son point process, reversible jump Markov Chain Monte Carlo.

MSC: 62M05.

1 Introduction

Data sets for times to occurrences of a specific event, such as recurring tumors, infections, and hospitalizations have been collected in many medical studies. For most of these studies, the interest is on assessing whether the treatment is effective in reducing the overall occurrence of events.

For analysis of this kind of data, it is important to account for within subject correlations, because data collected for the same subject are likely to be more related. Failure to do this may cause severe underestimation for the variance of the treatment effect. Commonly used approaches for modeling within subject correlations include marginal, conditional, and frailty models (Vaupel, Manton, and Stallard, 1979). For marginal models, both working independent and working correlation models are considered, and the robust 'sandwich' method (Lin and Wei, 1989, Wei, Lin, and Weissfeld 1989) is applied for the variance estimate adjustment. Conditional models for recurrent events are based on event ranks. The baseline risk is the same for events with the same number of previous events, and different for events with different previous event numbers. The Prentice, Williams, and Peterson (1981) model is an example of a conditional model. Frailty models assume an unobservable random effect shared by events from the same subject. Within-subject correlation is accounted for using the random effect. If a subject has a large value of frailty, then this subject is at a higher risk of a new event compared to a subject with a small value of frailty.

We consider frailty models in this paper. One of the models is called the shared frailty model which assumes the frailty term is common to recurrent events of the same subject and is constant over time. For examples of Bayesian analyses involving shared frailty models, see Sinha (1993), Sahu *et al.* (1997), Aslanidou, Dey, and Sinha (1998), Sinha (1998), Dunson and Dinse (2000), Chen, Ibrahim, and Sinha (2002), and Dunson and Chen (2004). To relax the constant frailty assumption, dynamic models where random effects can vary stochastically over time

or event sequence have been considered. They include Paik, Tsai, and Ottman (1994), Yue and Chan (1997), Yau and MacGilchrist (1998), Fong, Lam, Lawless, and Lee (2001), Lam, Lee, and Leung (2002), Henderson and Shimakura (2003) and Pennell and Dunson (2006). Dynamic frailty models will allow us to investigate the possible changes in the subject's failure risk and serial dependence of the frailty according to time since the beginning of the experiment. We will focus on a time-varying dynamic frailty model in this paper because it is more appropriate to assume the subject-specific risk changes according to time as opposed to number of events in the tumor carcinogenesis studies.

It is common to use a nonhomogeneous Poisson counting process to model the number of recurring events (Lawless and Nadeau, 1995). The intensity function of the Poisson process for each subject can be modeled as a multiplicative function of the common baseline intensity, the treatment effect and the subject-specific frailty. Both the baseline intensity and the subject-specific frailty functions can be modeled by piecewise constant functions with a known number of change points. The piecewise functions are more flexible than parametric models to accommodate the evolution changes of the intensity function. In many situations, determining a desirable number of change points is a challenge. If we set the number too large, this will make the model complex. The extra parameters may increase estimation uncertainty. If we set the number too small, it may not fit the data well. Therefore it is desirable to make this number random. So we extend shared and dynamic frailty models with a fixed number of change points in the baseline to that with a random number. The new models have the advantage of making the locations and the number of change points to be data-dependent, and therefore they are more flexible and adaptive for model fitting. We develop a reversible jump Markov Chain Monte Carlo method (RJMCMC; Green, 1995) to handle the change of the dimension of the parameter space. The algorithm can be extended to a random number of change points for each subject-specific frailty function. However, our analysis in this paper did not incorporate this extension due to the data set we have selected does not have enough data points for reliable inference for the dimension change of each subject's specific frailty.

To illustrate the application of our models, we use the data on times to mammary tumors of 59 rats from Grubbs *et al.* (1991). This data

set has been analyzed before. For example, Kokoska *et al.* (1993) use an expanded data set and compare four treatment groups based on a parametric analysis for the number of tumors per animal and the time to each tumor detection. Dunson and Dinse (2000) model the mean and hazard function using Poisson and logistic regression; and Pennell and Dunson (2006) model the mean and hazard function using a dynamic frailty model with a nonparametric Dirichlet process.

We evaluate four types of models including constant shared or dynamic subject-specific frailty models and dynamic baseline models with a fixed or random number of change points using the logarithm of the pseudo-marginal likelihood (LPML). This is the logarithm of the product of the conditional predictive ordinates (CPO) (Gelfand, Dey, and Chang, 1992) over all subjects. For the data set by Grubbs *et al.*, we show that most of the dynamic frailty models fit the data better than the constant shared frailty models, and models with a random number of change points in the baseline are superior to that with a fixed number.

We would like to point out that our model is related to Pennell and Dunson (2006) except we model the subject-specific frailty and the baseline intensity directly by two separate dynamic piecewise constant functions without the unknown Dirichlet distribution as hyper-parameter. Furthermore, they assume fixed knots in the piecewise dynamic gamma models, we added unknown variable knots in them to be learned from the data. Different from the approach used in Pennell and Dunson, we need to employ RJMCMC to handle the unknown variable knots problem. We also address model determination issues in our paper for the various models we explore.

2 Model Specification

The typical data set contains time epochs of recurrent events for each subject and a right-censored time for each subject. Some of these times epochs can have ties. We will first transform the data set into a counting process data set. The underlying process for the number of recurrent events could be modeled as a nonhomogeneous Poisson process. We use $N_i(t)$ to denote the number of events of the i^{th} subject for the small time interval around time t, and $\lambda_i(t)$ to denote the intensity function for the i^{th} subject at time t. The intensity function could be modeled

as a multiplicative function of the subject-specific frailty, an exponential function of the covariates and the baseline intensity. The intensity function $\lambda_i(t)$ for subject i at time t could be expressed as

$$\lambda_i(t) = I_i(t) w_i(t) \exp\{\boldsymbol{\beta}^T z_i(t)\} \mathcal{B}(t), \tag{1}$$

where $I_i(t)$ is an indicator function with value 1 when the subject i is at risk at time t and 0 otherwise; $w_i(t)$ denotes the frailty of subject i at time t; $z_i(t)$ is a p dimensional time-dependent vector of covariates evaluated at time t for subject i and $\boldsymbol{\beta}$ is the p-dimensional regression coefficient; and $\mathcal{B}(t)$ denotes the baseline intensity at time t common to all subjects.

Therefore, three components, $\boldsymbol{\beta}$, $\mathcal{B}(t)$, and $w_i(t)$, for $i = 1, \ldots, I$ need to be estimated. Nonparametric estimation of $\mathcal{B}(t)$ and $w_i(t)$ may be undesirable due to limited number of data points. To make the estimation of $\mathcal{B}(t)$ and $w_i(t)$ more computationally tractable and without sacrificing too much generality, we assume piecewise constant models as an intermediate between parametric and nonparametric models. The details are given next.

2.1 Subject-Specific Frailty

We consider both time-independent shared frailty and time-dependent dynamic frailty models.

In the time independent shared frailty models, $w_i(t)$ is assumed to be a constant w_i, which does not evolve with time.

In dynamic frailty models, $w_i(t)$ is modeled by a time-varying piecewise constant function. In particular, we divide the study period $(0, T]$ into $m+1$ segments with $t_0 = 0 < t_1 < t_2 < \cdots < t_{j-1} < t_j < \cdots < t_{m+1} = T$, where the j^{th} segment is defined by $(t_{j-1}, t_j]$. The frailty for each subject can be modeled by a step function

$$w_i(t) = \sum_{j=1}^{m+1} w_{ij} I(t_{j-1} < t \leq t_j), \tag{2}$$

where w_{ij} denotes the frailty of the i^{th} subject at the j^{th} segment. Note that we assume for simplicity that all subjects, $i = 1, \ldots, I$, share the same pre-specified change points at t_j, $j = 1, \ldots, m$, but with different unknown frailty magnitudes.

The evolution of frailties over segment j could be defined by $w_{ij} = w_{i(j-1)}\phi_{ij}$, where ϕ_{ij} is the multiplicative frailty innovation for the i^{th} subject over segment j. Note $w_{ij} = \prod_{g=1}^{j}\phi_{ig}$. This multiplicative dynamic structure has a convenient way of modeling between subject and within subject variations. In a model with $\phi_{ij} \overset{i.i.d}{\sim} G(\psi,\psi)$, a gamma distribution with mean 1 and variance $\frac{1}{\psi}$, we assume that heterogeneities between and within subjects have the same magnitude $1/\psi$. Then the correlation between frailties of the segments j and $j+d$ (for a positive integer d) for each subject could be calculated similarly as in Pennell and Dunson (2006):

$$\begin{aligned}
\mathrm{Corr}(w_{ij}, w_{i(j+d)}) &= \frac{E(\Pi_{g=1}^{j}\phi_{ig}^2\Pi_{g=j+1}^{j+d}\phi_{ig}) - E(\Pi_{g=1}^{j}\phi_{ig})E(\Pi_{g=1}^{j+d}\phi_{ig})}{\sqrt{V(\Pi_{g=1}^{j}\phi_{ig})V(\Pi_{g=1}^{j+d}\phi_{ig})}} \\
&= \frac{E(\Pi_{g=1}^{j}\phi_{ig}^2)E(\Pi_{g=j+1}^{j+d}\phi_{ig}) - E(\Pi_{g=1}^{j}\phi_{ig})E(\Pi_{g=1}^{j+d}\phi_{ig})}{\sqrt{V(\Pi_{g=1}^{j}\phi_{ig})V(\Pi_{g=1}^{j+d}\phi_{ig})}} \\
&= \frac{(1+1/\psi)^j - 1}{\sqrt{((1+1/\psi)^j - 1)((1+1/\psi)^{j+d} - 1)}} \\
&= \sqrt{\frac{(1+1/\psi)^j - 1}{(1+1/\psi)^{j+d} - 1}}. \qquad (3)
\end{aligned}$$

So we can verify the autocorrelation is decreasing in d and increasing in j, both are expected. Moreover, it is a rational function of $1+1/\psi$, where $1/\psi$ measures the temporal heterogeneity within each subject. When $\psi \to 0$, then the autocorrelation approaches 0; when $\psi \to \infty$, then the autocorrelation approaches $\sqrt{j/(j+d)}$. We can also assume a slightly more general version with $\phi_{i1} \overset{i.i.d.}{\sim} G(\psi_1,\psi_1)$ for all i for the between heterogeneity as in Pennell and Dunson, which has a different parameter than the within heterogeneity ψ. For our data, we found the first version is sufficient.

2.2 Baseline Intensity

The baseline intensity $\mathcal{B}(t)$ is modeled by a piecewise constant function with

$$\mathcal{B}(t) = \sum_{h=1}^{k+1}\mathcal{B}_h I(s_{h-1} < t \le s_h), \qquad (4)$$

with change points at $s_h, h = 1, \ldots, k$, where $0 < s_1 < s_2 < \cdots < s_h < \cdots < s_{k+1} = T$. We set the number of change points k in the baseline intensity separately from that in dynamic frailty models denoted by m, because we consider also models where the number k is unknown while m is always pre-specified.

We define the evolution of the baseline intensity over segment h as $\mathcal{B}_h = \mathcal{B}_{h-1}\delta_h$, where $\mathcal{B}_h$ denotes the baseline intensity at segment h, and δ_h denotes the evolution innovation of the baseline intensity over the h^{th} segment. Equivalently we have for each h, $\mathcal{B}_h = \prod_{g=1}^{h} \delta_g$, as in Pennell and Dunson. The auto-correlation between $\mathcal{B}_h$ and $\mathcal{B}_{h+d}$ can be built similarly as that of the dynamic frailties in (3).

2.3 Different Models

We consider four types of models. Model I includes a constant shared subject-specific frailty model ($m = 0$ in eq.(2)) and a fixed number of change points k in the baseline. Model II includes dynamic models for both subject-specific frailty model and the baseline with a fixed number of change points. Models III and IV are similar to models I and II respectively, except the number of change points in the baseline is assumed to be random.

2.4 Priors

For Bayesian inference, priors for parameters need to be specified. We assume the baseline innovation $\delta_h \overset{i.i.d}{\sim} G(\nu_1, \nu_1)$ for $h = 1, \ldots, k+1$ and the multiplicative frailty innovation $\phi_{ij} \overset{i.i.d}{\sim} G(\nu_2, \nu_2)$ for both i and j indices, where ν_1 and ν_2 are given. We assume the change points $s_1, s_2, \ldots, s_k$ have an ordered discrete uniform distribution on (0,T]. In models III and IV, the number of change points k has a Poisson distribution prior with mean ξ. In models I and II, k is pre-specified. The β parameter in our example has only one dimension, so we have chosen $e^{\beta} \sim G(\nu_3,\ \nu_3)$ with known ν_3 to facilitate the posterior sampling of β. For the general problem of p dimensional $\boldsymbol{\beta}$ parameter, a multivariate normal prior has been chosen in the literature.

3 Inference

To obtain parameter estimates, we develop Markov Chain Monte Carlo (MCMC) (Gelfand and Smith, 1990) algorithms for models I and II with a fixed number of change points k. Parameter updating for models III and IV with a random k is implemented by the reversible jump Markov Chain Monte Carlo algorithm (RJMCMC) (Green, 1995).

3.1 Likelihood Function

3.1.1 Likelihood Function for Model II

We describe the likelihood function for model II first. We divide the total study period $(0, T]$ into sufficiently narrow intervals and make the assumptions that $\lambda_i(t) = \lambda_{il}$ a constant for all time t in the interval l. We use $N_i(l)$ denotes the number of tumors observed for the i^{th} rat in the l^{th} interval. We denote the count information from the data by $\boldsymbol{N} = (\boldsymbol{N}_1, \boldsymbol{N}_2, ..., \boldsymbol{N}_I)$, where $\boldsymbol{N}_i = \{N_i(l)\}_{l=1,...,L}$ denotes the vector of the counts of the tumor incidence for subject i collected over the study period.

Let Φ denote the vector of ϕ_{ij}, Δ denote the vector of δ_h, and $\boldsymbol{\tau}$ denote the vector of s_h. When the censoring is non-informative, the likelihood function is given by

$$\begin{aligned} & L(\boldsymbol{\beta}, \Phi, \Delta, \boldsymbol{\tau}; \boldsymbol{N}) \\ &= \prod_{i=1}^{I} \prod_{l} f(N_i(l) | \boldsymbol{\beta}, \Phi, \Delta, \boldsymbol{\tau}) \\ &= \prod_{i=1}^{I} \prod_{l} \exp(-\lambda_{il}) \lambda_{il}^{N_i(l)} / N_i(l)! \qquad (5) \\ &= \prod_{i=1}^{I} \prod_{l} \exp\{-I_i(l) w_i(l) \theta_i(l) \mathcal{B}(l)\} [I_i(l) w_i(l) \theta_i(l) \mathcal{B}(l)]^{N_i(l)} / N_i(l)!. \end{aligned}$$

In the formula,

$$\begin{aligned} w_i(l) &= \sum_{j=1}^{m+1} w_{ij} I(t_{j-1} < l \leq t_j) = \sum_{j=1}^{m+1} (\Pi_{g=1}^{j} \phi_{ig}) I(t_{j-1} < l \leq t_j), \\ \theta_i(l) &= \exp\{\boldsymbol{\beta}^T z_i(l)\}, \end{aligned}$$

$$\text{and } \mathcal{B}(l) = \sum_{h=1}^{k+1} \mathcal{B}_h I(s_{h-1} < l \leq s_h) = \sum_{h=1}^{k+1} (\Pi_{g=1}^{h} \delta_g) I(s_{h-1} < l \leq s_h).$$

For the data set considered in this study, l denotes the l^{th} week, and $N_i(l)$ denotes the number of tumors observed for the i^{th} rat in the l^{th} week. So we have evaluated the likelihood at a finer time grid than both that of the baseline and of the frailty functions.

3.1.2 Likelihood Functions for Other Models

The likelihood function for model I is a special case of the model II with $m = 0$. The partial likelihood of model III (IV) conditioning on the given number of change points in the baseline to be k is that of model I (II).

3.2 Updating Algorithm

We illustrate the updating algorithm for model IV first. It is the dynamic frailty and dynamic baseline model with random k in the baseline. Updating for models I-III will be discussed later.

3.2.1 RJMCMC for Model IV

Suppose we are at the iteration step with k change points. There are three possible moves S, D, and B for the next iteration step. The move S denotes stay with no changes of k in the dimension for baseline. The move D denotes death, reducing k by 1. The move B denotes birth, increasing k by 1. Parameter updating is done by randomly selecting one of the three moves (S, D, B) with probabilities c_k, d_k and b_k respectively. So we first calculate the probabilities of each move. Consider $b_k = \gamma \min\{1, P_\xi(k+1)/P_\xi(k)\}$ and $d_k = \gamma \min\{1, P_\xi(k-1)/P_\xi(k)\}$. Here $P_\xi(k)$ denotes the probability of an outcome k (change points) from a Poisson distribution with mean ξ. The constant γ is chosen as large as possible subject to $b_k + d_k \leq 0.9$ for each k. The probability of stay c_k is simply $1 - b_k - d_k$.

We will describe the possible three moves in more details:

(*S Steps*) If the move type is S, there are no changes of parameter dimension. We update all parameters using the MCMC algorithm from the following posterior distributions.

(S.1) Generate the regression coefficient $\boldsymbol{\beta}$ from the density

$$[\boldsymbol{\beta}|\Phi,\Delta,\boldsymbol{\tau},N] \propto \left[\prod_{i=1}^{I}\prod_{l}\exp\{-I_i(l)w_i(l)\theta_i(l)\mathcal{B}(l)\}[I_i(l)\theta_i(l)]^{N_i(l)}\right]\pi(\boldsymbol{\beta}),$$

where $[I_i(l)\theta_i(l)]^{N_i(l)}=1$ if $I_i(l)=0$. For the general multivariate $\boldsymbol{\beta}$ formulation, $\boldsymbol{\beta}$ is typically sampled using the Metropolis-Hastings (Hastings 1970) algorithm with a normal density proposal. For the special case given in our example with the control group being 0, treatment group being 1 and a gamma prior for $\theta=e^{\beta}$, then we sample e^{β} directly from an updated gamma distribution.

(S.2) Dynamic updating of the frailty evolution parameter $\{\phi_{ij}\}_{j=1,\ldots,m+1}$ is done sequentially from $j=1$ to $j=m+1$ for each subject i, where ϕ_{ij} is generated from $G(\nu_2+N^*_{ij},\nu_2+M^*_{ij})$ with $N^*_{ij}=\sum_{l>t_{j-1}}N_i(l)$ and $M^*_{ij}=\sum_{l>t_{j-1}}I_i(l)w^*_i(l)\theta_i(l)\mathcal{B}(l)$ with $w^*_i(l)$ calculated by $\sum_{g=1}^{m+1}(\Pi^{g}_{a=1,a\neq j}\phi_{ia})I(t_{g-1}<l\leq t_g)$. Derivation of this updating is based on the density

$$[\phi_{ij}|\phi_{ig,g\neq j}{}'s,\Delta,\boldsymbol{\beta},\boldsymbol{\tau},N] \propto \left[\prod_{l>t_{j-1}}\exp\{-I_i(l)w_i(l)\theta_i(l)\mathcal{B}(l)\}[I_i(l)\phi_{ij}]^{N_i(l)}\right]\pi(\phi_{ij}).$$

(S.3) For a random chosen index h from 1 to k, generate δ_h for the baseline intensity from $G(\nu_1+N^*_h,\nu_1+M^*_h)$, where $N^*_h=\sum_{i=1}^{I}\sum_{l>s_{h-1}}N_i(l)$ and $M^*_h=\sum_{i=1}^{I}\sum_{l>s_{h-1}}I_i(l)w_i(l)\theta_i(l)\mathcal{B}^*(l)$ with $\mathcal{B}^*(l)$ calculated by $\sum_{g=1}^{k+1}(\Pi^{g}_{a=1,a\neq h}\delta_a)I(s_{g-1}<l\leq s_g)$. Derivation of the updating is based on the function

$$[\delta_h|\delta_{g,g\neq h}{}'s,\Phi,\boldsymbol{\beta},\boldsymbol{\tau},N] \propto \left[\prod_{i=1}^{I}\prod_{l>s_{h-1}}\exp\{-I_i(l)w_i(l)\theta_i(l)\mathcal{B}(l)\}[I_i(l)\delta_h]^{N_i(l)}\right]\pi(\delta_h).$$

(S.4) To update the time s_h, propose another s'_h by drawing an integer randomly between (s_{h-1},s_{h+1}). Accept s'_h with probability

$$\min\{1,(\text{likelihood ratio})\times(\text{prior ratio})\times(\text{proposal ratio})\}.$$

The likelihood ratio is derived by replacing s_h with s'_h in the likelihood function defined in (5) and then divided by the original likelihood. The prior ratio is

$$\text{prior ratio} = \frac{(s_{h+1} - s'_h)(s'_h - s_{h-1})}{(s_{h+1} - s_h)(s_h - s_{h-1})},$$

and the proposal ratio is 1 because the new point is generated from a discrete uniform distribution.

If the move is a type B, then we implement the following steps to update the parameters in the baseline. The other parameters are updated with the same algorithm as before.

(*B Steps*) We first draw a new jump time s^* from $(0, T)$. Suppose $s^* \in (s_{h-1}, s_h)$. The new time introduces two new levels of the evolution parameters δ_h^* and δ_{h+1}^* for the baseline intensity by using transformations from δ_h and ϵ, where ϵ is drawn randomly from the proposal density $G(\alpha, \alpha)$ with the density function denoted by $g_\alpha(\epsilon)$. We use a transformation of $\delta_h^* = \delta_h/\epsilon$ and $\delta_{h+1}^* = \epsilon$, and accept δ_h^* and δ_{h+1}^* with probability

$$\min\{1, (\text{likelihood ratio}) \times (\text{prior ratio}) \times (\text{proposal ratio}) \times J\},$$

where J is the Jacobian of the transformation. Simple derivation will show

$$\text{likelihood ratio} = \frac{\prod_{i=1}^{I} \prod_{l=s_{h-1}+1}^{s^*} \exp\{-\lambda_i^*(l)\} \lambda_i^*(l)^{N_i(l)}}{\prod_{i=1}^{I} \prod_{l=s_{h-1}+1}^{s^*} \exp\{-\lambda_i(l)\} \lambda_i(l)^{N_i(l)}},$$

where $\lambda_i^*(l) = I_i(l) w_i(l) \theta_i(l) (\Pi_{g=1}^{h-1} \delta_g) \delta_h^*,\ l \in (s_{h-1}, s^*]$;

$$\begin{aligned} \text{prior ratio} \quad = \quad & \frac{P_\xi(k+1)}{P_\xi(k)} \frac{(2k+2)(2k+3)}{T^2} \\ & \frac{(s_{h+1} - s^*)(s^* - s_{h-1})}{(s_{h+1} - s_{h-1})} \frac{\pi(\delta_h^*)\pi(\delta_{h+1}^*)}{\pi(\delta_h)}, \end{aligned}$$

where $\pi(\delta_h)$ denotes the gamma density with $\delta_h \sim G(\nu_1, \nu_1)$; and

$$\text{proposal ratio} = \frac{d_{k+1} T}{b_k (k+1) g_\alpha(\epsilon)} = \frac{d_{k+1} T}{b_k (k+1)} \frac{\Gamma(\alpha) \exp\{\alpha\epsilon\}}{\alpha^\alpha \epsilon^{\alpha-1}};$$

and the Jacobian of the transformation can be calculated to be $J = |\delta_{h+1}^*|$. Details for deriving prior ratio and proposal ratio were given by Green (1995).

(*D Steps*) For a move of type D, we choose an h randomly from $\{1, 2, 3, \ldots, k\}$ for the step involving the removal of the change point h in the baseline. In particular, we propose a new parameter δ_h^* in the segment (s_{h-1}, s_{h+1}) to replace existing evolution parameters δ_h and δ_{h+1} at segment (s_{h-1}, s_h) and (s_h, s_{h+1}). The parameter δ_h^* is defined to be $\delta_h^* = \delta_h * \delta_{h+1}$, which is the reverse of the B move. The likelihood ratio, prior ratio, proposal ratio and Jacobian of the transformation are simply the reverse of the ratios defined in the B move.

3.2.2 Updating For Models I-III

Model III is a special case of model IV where the number of change points m in frailties is 0, and the algorithm is the same as that of the model IV with $m = 0$. Therefore, in the frailties we have $w_i(l) = w_i = \phi_{i1}$ for all l.

Model II is a special case of model IV where the probability of move S is 1. So the algorithm for model II is the same as the algorithm of the move S in model IV.

Model I is a special case of the dynamic frailty model with the number of change points in the subject-specific frailty $m = 0$. The algorithm is exactly the same as the algorithm of model II with $m = 0$.

4 Model Evaluation Criterion

We use a summary statistics which is the logarithm of the pseudo-marginal likelihood (LPML) based on the conditional predictive ordinates (CPO; Gelfand, Dey and Chang, 1992) to compare the proposed models. We prefer this criterion because of the cross-validated idea for the predictive density that provides a meaningful measure of the entertained models with the current data. Pseudo-marginal likelihood is less sensitive to prior choices than the Bayes factor, so it is more useful in practice. Moreover, it is easy to estimate CPOs from the posterior samples using the harmonic mean estimates formula (Gelfand and Dey, 1994), whereas the harmonic mean method is known to have problems in estimating the marginal likelihood used in the Bayes factor (Xie et al. 2011).

The CPO_i denotes the predictive probability of $\boldsymbol{N}_i$ based on $\boldsymbol{N}_1, ...,$

$\boldsymbol{N}_{i-1}, \boldsymbol{N}_{i+1}, ..., \boldsymbol{N}_I$, and

$$CPO_i = \Pr(\boldsymbol{N}_i | z_i, \boldsymbol{N}^{(-i)}),\ i = 1, \ldots, I,$$

where $\boldsymbol{N}^{(-i)}$ denotes the data set with the i^{th} subject data $\boldsymbol{N}_i$ deleted. So CPO_i is the predictive probability of $\boldsymbol{N}_i$ based on the rest of the data. In Monte Carlo simulation, CPO_i can be approximated by the following harmonic mean estimator of the likelihood function based on the full data:

$$\widehat{CPO_i} = \left\{ \frac{1}{R} \sum_{r=1}^{R} \frac{1}{f(\boldsymbol{N}_i | \boldsymbol{\vartheta}_r, z_i)} \right\}^{-1},$$

where R denotes the number of Monte Carlo replications, and $\boldsymbol{\vartheta}_r$ denotes the value of $\boldsymbol{\vartheta}$ for the rth Monte Carlo replicate. Note $\boldsymbol{\vartheta} = (\boldsymbol{\beta}, \boldsymbol{\tau}, \Delta)$ denotes the vector of all parameters except Φ. In our model, f is the Poisson probability function defined in (5). The frailties are treated as latent variables. So we calculate CPO_i by integrating out the random effects. For example, for the constant shared frailty models, we can show

$$\begin{aligned} f(\boldsymbol{N}_i | \boldsymbol{\vartheta}_r, z_i) &= \frac{\nu_2^{\nu_2}}{\Gamma(\nu_2)} \frac{\Gamma(\nu_2 + \sum_l N_i(l))}{\{\nu_2 + \sum_l I_i(l)\theta_i(l)\mathcal{B}(l)\}^{\nu_2 + \sum_l N_i(l)}} \\ &\quad \times \left[\prod_l [\theta_i(l) I_i(l) \mathcal{B}(l)]^{N_i(l)} / N_i(l)! \right]. \end{aligned}$$

For dynamic frailty models, there are no closed forms for the integration. We then use Monte Carlo simulations to integrate out the random effects for each individual.

While CPO_i explains the goodness of model fit for the i^{th} subject based on the remaining data, an overall measure for prediction can be summarized by the product of CPOs, or equivalently its logarithm, called the logarithm of the pseudo-marginal likelihood (LPML):

$$LPML = \sum_{i=1}^{I} \log(CPO_i).$$

5 Analysis of Chemoprevention Data

5.1 Model Comparisons

We first describe the data set given by Grubbs *et al.* (1991). In the experiment, 119 rats were randomized into four groups with 30 rats in each

of the three treatment groups with diet supplements and 29 rats in the control group to examine the effect of canthaxanthin (a carotenoid found in many vegetables and fruits) on chemically induced mammary carcinogenesis. We use part of the data (Dunson and Dinse, 2000, p.1072) consisting of a treatment group and a control group. Each rat in the treatment group received daily diet supplement with 3390 mg/kg canthaxanthin from 34 days old until 54 days old. All rats were given 15 mg of the carcinogen dimethylbenzanthracene by gavage at 55 days old. Then each rat was followed for 180 days, and the times of occurrence of mammary tumors (detected by palpations) were recorded. The rats were not considered to be at risk of developing tumors for the first three weeks, so regular palpations started when a rat was 75 days old. A graphical representation of the data that plots the times and numbers of tumors for each rat is shown in Figure 1. For example, we see Rat #1 has no tumors and is censored at week 26, and Rat #19 has six tumors occurring at weeks 7, 13, 14, 14, 17, 19 and dies at week 19th.

To exemplify our method, we applied it to the data set in Figure 1. As a first step we compared the fit of Models (I)-(IV) to determine which one was most appropriate for the data. The four types of models are (I) constant shared subject-specific frailty models with dynamic baseline models with a fixed number of change points k, (II) dynamic frailty models with fixed change points with size m and dynamic baseline models with the same fixed points, (III) constant shared subject-specific frailty and dynamic baseline with a random k and (IV) dynamic subject-specific frailty with a fixed m and dynamic baseline with random k. Models of type I and II depend on a pre-specified parameter k for the baseline. We evaluate four different models with k to be 2, 4, 6, and 8 respectively. Models of type II and IV depend on the pre-specified parameter m for subject-specific frailty; we also set m to be 2, 4, 6, and 8 for the four different models under each type. When $m = 2$, t_1 and t_2 are chosen to be 5 and 20. When $m = 4$, t_1 to t_4 are chosen to be 5, 10, 15 and 20. When $m = 6$, t_1 to t_6 are chosen to be 5, 8, 12, 16, 20 and 24. When $m = 8$, t_1 to t_8 are chosen to be 5, 7, 9, 12, 15, 18, 21 and 24. The locations of the change points are chosen so that each interval between change points contains a meaningful number of events. We set ξ, the prior mean on the number of change points, to 5.0 for models III and IV, the variance of the baseline evolution parameter $1/\nu_1$ to 1, and

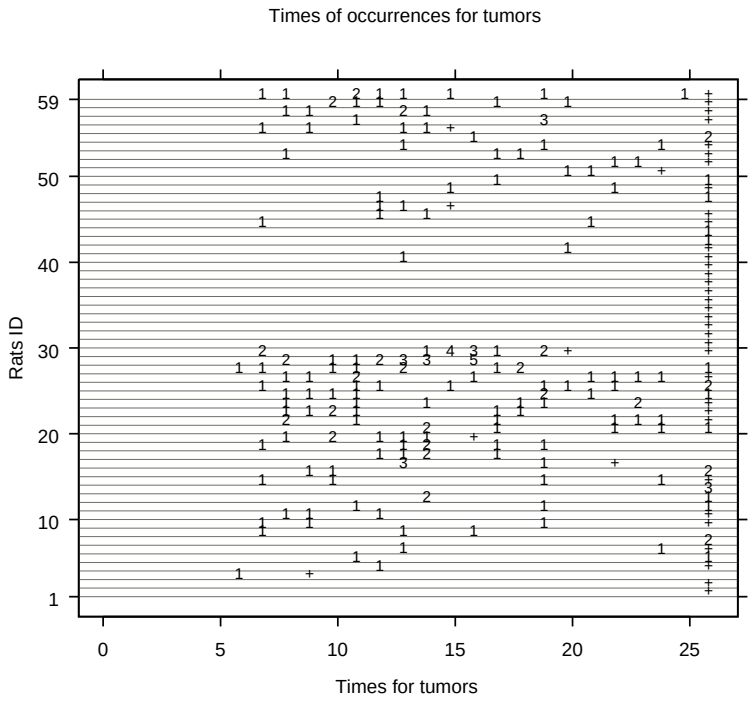

Figure 1: The plot of times to tumors for each rat.
The x-axis denotes the study week for the study, and the y-axis denotes the rats ID. Times and numbers of tumors are plotted for each rat. The number of tumor marked by 1 indicates that one tumor is found at the location marked by 1; 2 indicates that two tumors found at the location marked by 2, etc. The censoring times are denoted by +. Rats #1 to #29 belong to the vehicle control group, and rats #30 to #59 belong to the treatment group.

variance of the subject-specific frailty evolution $1/\nu_2$ to 0.5 respectively for all four models. We have also chosen $e^{\beta} \sim G(0.01, 0.01)$ (with mean of 1 and variance of 100 as the prior for e^{β}.) To implement each of the MCMC methodology for models I and II and the RJMCMC methodology for models III and IV, we ran a single chain of $40,000$ iterations for each type with the first $10,000$ iterations as burn-ins and the remaining $30,000$ iterations used to summarize the posterior characteristics. For each type in Model IV, it takes about 10 minutes for each run on a Dell Optiplex 990 Minitower with an Intel Core i7 2600 Processor (3.4GHz, 8M), 32 GB RAM and a Linux operating system.

Table 1: Results for Model Fitting.

	PM(β)	PSD(β)	HPD(β)	LPML
Type I: Constant shared subject-specific frailty and dynamic baseline with a fixed change points				
k=2	-1.19	0.24	(-1.67,-0.73)	-577
k=4	-1.19	0.24	(-1.66,-0.71)	-579
k=6	-1.18	0.24	(-1.64,-0.70)	-579
k=8	-1.19	0.24	(-1.66,-0.70)	-581
Type II: Dynamic subject-specific frailty and dynamic baseline, both with fixed change points				
k=m=2	-1.09	0.30	(-1.69,-0.51)	-574
k=m=4	-1.11	0.32	(-1.75,-0.49)	-577
k=m=6	-1.12	0.34	(-1.79,-0.45)	-579
k=m=8	-1.12	0.37	(-1.87,-0.43)	-586
Type III: Constant shared subject-specific frailty and dynamic baseline with unknown change points				
Random k	-1.10	0.24	(-1.59,-0.62)	-572
Type IV: Dynamic subject-specific frailty model with fixed change points and dynamic baseline with unknown change points				
Random k, m=2	-1.06	0.32	(-1.69,-0.44)	-571
Random k, m=4	-1.07	0.35	(-1.79,-0.40)	-569
Random k, m=6	-1.04	0.38	(-1.79,-0.33)	-574
Random k, m=8	-1.02	0.40	(-1.83,-0.24)	-580

PM(β) denotes the posterior mean of β, PSD(β) denotes the posterior standard deviation of β, HPD(β) denotes the 95% highest posterior density interval of β, and LPML denotes the logarithm of the pseudo-marginal likelihood.

Table 1 shows the results for the four types of models. The estimates for the posterior mean, the posterior standard deviation, the 95% highest posterior density intervals of β and model fitting score LPML are displayed for each model. Note we do not use the boldface β here because we only have one covariate (treatment) in this data set.

As we can see from the table, the point estimates of β are not sensitive to the choice of the number of change points in both the frailty function and the baseline. The posterior standard deviation for treatment effect becomes larger as the model becomes more complex. For example, the constant shared frailty models have smaller posterior standard deviation of β than that of the dynamic frailty models. The values for model fitting criterion LPML are similar for constant shared frailty models with different choices of k. But this is not the case for the dynamic frailty models. LPML varies greatly with different choices of k and m. With good choices of k and m, dynamic frailty models can fit the data better than the constant shared frailty models, which indicates that the dynamic frailty models could explain the variability of the data better than shared frailty models. Both the constant shared frailty model and the dynamic frailty model with random k show an improvement over models with a fixed k. This supports the advantage of model fitting using a random k in terms of the LPML criterion.

5.2 Results of Best Fit Model

We perform an expanded analysis based on the best model given in Table 1, that is the dynamic frailty model with random k in the baseline, with $m = 4$ for each subject-specific frailty with change points located at 5, 10, 15, and 20, and $\nu_2 = 2$ as the prior mean for the subject-specific frailty multiplicative innovation; and with $\xi = 5$ as the prior mean number of change points, and $\nu_1 = 1$ for the prior mean of the baseline multiplicative innovation.

The convergence of several parameters in the RJMCMC algorithm has been checked by trace plots and autocorrelation plots. To illustrate the performance of the number of change points k in the model, we plotted the posterior histogram of the number of change points k in Figure 2. As we can see from the graph, six or seven change points are most frequently supported by the data.

Table 2: Sensitivity Analysis for Model Type 4 with $m = 4$.

	PM(β)	SD(β)	HPD(β)	LPML
Models with varying ν_1				
$\nu_1 = 0.5, \nu_2 = 2, \xi = 5$	-1.04	0.36	(-1.74,-0.34)	-569
$\nu_1 = 1, \nu_2 = 2, \xi = 5$	-1.07	0.35	(-1.79,-0.40)	-569
$\nu_1 = 2, \nu_2 = 2, \xi = 5$	-1.12	0.35	(-1.79,-0.40)	-570
Models with varying ν_2				
$\nu_1 = 1, \nu_2 = 1, \xi = 5$	-1.03	0.46	(-1.92,-0.12)	-587
$\nu_1 = 1, \nu_2 = 2, \xi = 5$	-1.07	0.35	(-1.79,-0.40)	-569
$\nu_1 = 1, \nu_2 = 5, \xi = 5$	-1.10	0.26	(-1.62,-0.59)	-569
Models with varying ξ				
$\nu_1 = 1, \nu_2 = 2, \xi = 5$	-1.07	0.35	(-1.79,-0.40)	-569
$\nu_1 = 1, \nu_2 = 2, \xi = 9$	-1.08	0.35	(-1.74,-0.38)	-569
$\nu_1 = 1, \nu_2 = 2, \xi = 12$	-1.11	0.35	(-1.79,-0.42)	-569

PM(β) denotes the posterior mean of β, PSD(β) denotes the posterior standard deviation of β, HPD(β) denotes the 95% highest posterior density interval of β, and LPML denotes the logarithm of the pseudo-marginal likelihood.

Figure 3 shows the empirical and the estimated baseline intensity functions for the control (on the left panel) and for the treatment group (on the right panel). The empirical baseline intensity function for the control (treatment) group is calculated by taking the ratio of the observed number of tumors over the number of rats at risk for each week for the control (treatment) group. As we can see, the estimated baseline function is a lot smoother than the empirical one. Smoothing is induced by the model which describes evolution of the baseline over time. The baseline intensity is quite low at the beginning of the study. After the fifth week, it increases dramatically. It becomes more stable in the middle of the study with a few change points. At the end of the study, there is an increasing trend for the baseline. This is caused by the large

number of tumors found in the last week.

Figure 4 displays the estimated cumulative intensity function (solid line) and observed cumulative number of tumors for each rat as a function of time (dotted line). The cumulative intensity function was estimated with the random effects integrated out. We observe that the model estimates the intensity function well for most of the rats with only a few exceptions, where the intensity increases a lot due to many more tumors were found in the last week for a few rats.

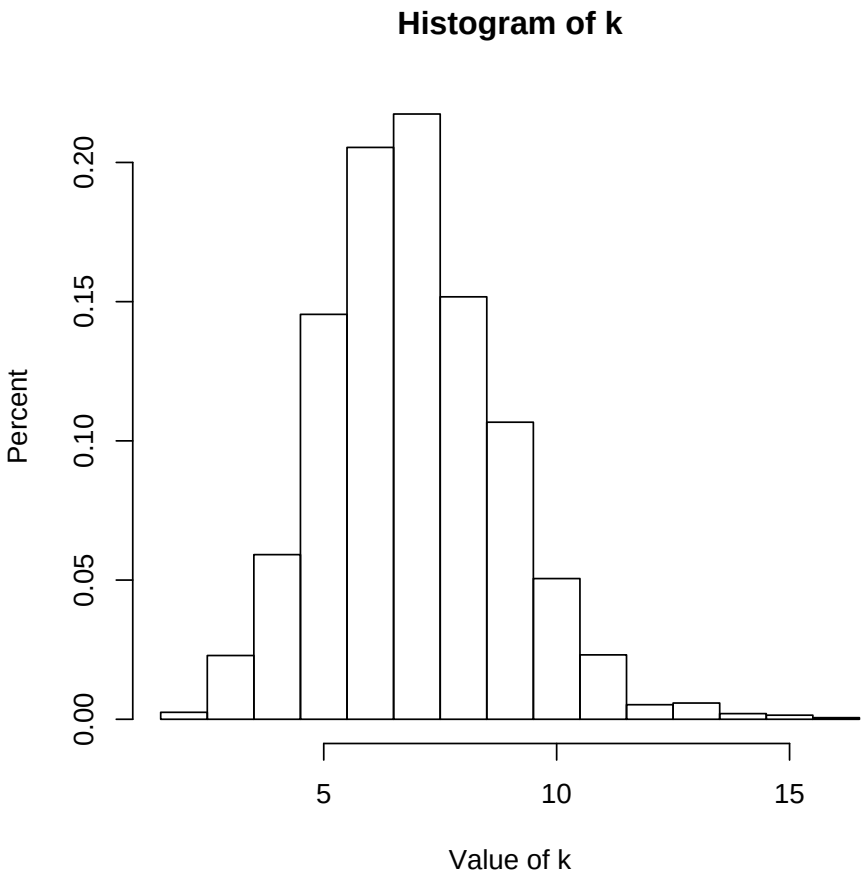

Figure 2: The posterior distribution for the number of change points k.

The treatment is clearly significant in reducing the number of tumors for the rats. In particular, the point estimate for the treatment effect

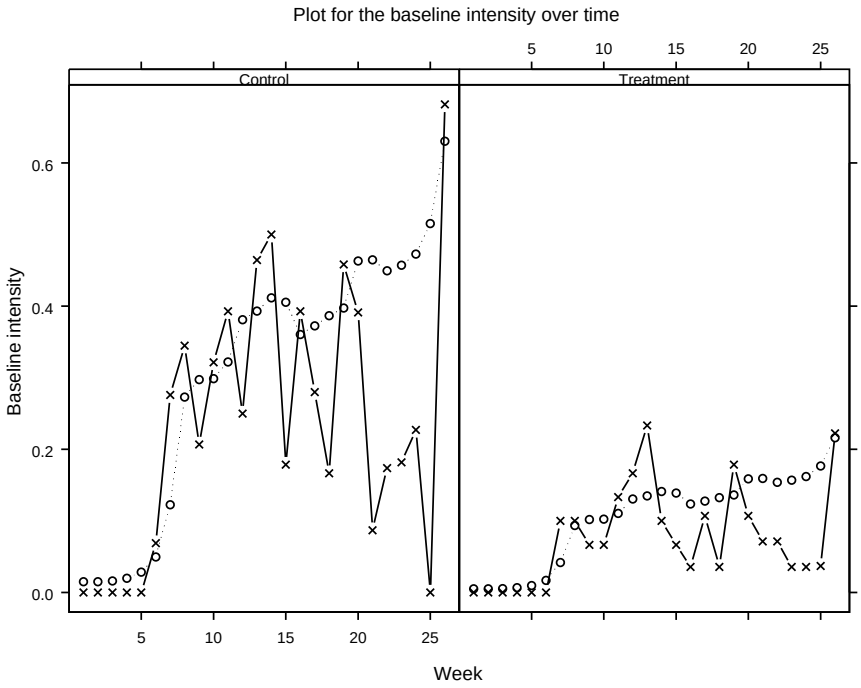

Figure 3: The baseline intensity function over time.
The left panel displays the baseline intensity functions. The right panel displays them with the treatment effect. The symbol × indicates the empirical intensity functions, and o indicates the estimated ones.

β is -1.07, which indicates that the expected number of tumors for the rats in the treatment group is 66% less than those in the control group. The 95% highest posterior density interval is (-1.79, -0.40), which is completely to the left of 0. The result is comparable to the result -1.13 given by Dunson and Dinse obtained (2000) from a shared frailty model. Compared to their model, our model (in particular model IV) with more complex frailty and baseline structure leads to a deeper understanding of the risk evolution and also a more appropriate interpretation due to its better fit. We also consider our model less complex than that of Pennell and Dunson (2006), because we don't have the Dirichlet processes as the hyperparameters for the frailty distribution and the baseline distribution. We model the frailty distribution and baseline distribution

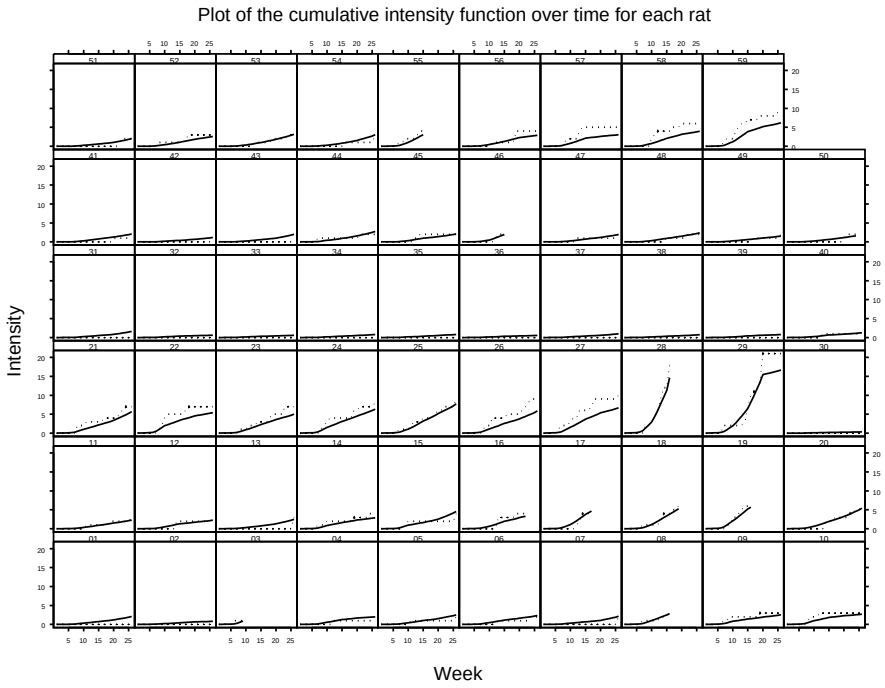

Figure 4: The observed and predicted intensity functions for each rat. The solid line denotes the estimated cumulative intensity function and the dotted line denotes the observed cumulative number of tumors for each rat.

directly using dynamic piecewise constant functions. Pennell and Dunson partition the time into 24 small intervals. We assume the number of knots and the location of knots in the baseline to be unknown so they are estimated by data sharing. According to our pseudo-marginal likelihood criterion, much smaller number of partitions would fit the data better. In summary, our approach is flexible enough to handle the dynamic evolution of the frailty and the baseline, yet simple enough to interpret their evolutions.

5.3 Sensitivity Analysis

Next, we evaluate the sensitivity of the model type IV with $m = 4$ (the one with the largest LPML from Table 1) to the choice of priors, we choose different values of ν_1, ν_2 and ξ for further evaluations. As we can

see from Table 2, the results are not sensitive to the choice of ν_1 and ξ, where the posterior mean, standard deviation, highest posterior density interval of β and LPML do not change much with different choices of ν_1 and ξ. The model fitting could be affected by the choice of ν_2, which is the inverse of the variance of random effects. This is not surprising because it is not reasonable to make the variance of random effects very large. We recommend trying some different values and choose the best fitting one by LPML, or just put a hyperprior on ν_2 as in Pennell and Dunson (2006) and estimate its value.

6 Discussion

We use a nonhomogeneous Poisson process to model the recurrent events for each subject, where tied events can be handled easily by counting the ties events in the Poisson process. The intensity functions for all subjects share the same baseline function, the same regression coefficients for the treatment effects, but with individual frailty functions. A frailty function is used to model the correlation among recurrent events of the same subject. We consider it as a constant function or a piecewise constant function. We construct a correlated process by specifying the dynamic evolution of frailties over the piecewise segments. Moreover, we construct a correlated process for the baseline intensity similarly. The piecewise constant approximation models are usually more appropriate and flexible than simple parametric methods. For example, in cases where a constant frailty for a subject exists, the piecewise dynamic frailty models could still provide an unbiased estimate for the treatment effect and a conservative estimate of its variance. While in a complex process, a model with constant frailty may underestimate the variance of the treatment effect and yield an inappropriate analysis. So the dynamic formulation of frailties and baseline intensity will be more appropriate for describing the evolution of risk that changes with time.

In this study, we implement MCMC algorithms for Bayesian inferences for models with stepwise frailty and baseline intensity functions. We also develop RJMCMC for models with a random number of change points in the baseline intensity function. Compared to other models, this model could adjust the number of change points to an optimal one and make valid inferences at the same time. While this may add model com-

plexity, we demonstrate that the computation is still tractable by the implementation of the RJMCMC algorithm. It provides insight into any significant dynamic changes of the baseline intensity during the study period. Using the pseudo-marginal likelihood criterion applied to a data set, we show it is desirable to extend models with a stepwise baseline intensity function from a fixed number of change points to a random one.

Acknowledgements

The authors would like to thank the guest editor and the referees for their constructive suggestions. No official support or endorsement by the Food and Drug Administration of this paper should be inferred.

References

Aslanidou, H., Dey, D. K., and Sinha, D., (1998), Bayesian analysis of multivariate survival data using Monte Carlo methods. Canadian Journal of Statistics, **26**, 33-48.

Chen, M. H., Ibrahim, J. G., and Sinha, D. (2002), Bayesian inference for multivariate survival data with a cure fraction. Journal of Multivariate Analysis, **80**, 101-126.

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

Dunson, D. B. and Chen, Z. (2004), Selecting factors predictive of heterogeneity in multivariate event time data. Biometrics, **60**, 352-358.

Dunson, D. B. and Dinse, G. E. (2000), Distinguishing effects on tumor multiplicity and growth rate in chemoprevention experiments. Biometrics, **56**, 1068-1075.

Fong, D. Y. T., Lam, K. F., Lawless, J. F., and Lee, Y. W. (2001), Dynamic random effects models for times between repeated events. Lifetime Data Analysis, **7**, 345-362.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992), Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In Bayesian Statistics 4, (Eds. J.M. Bernado, J.O. Berger, A.P. David, and A.F.M. Smith), Oxford: Oxford University Press.

Gelfand, A. E. and Dey, D. K. (1994), Bayesian model choice: asymptotics and exact calculations. Journal of the Royal Statistical Society, Series B, (Methodological), **56**, 501-514.

Gelfand, A. E. and Smith, A. F. M. (1990), Sampling based approaches to calculating marginal densities. Journal of the American Statistical Association, **85**, 398-409.

Green, P. J. (1995), Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. Biometrika, **82**, 711-732.

Grubbs, C. J., Eto, I., Juliana, M. M., and Whitaker, L. M. (1991), Effect of canthaxanthin on chemically induced mammary carcinogenesis. Oncology, **48**, 239-245.

Hatings, W. K. (1970), Monte Carlo sampling methods using Markov Chain and their applications. Biometrika, **57**, 97-109.

Henderson, R., and Shimakura, S. (2003), A serially correlated gamma frailty model for longitudinal count data. Biometrika, **90**, 355-366.

Kokoska, S. M., Hardin, J. M., Grubbs, C. J., and Hsu, C. (1993), The statistical analysis of cancer inhibition/promotion experiments. Anticancer Research, **13**, 1357-1364.

Lam, K. F., Lee, Y. W., and Leung, T. L. (2002), Modeling multivariate survival data by a semiparametric random effects proportional odds model. Biometrics, **58**, 316-323.

Lawles, J. F. and Nadeau, C. (1995), Some simple robust methods for the analysis of recurrent events. Technometrics, **37**, 158-168.

Lin, D. and Wei, L. (1989), The robust inference for the proportional hazards model. Journal of the American Statistical Association, **84**, 1074-1078.

Paik, M. C., Tsai, W. Y., and Ottman, R. (1994), Multivariate survival analysis using piecewise gamma frailty. Biometrics, **50**, 975-988.

Pennell, M. L. and Dunson, D. B. (2006), Bayesian semiparametric dynamic frailty models for multiple event time data. Biometrics, **62**, 1044-1052.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981), On the regression analysis of multivariate failure time data. Biometrika, **68**, 373-379.

Sahu, S. K., Dey, D. K., Aslanidou, H., and Sinha, D. (1997), A Weibull regression model with gamma frailties for multivariate survival data. Lifetime Data Analysis, **3**, 123-137.

Sinha, D. (1993), Semiparametric Bayesian analysis of multiple time event data. Journal of the American Statistical Association, **88**, 979-983.

Sinha, D. (1998), Posterior likelihood methods for multivariate survival data. Biometrics, **54**, 1463-1474.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979), The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography, **16**, 439-454.

Wei, L., Lin, D., and Weissfeld, L. (1989), Regression analysis of multivariate incomplete failure time data by modeling marginal distribution. Journal of the American Statistical Association, **84**, 1065-1073.

Xie, W., Lewis, P.O., Fan, Y., Kuo, L., and Chen, M.-H. (2011), Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Systematic Biology, **60**, 150-160.

Yue, H. and Chan, K. S. (1997), A dynamic frailty model for multivariate survival data. Biometrics, **53**, 785-793.