



**University of Dundee**

**Multivariate frailty models for multi-type recurrent event data and its application to cancer prevention trial**

Bedair, Khaled; Hong, Yili; Li, Jie; Al-Khalidi, Hussein R.

*Published in:*  
Computational Statistics and Data Analysis

*DOI:*  
[10.1016/j.csda.2016.01.018](https://doi.org/10.1016/j.csda.2016.01.018)

*Publication date:*  
2016

*Document Version*  
Peer reviewed version

[Link to publication in Discovery Research Portal](#)

*Citation for published version (APA):*

Bedair, K., Hong, Y., Li, J., & Al-Khalidi, H. R. (2016). Multivariate frailty models for multi-type recurrent event data and its application to cancer prevention trial. *Computational Statistics and Data Analysis*, 101, 161-173. DOI: 10.1016/j.csda.2016.01.018

**General rights**

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Multivariate Frailty Models for Multi-type Recurrent Event Data and Its Application to Cancer Prevention Trial

Khaled Bedair<sup>1,2</sup>, Yili Hong<sup>3</sup>, Jie Li<sup>3</sup>, and Hussein R. Al-Khalidi<sup>4</sup>

<sup>1</sup>Faculty of Commerce, Tanta University, Tanta, Egypt

<sup>2</sup>School of Medicine, University of Dundee, Dundee, UK

<sup>3</sup>Department of Statistics, Virginia Tech, Blacksburg, VA, USA

<sup>4</sup>Department of Biostatistics & Bioinformatics, Duke University, Durham, NC, USA

## Abstract

Multi-type recurrent event data arise in many situations when two or more different event types may occur repeatedly over an observation period. For example, in a randomized controlled clinical trial to study the efficacy of nutritional supplements for skin cancer prevention, there can be two types of skin cancer events occur repeatedly over time. The research objectives of analyzing such data often include characterizing the event rate of different event types, estimating the treatment effects on each event process, and understanding the correlation structure among different event types. In this paper, we propose the use of a proportional intensity model with multivariate random effects to model such data. The proposed model can take into account the dependence among different event types within a subject as well as the treatment effects. Maximum likelihood estimates of the regression coefficients, variance-covariance components, and the nonparametric baseline intensity function are obtained via a Monte Carlo Expectation Maximization (MCEM) algorithm. The expectation step of the algorithm involves the calculation of the conditional expectations of the random effects by using the Metropolis-Hastings sampling. Our proposed method can easily handle recurrent event data that have more than two types of events. Simulation studies were used to validate the performance of the proposed method, followed by an application to the skin cancer prevention data.

**Key Words:** Correlated frailty; MCMC; Proportional hazards model; Random effects; EM algorithm; Skin cancer.

# 1 Introduction

In many research fields, it is common to observe processes that generate events repeatedly over the follow-up time for a given subject. Such processes are called recurrent event processes and the generated data are referred to as recurrent event data. In clinical studies, patients may experience transient clinical events repeatedly over an observation period, such as occurrences of heart attack in cardiovascular studies, epileptic seizures in neurology studies, fractures in osteoporosis studies, and recurrence of bladder cancer tumors in oncology studies. Multi-type recurrent event data arise when two or more different kinds of events may occur repeatedly over an observation period. For example, in bone marrow transplantation, different types of recurrent infections (e.g., bacterial, fungal, and viral infections) can occur after the surgery.

The analyses of such multi-type recurrent event data often aim to answer scientific questions such as: what are the effects of explanatory variables (e.g., treatment) on the event process of different types, how to characterize the individual-to-individual difference in the event processes, and what is the correlation among the event process of different types. For example, in the Nutritional Prevention of Cancer Trial conducted by Arizona Cancer Center (see Section 5 for more details), the investigators were interested in the efficacy of nutritional supplement of selenium on preventing skin cancer of different types, as well as the heterogeneity and correlation among the recurrence of skin cancer of different types. Thus, a general method to analyze multi-type recurrent event data needs to be developed to address these scientific questions.

Multivariate frailty model are widely used to model recurrent event data (e.g., Duchateau et al. 2003, Manda and Meyer 2005, and McGilchrist and Yau 2008). Although there has been many developments on recurrent event data analysis, existing methods are not general or flexible enough to model and characterize multi-type recurrent event data. In this paper, we propose the use of a semiparametric proportional hazards model with correlated random effect to characterize multi-type recurrent event data. Monte Carlo Expectation Maximization (MCEM) algorithm is used to estimate unknown parameters in the model.

For a literature review, a general description of models for recurrent event data can be found in literature, for example, Cook and Lawless (2007). In terms of baseline modeling, Abu-Libdeh, Turnbull, and Clark (1990), and Cook et al. (1999) considered a frailty model with parametric baseline functions. Chen et al. (2005), Moreno (2008), and Chen and Cook (2009) used a piecewise constant baseline intensity function to analyze multi-type recurrent event data. For piece-wise constant function models, one needs to specify the locations and number of pieces for the intensity function. As it is pointed out by Friedman (1982), the estimator will be biased if the locations and the number of pieces are not correctly

specified. Thus, in comparison to the parametric and piecewise constant assumptions, the nonparametric baseline assumption used in this paper is more flexible.

For modeling of multiple types of events, Cai and Schaubel (2004) proposed a class of semi-parametric marginal means/rates models for multiple type recurrent event data, without random effects. Cook, Lawless, and Lee (2010) introduced a bivariate mixed Poisson model using a copula function to describe the correlation between frailties of the two types of events. They applied the Expectation-Maximization (EM) algorithm and used numerical integration in the expectation step (E step) to calculate the conditional expectations. However, numerical integration can be unstable, even when the number of event types is two. When the number of event types is more than two, it is challenging to make numerical integration work. The Monte Carlo sampling used in our method is more flexible than numerical integration in dealing with multi-type events, even for large number of event types. Rondeau, Mazroui, and Gonzalez (2012) developed an R package for the analysis of correlated survival data with frailty models. Mazroui et al. (2013) considered multivariate frailty models for two types of recurrent events with parametric baseline and spline based baseline functions. Mazroui et al. (2015) considered multivariate frailty models for two types of recurrent events with time-varying coefficients.

Although MCEM algorithms were used in parameter estimation for frailty models of single-type recurrent event data analysis (Vaida and Xu 2000, and Ripatti, Larsen, and Palmgren 2002), the generalization of the MCEM algorithm to multi-type recurrent event data is not a trivial task. The challenges mainly arise from the need to estimate multiple baseline functions, correlated frailty, and parameters in the frailty distribution, which are incorporated in this paper. In summary, the proposed method provides a flexible and general approach that can be directly implemented to analyze multi-type recurrent event data.

The remainder of this paper is organized as follows. Section 2 describes the data setup, the model for the intensity function, and the model for the random effects. Section 3 describes the MCEM algorithm that is used to estimate the unknown parameter in the model and statistical inference procedures. Section 4 uses simulations to validate the estimation and inference procedure developed in this paper. Section 5 applies the proposed method to the skin cancer dataset. Section 6 gives some concluding remarks and areas for further research.

## 2 Data and Model

In this section, we introduce the notation for the data and the models to describe the data.

## 2.1 Data

The  $k^{\text{th}}$  event time for subject  $i$  of event type  $j$  is denoted by  $t_{ijk}$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, J$ , and  $k = 1, \dots, N_{ij}(\tau_i)$ . Here  $m$  is the number of subjects under the study,  $J$  is the total number of event types,  $\tau_i$  is the length of follow-up time for subject  $i$ , and  $N_{ij}(t)$  is defined to be the number of type  $j$  events occurred over time interval  $(0, t]$  for subject  $i$ . The censoring indicator  $\delta_{ijk}$  equals to 1 if event type  $j$  is observed for subject  $i$  at time  $t_{ijk}$ , and  $\delta_{ijk} = 0$  otherwise. We also have information on covariates for subject  $i$ , denoted by  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ , where  $p$  is the number of covariates. Let  $\mathbf{t}_i$  denote the data for subject  $i$  which includes observed event times, censoring indicator and the covariates. We use  $\mathbf{t} = \{\mathbf{t}_1, \dots, \mathbf{t}_m\}$  to denote the dataset for all subjects.

## 2.2 Model for Event Intensities

Let  $Y_i(t) = I(t \leq \tau_i)$  be the at-risk process for subject  $i$ ,  $i = 1, \dots, m$ . The event history for subject  $i$  up to time  $t$  is defined as  $\mathcal{H}_i(t) = \{N_{ij}(s), \mathbf{x}_i, Y_i(s); j = 1, \dots, J, 0 \leq s < t\}$ . The intensity function for the type  $j$  events of subject  $i$  is defined as (e.g., Cook and Lawless 2007),

$$\lambda_{ij}[t|\mathcal{H}_i(t)] = \lim_{\Delta t \rightarrow 0} \frac{\Pr [N_{ij}(t + \Delta t) - N_{ij}(t) = 1 | \mathcal{H}_i(t)]}{\Delta t}.$$

In this paper, we use the following form to model the event intensity for type  $j$

$$\lambda_{ij}(t) = \lambda_{0j}(t) \exp(\mathbf{x}'_i \boldsymbol{\beta}_j + w_{ij}), \quad (1)$$

where  $i = 1, \dots, m$ , and  $j = 1, \dots, J$ . Here,  $\lambda_{0j}(t)$  is the baseline intensity function for event type  $j$ . The vector  $\boldsymbol{\beta}_j$  with dimension  $p \times 1$  contains the fixed effect parameters for the type  $j$  events, and  $w_{ij}$  is the random effect for the  $j^{\text{th}}$  event type of the  $i^{\text{th}}$  subject. The cumulative intensity function can be written as

$$\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(s) ds = \Lambda_{0j}(t) \exp(\mathbf{x}'_i \boldsymbol{\beta}_j + w_{ij}), \quad (2)$$

where  $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(s) ds$  is the baseline cumulative intensity function. In this paper, the baseline cumulative intensity function is left to be unspecified. That is we use a nonparametric form to describe the function  $\Lambda_{0j}(\cdot)$ ,  $j = 1, \dots, J$ . Note that the intensity function (1) can also be rewritten as

$$\lambda_{ij}(t) = \lambda_{0j}(t) u_{ij} \exp(\mathbf{x}'_i \boldsymbol{\beta}_j),$$

where  $u_{ij} = \exp(w_{ij})$  is called the frailty for the  $j^{\text{th}}$  event type of the  $i^{\text{th}}$  subject.

## 2.3 Model for Random Effects

Let  $\mathbf{w}_i = (w_{i1}, \dots, w_{iJ})'$  be the random effects vector for the  $i^{\text{th}}$  subject. A multivariate normal distribution was used to model the random effect  $\mathbf{w}_i$ . In particular, the random effects are independent and identically distributed (i.i.d.) with a multivariate normal distribution  $\text{MVN}(\mathbf{0}, \Sigma)$ . The variance-covariance matrix  $\Sigma$  for the  $i^{\text{th}}$  subject is

$$\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1J} \\ \vdots & \ddots & \vdots \\ \sigma_{J1} & \cdots & \sigma_{JJ} \end{bmatrix}.$$

Let  $\mathbf{w} = (\mathbf{w}'_1, \dots, \mathbf{w}'_m)'$  be a vector of random effects for all subjects. Then,  $\mathbf{w} \sim \text{MVN}(\mathbf{0}, \Sigma_{\mathbf{w}})$ , where  $\mathbf{0}$  is a zero vector, and the variance-covariance matrix  $\Sigma_{\mathbf{w}}$  can be defined as a block-diagonal matrix with all elements in the diagonal equal to  $\Sigma$ . We write  $\Sigma_{\mathbf{w}} = \text{diag}(\Sigma, \dots, \Sigma)$ .

## 3 Parameter Estimation

### 3.1 The Likelihood Function

Let  $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_J)'$  be the vector for the regression parameters for all types of events. We use  $\boldsymbol{\eta} = \{\boldsymbol{\beta}, \Lambda_{01}(\cdot), \dots, \Lambda_{0J}(\cdot)\}$  to denote the unknown parameters in the event intensity model in (1). Let  $\boldsymbol{\theta} = \text{vech}(\Sigma)$  be the vector for the parameters in  $\Sigma$ , which is the variance-covariance matrix for the distribution of the random effects. Here,  $\text{vech}(\Sigma)$  is the half vectorization of the matrix  $\Sigma$ . The set of unknown parameters included in the model is denoted  $\boldsymbol{\xi} = \{\boldsymbol{\eta}, \boldsymbol{\theta}\}$ . The likelihood function for  $\boldsymbol{\xi}$  based on data  $\mathbf{t}$  and random effects  $\mathbf{w}$  is

$$L(\boldsymbol{\xi}|\mathbf{t}, \mathbf{w}) = \prod_{i=1}^m \left[ \prod_{j=1}^J \left\{ \prod_{k=1}^{N_{ij}(\tau_i)} [\lambda_{ij}(t_{ijk})]^{\delta_{ijk}} \right\} \exp[-\Lambda_{ij}(\tau_i)] f(\mathbf{w}_i; \boldsymbol{\theta}) \right],$$

where  $f(\mathbf{w}_i; \boldsymbol{\theta})$  is the probability density function (pdf) of  $\mathbf{w}_i$ . Here  $\lambda_{ij}(\cdot)$  and  $\Lambda_{ij}(\cdot)$  are defined in (1) and (2), respectively. The log-likelihood function is  $\mathcal{L}(\boldsymbol{\xi}|\mathbf{t}, \mathbf{w}) = \log[L(\boldsymbol{\xi}|\mathbf{t}, \mathbf{w})]$ . In particular,

$$\mathcal{L}(\boldsymbol{\xi}|\mathbf{t}, \mathbf{w}) = \mathcal{L}_1(\boldsymbol{\eta}|\mathbf{t}, \mathbf{w}) + \mathcal{L}_2(\boldsymbol{\theta}|\mathbf{w}), \quad (3)$$

where

$$\mathcal{L}_1(\boldsymbol{\eta}|\mathbf{t}, \mathbf{w}) = \sum_{i=1}^m \sum_{j=1}^J \left( \sum_{k=1}^{N_{ij}(\tau_i)} \delta_{ijk} \{ \log[\lambda_{0j}(t_{ijk})] + (\mathbf{x}'_i \boldsymbol{\beta}_j + w_{ij}) \} - \Lambda_{0j}(\tau_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_j + w_{ij}) \right), \quad (4)$$

and

$$\mathcal{L}_2(\boldsymbol{\theta}|\mathbf{w}) = -\frac{Jm}{2} \log(2\pi) - \frac{m}{2} \log(|\Sigma|) - \frac{1}{2} (\mathbf{w}' \Sigma_{\mathbf{w}}^{-1} \mathbf{w}).$$

The pdf of the conditional distribution of  $\mathbf{w}$  given  $\mathbf{t}$  is denoted by  $g(\mathbf{w}|\mathbf{t})$ ,

$$g(\mathbf{w}|\mathbf{t}) = \frac{L(\boldsymbol{\xi}|\mathbf{t}, \mathbf{w})}{\int_{-\infty}^{\infty} L(\boldsymbol{\xi}|\mathbf{t}, \mathbf{w}) d\mathbf{w}}.$$

It can be shown that the pdf of  $\mathbf{w}|\mathbf{t}$  is

$$g(\mathbf{w}|\mathbf{t}) \propto \exp \left[ \sum_{i=1}^m \sum_{j=1}^J N_{ij}(\tau_i) w_{ij} - \Lambda_{0j}(\tau_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_j) \exp(w_{ij}) - \frac{1}{2} (\mathbf{w}'_i \boldsymbol{\Sigma}^{-1} \mathbf{w}_i) \right].$$

### 3.2 The EM Algorithm

The EM algorithm provides a tool for obtaining maximum likelihood estimates under models that yield analytically formidable likelihood equations. It is an iterative routine requiring two primary calculations at each iteration: computation of a particular conditional expectation of the log-likelihood (E step) and maximization this expectation over the corresponding parameters (M step). The Monte Carlo EM (MCEM), introduced by Wei and Tanner (1990), is a modification of the EM algorithm where the expectation in the E step is computed numerically through Markov chain Monte Carlo (MCMC) methods such as the Gibbs sampler and Metropolis-Hastings sampling as described in Robert and Casella (2009), and McLachlan and Krishnan (2007).

The EM algorithm will start with an initial value of  $\boldsymbol{\xi}$ , denoted by  $\widehat{\boldsymbol{\xi}}^{(0)}$ , which can be obtained by fitting a Gaussian frailty model separately for each type of events (i.e., set the correlations among different event types to be 0). Then it iteratively carries out the E step and M step as described below until convergence.

#### E Step:

The estimates for the parameters at the  $q^{th}$  iteration of the EM algorithm are represented by  $\widehat{\boldsymbol{\xi}}^{(q)}$ . In the E step, we need to find the conditional expectation of the full log-likelihood function with respect to the conditional distribution  $\mathbf{w}|\mathbf{t}$ . That is to find

$$Q(\boldsymbol{\xi}|\widehat{\boldsymbol{\xi}}^{(q)}) = E_{\mathbf{w}|\mathbf{t}}[\mathcal{L}(\boldsymbol{\xi}|\mathbf{t}, \mathbf{w})]. \quad (5)$$

Because there is no closed-form expression for  $g(\mathbf{w}|\mathbf{t})$ , we use MCMC to simulate sample from  $g(\mathbf{w}|\mathbf{t})$ . The Metropolis-Hastings algorithm steps are shown in **Algorithm 1**.

#### Algorithm 1:

1. Choose an initial value  $\mathbf{w}^{(0)} = (\mathbf{w}_1^{(0)'}, \dots, \mathbf{w}_m^{(0)'})'$ .
2. At iteration  $s$ , draw a candidate value  $\mathbf{w}_i^*$ , for the  $i^{th}$  component of  $\mathbf{w}$  from a proposal distribution  $q(\mathbf{w}_i^*|\mathbf{w}_i^{(s-1)})$ , where a standard proposal distribution would be the

multivariate normal distribution with mean equal to the current value  $\mathbf{w}_i^{(s-1)}$  and with some proposal variance.

3. Compute an acceptance ratio (probability)

$$r^{(s)} = \frac{g(\mathbf{w}_i^*|\boldsymbol{\xi}) q(\mathbf{w}_i^{(s-1)}|\mathbf{w}_i^*)}{g(\mathbf{w}_i^{(s-1)}|\boldsymbol{\xi}) q(\mathbf{w}_i^*|\mathbf{w}_i^{(s-1)})}.$$

4. Accept  $\mathbf{w}_i^*$  as the new value with probability  $\min(1, r^{(s)})$ ; otherwise,  $\mathbf{w}_i^{(s)} = \mathbf{w}_i^{(s-1)}$ . Repeat step 3 and 4 until finishing updating each element in  $\mathbf{w}$ .
5. Repeat steps 2 to 4 for generating  $\tilde{S}$  samples  $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(\tilde{S})}$ .

### M Step:

The goal of the M step is to find the value of  $\boldsymbol{\xi}$  that maximizes  $Q(\boldsymbol{\xi}|\hat{\boldsymbol{\xi}}^{(q)})$  in (5). The maximizer is denoted by  $\hat{\boldsymbol{\xi}}^{(q+1)}$ . Note that

$$E_{\mathbf{w}|\mathbf{t}}[\mathcal{L}(\boldsymbol{\xi}|\mathbf{t}, \mathbf{w})] = E_{\mathbf{w}|\mathbf{t}}[\mathcal{L}_1(\boldsymbol{\eta}|\mathbf{t}, \mathbf{w})] + E_{\mathbf{w}|\mathbf{t}}[\mathcal{L}_2(\boldsymbol{\theta}|\mathbf{w})].$$

Hence the maximization of  $Q(\boldsymbol{\xi}|\hat{\boldsymbol{\xi}}^{(q)})$  can be done by maximizing  $E_{\mathbf{w}|\mathbf{t}}[\mathcal{L}_1(\boldsymbol{\eta}|\mathbf{t}, \mathbf{w})]$  and  $E_{\mathbf{w}|\mathbf{t}}[\mathcal{L}_2(\boldsymbol{\theta}|\mathbf{w})]$  separately. Because  $E_{\mathbf{w}|\mathbf{t}}[\mathcal{L}_1(\boldsymbol{\eta}|\mathbf{t}, \mathbf{w})]$  involves both  $\boldsymbol{\beta}$  and nonparametric baselines  $\Lambda_{0j}(\cdot)$ , it is challenging to do a direct maximization. Thus we use the profile likelihood approach to estimate  $\boldsymbol{\beta}$  first. In this case the log profile likelihood function for  $\boldsymbol{\beta}$  is equivalent to the conditional expectation of the log partial likelihood function. In particular, the conditional expectation of log partial likelihood function is

$$E_{\mathbf{w}|\mathbf{t}}[\mathcal{P}\mathcal{L}(\boldsymbol{\beta})] = \sum_{j=1}^J \sum_{l=1}^{L_j} \left\{ \mathbf{x}'_i \boldsymbol{\beta}_j + E_{\mathbf{w}|\mathbf{t}}(w_{ij}) - \log \left[ \sum_{i \in R_{jl}} E_{\mathbf{w}|\mathbf{t}}(u_{ij}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_j) \right] \right\}. \quad (6)$$

Here,  $R_{jl}$  is the risk set at time  $t_{jl}$ , which contains all individuals that are still under study at time  $t_{jl}$ . Those time points  $t_{jl}$  are the ordered distinct event times for type  $j$  events. That is,  $t_{j1} < \dots < t_{jl} < \dots < t_{jL_j}$  where  $L_j$  is the number of distinct event time points for type  $j$ ,  $j = 1, \dots, J$ . Also, let  $d_{jl}$  be the total number of event of type  $j$  occurred at time  $t_{jl}$ .

The required conditional expectations  $E_{\mathbf{w}|\mathbf{t}}(w_{ij})$ , and  $E_{\mathbf{w}|\mathbf{t}}(u_{ij})$  in (6) are not available in closed forms. It is, however, possible to approximate these integrals by numerical methods or Monte Carlo simulation. Numerical integration was used for bivariate gamma frailties by Cook, Lawless, and Lee (2010). This solution is feasible, however, only for low-dimensional random vectors  $\mathbf{w}_i$  (Vaida and Xu 2000). The MCEM method used in this paper, however, provides a flexible way to handle more than two types of events.



At E step, we obtain  $S$  samples after burn-in and thinning from the conditional distribution of the random effects. Based on graphical checks of the MCMC trace plot, we discarded 5,000 samples for burn-in in the simulation study and data analysis. We use the autocorrelation function (ACF) plot to diagnose the autocorrelations among the MCMC samples. For the simulation study and data analysis, we took a sample of every 10 samples based on the ACF plots. For the computation of the conditional mean and the variance-covariance matrix, we used  $S = 10,000$  thinned samples, which is typically large enough to make the MCMC error negligible. Because the MCMC algorithm was implemented in C, the MCMC sampling is efficient computationally.

With the thinned samples, the conditional expectations  $E_{\mathbf{w}|t}(w_{ij})$  and  $E_{\mathbf{w}|t}(u_{ij})$  can be estimated by

$$E_{\mathbf{w}|t}(w_{ij}) = \frac{1}{S} \sum_{s=1}^S w_{ij}^{(s)}, \text{ and } E_{\mathbf{w}|t}(u_{ij}) = \frac{1}{S} \sum_{s=1}^S \exp(w_{ij}^{(s)}).$$

The estimate,  $\widehat{\boldsymbol{\beta}}^{(q+1)} = (\widehat{\boldsymbol{\beta}}_1^{(q+1)'}, \dots, \widehat{\boldsymbol{\beta}}_J^{(q+1)'})'$ , can thus be obtained by maximizing  $E_{\mathbf{w}|t}[\mathcal{PL}(\boldsymbol{\beta})]$  in (6). Because there is no closed-form expression for the maximum likelihood (ML) estimate  $\widehat{\boldsymbol{\beta}}^{(q+1)}$ , numerical optimization procedures are used. In particular, we used the `optim()` function in R (2013) for optimization.

For estimating the baseline cumulative intensity functions, we used the profile likelihood approach at which we fix  $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^{(q+1)}$  in (4), and used a step function to estimate the baseline cumulative intensity function, which only jumps at the distinct event time points. This leads the Breslow type of estimator. In particular, the estimate of the  $\Lambda_{0j}(t)$  is given by

$$\widehat{\Lambda}_{0j}^{(q+1)}(t) = \sum_{t_{jl} \leq t} \frac{d_{jl}}{\sum_{i \in R_{jt}} E_{\mathbf{w}|t}(u_{ij}) \exp(\mathbf{x}'_i \widehat{\boldsymbol{\beta}}_j^{(q+1)})}, \quad j = 1, \dots, J.$$

By maximizing  $E_{\mathbf{w}|t}[\mathcal{L}_2(\boldsymbol{\theta}|\mathbf{w})]$ , the maximum likelihood estimate for the variance-covariance matrix can be given in the form

$$\widehat{\Sigma} = \frac{1}{m} \sum_{i=1}^m E_{\mathbf{w}|t}(\mathbf{w}_i \mathbf{w}_i') = \frac{1}{Sm} \sum_{s=1}^S \sum_{i=1}^m \mathbf{w}_i^{(s)} \mathbf{w}_i^{(s)'}$$

Note in this case,  $\boldsymbol{\theta}$  is the unknown parameters in  $\Sigma$ .

We summarize the MCEM algorithm as follows.

**Algorithm 2:**

1. Choose a starting point  $\widehat{\boldsymbol{\xi}}^{(0)}$ .
2. (E step), at step  $q$ , using the current parameter values  $\widehat{\boldsymbol{\xi}}^{(q)}$ , generate  $S$  thinned samples  $(\mathbf{w}^{(1)'}, \dots, \mathbf{w}^{(S)'})'$  from the conditional distribution of  $g(\mathbf{w}|\boldsymbol{\xi})$  using the Metropolis-Hastings algorithm in **Algorithm 1**.

3. (M step), given the MCMC samples from the conditional distribution in Step 2, then
  - (a) Find estimates for  $\widehat{\boldsymbol{\beta}}^{(q+1)}$  that maximize  $E_{\mathbf{w}|t}[\mathcal{P}\mathcal{L}(\boldsymbol{\beta})]$ .
  - (b) Given  $\widehat{\boldsymbol{\beta}}^{(q+1)}$  in Step 3(a), compute the estimates for the baseline cumulative intensity function  $\widehat{\Lambda}_{0j}^{(q+1)}(\cdot)$ ,  $j = 1, \dots, J$ .
  - (c) Maximize  $E_{\mathbf{w}|t}[\mathcal{L}_2(\boldsymbol{\theta}|\mathbf{w})]$  to obtain  $\widehat{\boldsymbol{\theta}}^{(q+1)}$ .
4. If convergence is achieved, declare the current values to be the maximum likelihood parameter estimates; otherwise return to Step 2.

When the relative change in parameter values from two successive iterations is small then it is considered as convergence achieved. That is, one can apply the stopping rule suggested by Booth and Hobert (1999), which is

$$\max_d \left( \left| \frac{\widehat{\boldsymbol{\xi}}_d^{(s)} - \widehat{\boldsymbol{\xi}}_d^{(s-1)}}{\widehat{\boldsymbol{\xi}}_d^{(s-1)} - \delta_1} \right| \right) < \delta_2, \quad (7)$$

where  $(d = 1, \dots, D)$  is the number of parameters in the parameter vector  $\boldsymbol{\xi}$ , and  $\delta_1$  and  $\delta_2$  are predetermined values. Alternatively, Levine and Casella (2001) suggested a stopping rule after a fixed number of iterations, which is a useful stopping rule for practice. Graphical tools can be used for checking the convergence of the algorithm.

### 3.3 Statistical Inference

The EM algorithm provides only the parameter estimates. For statistical inference, the variance-covariance matrix is often needed. Because the baseline cumulative intensity function is nonparametric, we use the discrete local information matrix for inference. The discrete local information matrix is usually used for the statistical inference of frailty models (e.g., Chapter 5 of Duchateau and Janssen 2008).

In the discrete local information matrix, the parameters are  $\boldsymbol{\beta}$ ,  $\boldsymbol{\theta}$ , and the amount of jump in the step functions for the baseline. Those jumps are denoted by  $\lambda_{jl} = \Lambda_{0j}(t_{jl}) - \Lambda_{0j}(t_{j,l-1})$ ,  $l = 1, \dots, L_j$ ,  $j = 1, \dots, J$ . Let  $\boldsymbol{\lambda}_j = (\lambda_{j1}, \dots, \lambda_{jL_j})'$ , and  $\boldsymbol{\lambda} = (\boldsymbol{\lambda}'_1, \dots, \boldsymbol{\lambda}'_J)'$ . We use  $\boldsymbol{\zeta} = (\boldsymbol{\beta}', \boldsymbol{\theta}', \boldsymbol{\lambda}')$  to denote all the discrete parameter. Now we can compute the variance-covariance matrix corresponding to  $\boldsymbol{\zeta}$ . The Louis's formula in Louis (1982) is used for finding estimates of the variance-covariance matrix. In particular,

$$I_{\widehat{\boldsymbol{\zeta}}} = E_{\mathbf{w}|t} \left( -\frac{\partial^2 \mathcal{L}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta} \partial \boldsymbol{\zeta}'} \Big| \widehat{\boldsymbol{\zeta}} \right) - E_{\mathbf{w}|t} \left[ \frac{\partial \mathcal{L}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}} \frac{\partial \mathcal{L}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}'} \Big| \widehat{\boldsymbol{\zeta}} \right]. \quad (8)$$

The formulae for the first and second derivatives are available in Appendix A. One can compute the terms in (8) using samples of the last iteration of the EM steps to obtain  $I_{\hat{\zeta}}$ .

The variance-covariance matrix of  $\zeta$  is obtained as  $\Sigma_{\hat{\zeta}} = I_{\hat{\zeta}}^{-1}$ . Let  $\Sigma_{\hat{\beta}}$ ,  $\Sigma_{\hat{\theta}}$ , and  $\Sigma_{\hat{\lambda}_j}$  be the diagonal block matrices of  $\Sigma_{\hat{\zeta}}$  corresponding to parameter vectors  $\beta$ ,  $\theta$ , and  $\lambda_j$ , respectively. According to the asymptotic theory in Parner (1998), the asymptotic normality holds for the ML estimators for  $\hat{\beta}$  and  $\hat{\theta}$ . That is,  $\hat{\beta} \sim \text{MVN}(\beta, \Sigma_{\hat{\beta}})$ , and  $\hat{\theta} \sim \text{MVN}(\theta, \Sigma_{\hat{\theta}})$ , which can be used for statistical hypothesis testing and construction of confidence intervals for  $\beta$  and  $\theta$ .

The estimator of  $\Lambda_{0j}(t)$  can be expressed as  $\hat{\Lambda}_{0j}(t) = g_j(t, \hat{\lambda}_j)$ , where  $g_j(t, \lambda_j) = \sum_{t_{jl} \leq t} \lambda_{jl}$ ,  $j = 1, \dots, J$ . The asymptotic normality does not hold for the estimators for the jumps  $\hat{\lambda}$ . The estimator for the cumulative intensity function  $\hat{\Lambda}_{0j}(\cdot)$ , however, asymptotically follows a Gaussian process with the mean function  $\Lambda_{0j}(\cdot)$  and a variance function  $V_{0j}(\cdot)$  (Parner 1998). The variance function  $V_{0j}(\cdot)$  can be consistently estimated by

$$\hat{V}_{0j}(\cdot) = \left[ \frac{g_j(\cdot, \lambda_j)}{\partial \lambda_j'} \Big|_{\lambda_j = \hat{\lambda}_j} \right] \Sigma_{\hat{\lambda}_j} \left[ \frac{g_j(\cdot, \lambda_j)}{\partial \lambda_j} \Big|_{\lambda_j = \hat{\lambda}_j} \right].$$

The statistical hypothesis testing and construction of confidence interval for  $\Lambda_{0j}(t)$  can be based on the asymptotic normality of  $\hat{\Lambda}_{0j}(t)$ .

## 4 Simulation Studies

In this section, simulations were used to examine the empirical performance of our proposed method in finite samples.

### 4.1 Simulation Settings

To illustrate the flexibility of the proposed approach, we simulated data where three types of events may occur. There was one treatment variable with two levels (e.g., treatment vs. placebo) in the simulated data. Different sample sizes were also simulated to study the effect of sample size on estimation performance. The details of simulation are described as follows.

1. Set the study sample size  $m = 100, 200$ , and  $500$ , respectively. The regression coefficients were set to be  $\beta_1 = -1, \beta_2 = -0.5$ , and  $\beta_3 = -0.8$  for each of the three event types, respectively.
2. The multivariate random effect  $\mathbf{w}_i = (w_{i1}, w_{i2}, w_{i3})'$  is generated from a multivariate normal distribution with mean  $(0, 0, 0)'$  and variance-covariance matrix  $\Sigma$ . The

variance-covariance matrix is set to be

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & .4 & .4 \\ .4 & 1 & .4 \\ .4 & .4 & 1 \end{bmatrix}, \text{ and } \Sigma = \begin{bmatrix} 1 & .8 & .8 \\ .8 & 1 & .8 \\ .8 & .8 & 1 \end{bmatrix},$$

corresponding to independent, moderate, and high correlation cases. That is, the values of all variances  $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$  and the values of all covariance entries  $\sigma = \sigma_{12} = \sigma_{13} = \sigma_{23}$  are set to be 0, 0.4, and 0.8.

3. The treatment effect variable  $x_i$  is generated from a Bernoulli distribution with  $p = 0.5$ .
4. The maximum follow up time  $C$  is set to be 1 or 2 to represent different study lengths. The random censoring time  $C_i^*$  is assumed to be exponentially distributed with rate  $\alpha_c = 0.5$ . Here  $C_i$  is the censored time for subject  $i^{th}$  and  $C_i = \min \{C_i^*, C\}$ .
5. For each event type ( $j = 1, 2, 3$ ), for simplicity, set  $\lambda_{0j}(t) = 1$  which makes the recurrent event process to be a Poisson process. This means we can simulate the inter-arrival time,  $z_{ijl}$ , between two events from an exponential distribution with rate parameter  $\alpha_j = [\exp(\mathbf{x}'_i \boldsymbol{\beta}_j + w_{ij})]$ .
6. The observed event times  $t_{ijk} = \min(C_i, Y_{ijk})$ , where  $Y_{ijk} = \sum_{l=1}^k z_{ijl}$ . The censoring indicator  $\delta_{ijk} = 1$  if  $t_{ijk} = Y_{ijk}$  and  $\delta_{ijk} = 0$  otherwise.

## 4.2 Simulation Results

We used 1000 simulation runs. All calculations for the simulation studies were performed using parallel computing with R (2013). The results are obtained based on the algorithms of Section 3. The empirical mean square errors (MSE), variance, average of the estimated variance, and bias of the estimators were obtained. The coverage probability of the confidence interval procedures for unknown parameters were also obtained.

Due to the page limit, we only report several representative results here. The complete simulation results are available in supplementary Section 1. Figure 1 shows the empirical MSE for the estimators for the treatment effects  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$ , the parameters in the variance-covariance matrix. In Figure 1, there are six panels in  $2 \times 3$  format. The censoring times that we used are  $C = 1$  and  $C = 2$ . The values of the covariance entries are  $\sigma = 0, 0.4$ , and 0.8. The MSE is plotted as a function of number of subjects ( $m = 100, 200$ , and 500) in each panel for each parameter. Figure 2 shows the pointwise empirical MSE for the estimator of  $\Lambda_{01}(\cdot)$ . In each panel of Figure 2, the pointwise MSE is plotted as a function of time, for different values of  $m = 100, 200$ , and 500. Figure 3 shows the empirical coverage probability

for confidence interval procedures for  $\beta$  and  $\theta$ . Figure 4 shows the pointwise empirical coverage probability for the confidence interval procedure for  $\Lambda_{01}(\cdot)$ .

From the comprehensive simulation results, we observed the following:

- In general, the MSE decrease as the number of subjects increases, and longer follow-up time leads to a smaller MSE, because more events are observed. Higher correlation in the variance-covariance matrix of random effects tends to have a higher MSE.
- The results for the variance, estimated variance, and bias are similar. The estimated variance tends to be larger than the variance but they are getting closer when the number of subjects increase or the follow-up period is extended.
- The performance of the coverage probability is around the 0.95 nominal level, and it is getting closer to 0.95 when the number of subjects increases. For the baseline intensity functions  $\Lambda_{0j}(\cdot)$ 's, the improvement is more evident on the early part (i.e., when  $t$  is small) of the function, when  $m$  increases.

In summary, the simulation results for the regression coefficients  $\beta$ , variance-covariance parameters  $\theta$ , and the baseline intensity functions  $\Lambda_{0j}(\cdot)$ 's are as expected and show that our procedure works well.

## 5 Application to a Real Dataset

We applied our proposed model to the data collected as a part of the Nutritional Prevention of Cancer trial conducted by Arizona Cancer Center (e.g., Abu-Libdeh et al. 1990, Clark et al. 1996, and Duffield-Lillico et al. 2003) to study the efficacy of a nutritional supplement of selenium in the prevention of skin cancers among high-risk patients. Patients with previous history of either multiple basal cell carcinoma (BCC) or a squamous cell carcinoma (SCC) or otherwise at high risk for skin cancer were recruited into the trial in one of seven clinics. A total of 1,312 patients were enrolled in the study. There were 120 patients found to have incomplete follow-up data for the tumor types thus we exclude them from the analysis. We applied the analysis to a total of 1,192 patients with 606 and 586 randomly assigned to treatment (Selenium) group ( $x_i = 1$ ) and placebo group ( $x_i = 0$ ), respectively. Patients received a nutritional supplement of 200 micro-gram of selenium per day or a placebo. Patients were followed closely for a minimum of 4 years by means of frequent and regular clinic visits. They were also instructed on self-examination for appearance of skin lesions and told to report to the clinic if they observe any suspicious symptoms without waiting for the next scheduled appointment.

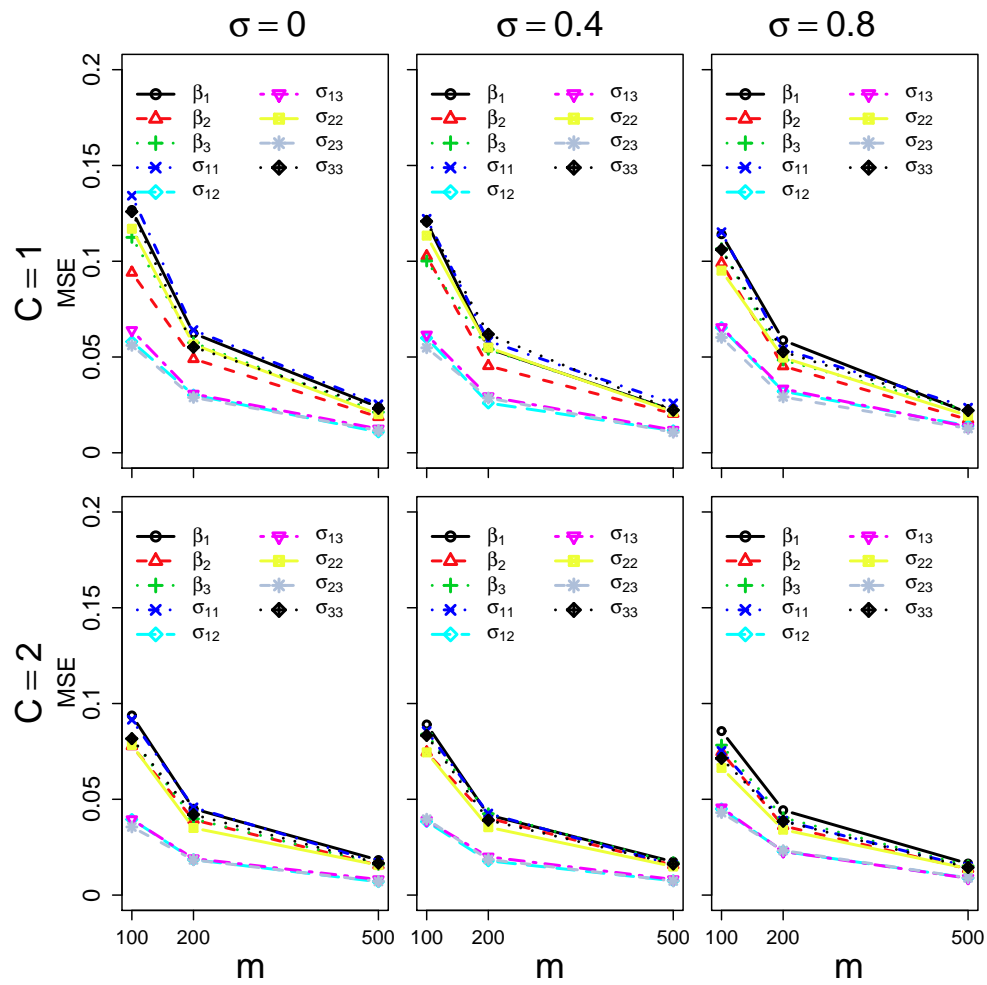


Figure 1: Empirical MSE for estimators of  $\beta$  and  $\theta$ .

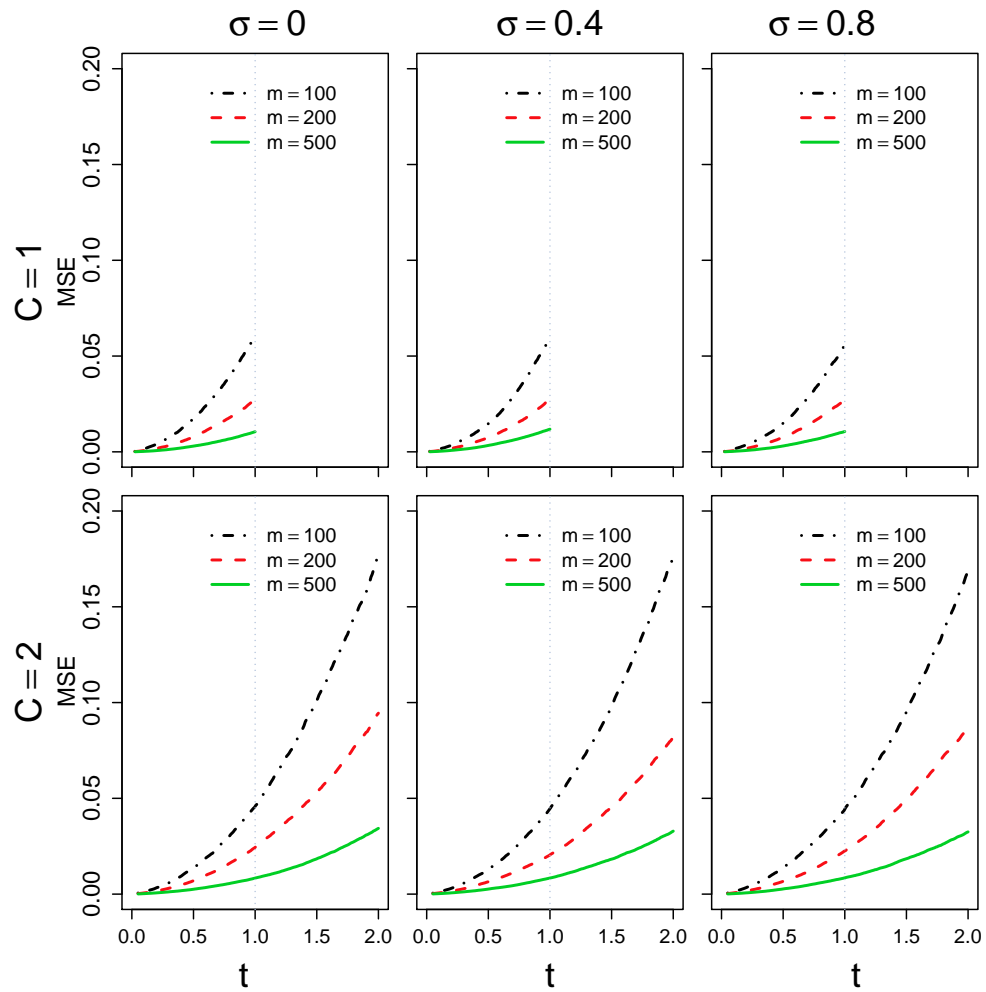


Figure 2: Pointwise empirical MSE for the estimator of  $\Lambda_{01}(\cdot)$ .

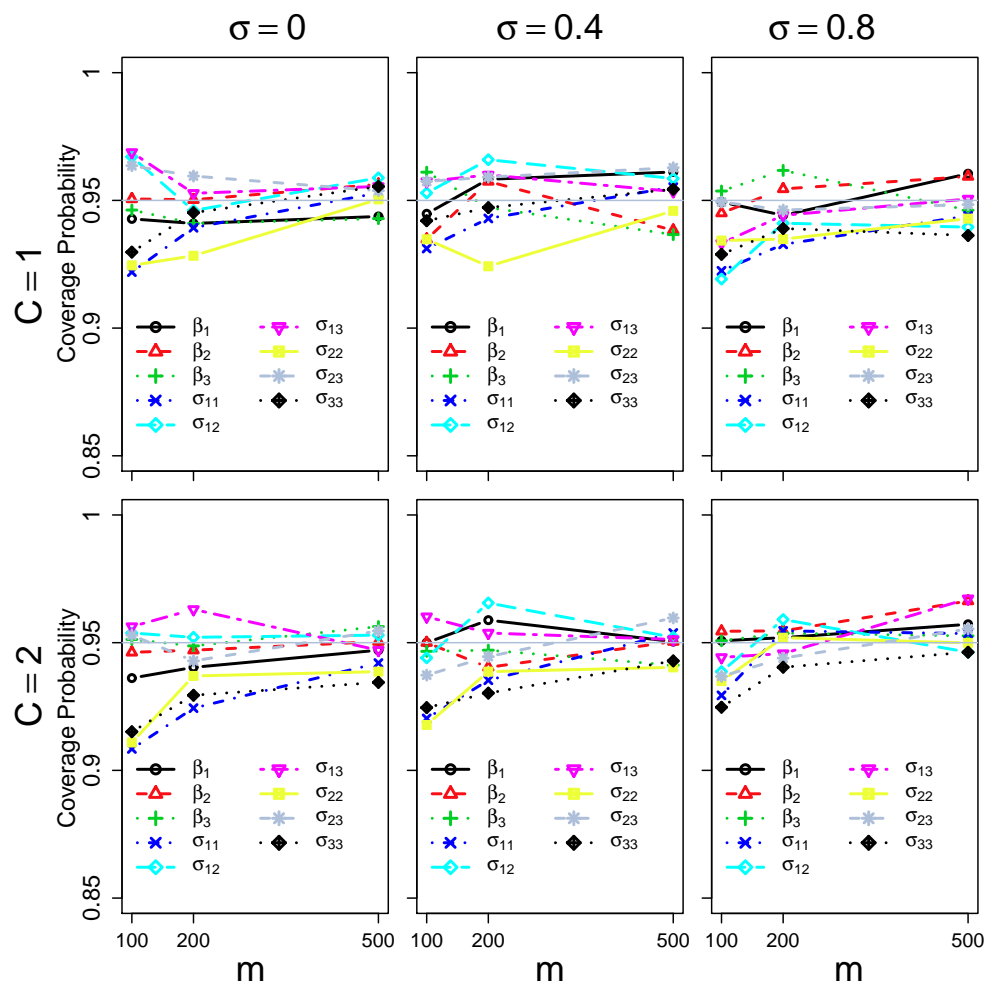


Figure 3: Empirical coverage probability for confidence interval procedures for  $\beta$  and  $\theta$ .



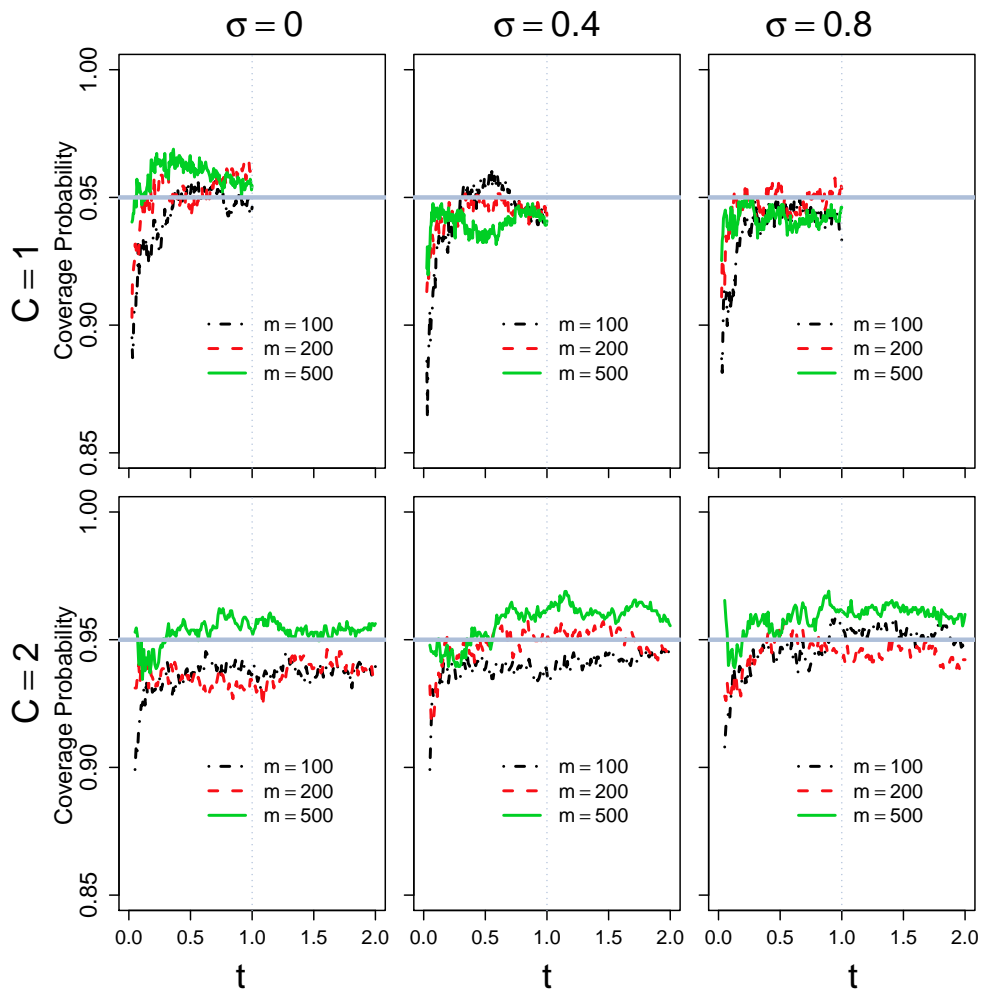


Figure 4: Pointwise empirical coverage probability for the confidence interval procedure for  $\Lambda_{01}(\cdot)$ .

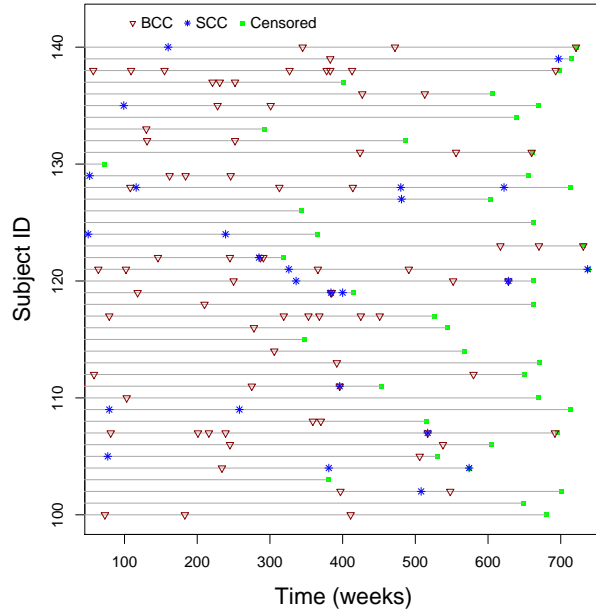


Figure 5: Multi-type event plot for tumor occurrence times and types for a sample of 40 patients of the skin cancer dataset.

Events of interest were the successive incidence times of the two types of tumors. The first type was the detection of one or more basal cell carcinoma (BCC) at a clinic visit, while the second type was the medical detection of one or more squamous cell carcinoma (SCC) at a clinic visit. Figure 5 shows the tumor occurrence times and types for a sample of 40 patients in the skin cancer dataset. Some patients did not develop any new events (e.g., subjects 101, 103, and 130), some patients were observed to develop only new events of a particular type (e.g., subjects 100, 110, and 124), and some experienced both types (e.g., subjects 102, 120, and 140). The last follow-up time point was treated as a censoring point. Descriptive statistics for the total number of observed events according to the cancer type and treatment group are given in Table 1.

We fitted the proposed multi-type model using the MCEM algorithm to the skin cancer data with two types of events (BCC - type I; SCC - type II). The initial values for  $\xi$  were the estimated values from the corresponding Cox model with Gaussian frailties separately for each type of event. In the E step of the MCEM algorithm, we used the Metropolis-Hastings algorithm with 100,000 samples after a burn-in sample of 5,000. The convergence of the MCMC was checked through visual inspection. Supplementary Figure 21 shows the MCMC trace plot, which shows the Monte Carlo chains converge well. We also used the autocorrelation function (ACF) plot to assess the autocorrelation among the MC samples. Supplementary Figure 21 also shows the ACF plots for the Monte Carlo samples. Based on

Table 1: Number of observed events ( $N$ ) and patients ( $m$ ) affected by cancer types according to the cancer type and treatment group.

Cancer type	Total (1192 subject)			Placebo (586 subject)			Selenium (606 subject)		
	$m$	$N$	$N/m$	$m$	$N$	$N/m$	$m$	$N$	$N/m$
BCC	856	3453	2.896	407	1582	2.699	449	1871	3.087
SCC	374	759	0.637	172	351	0.599	202	408	0.673
BCC or SCC	924	4212	3.533	434	1933	3.298	490	2279	3.760

Table 2: Estimates of the treatment effects of BCC and SCC event type for the proposed multi-type model.

Para.	EST.	S.E.	p-value	95 % CI		RR EST.	95 % CI	
				lower	upper		lower	upper
$\beta_1$	0.088	0.072	0.221	-0.053	0.229	1.092	0.948	1.258
$\beta_2$	0.122	0.128	0.340	-0.128	0.372	1.129	0.880	1.450
$\sigma_{11}$	1.027	0.072	< 0.001	0.887	1.167			
$\sigma_{12}$	0.457	0.089	< 0.001	0.282	0.632			
$\sigma_{22}$	2.044	0.205	< 0.001	1.642	2.446			

the ACF plots, we took a thinned sample of every 10 Monte Carlo samples. Thus, we used  $S = 10,000$  thinned samples in the MCEM algorithm.

Table 2 presents the estimates for the treatment effects and variance-covariance component parameters and the corresponding 95% confidence intervals based on the multi-type model. Figure 6 shows the estimated baseline cumulative intensity functions and corresponding 95% point-wise confidence intervals for the BCC and the SCC, based on the multi-type model. The results show that there is an estimated 9.2% increase in the rate of BCC in the Selenium group compared to the placebo group with an estimated relative risk  $RR(\beta_1) = 1.092$  (95% CI: [0.948, 1.258]; p-value: 0.221). There is an estimated 12.9% increase in the rate of SCC in the Selenium group compared to the placebo group with estimated relative risk  $RR(\beta_2) = 1.129$  (95% CI: [0.880, 1.450]; p-value: 0.340). The variances of the random effects are estimated as  $\hat{\sigma}_{11} = 1.027$  (95% CI: [0.887, 1.167]) and  $\hat{\sigma}_{22} = 2.044$  (95% CI : [1.642, 2.446]). The estimate of the covariance between random effects for the two event types is 0.457 (95% CI : [0.282, 0.632]). The correlation coefficient ( $\rho_{12}$ ) is estimated at 0.315 which indicates a mild to medium correlation between the occurrence of the two cancer types.

We did some additional analyses on the skin cancer data for some comparisons and to gain some more insights. Existing approaches include single-type Andersen-Gill (AG) analysis

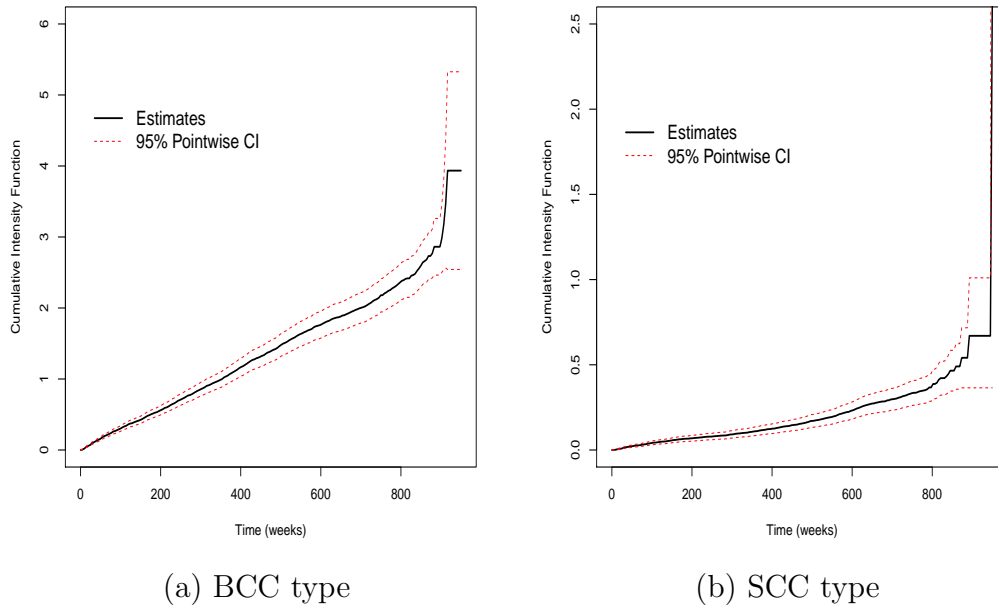


Figure 6: Cumulative intensity functions for BCC and SCC tumor types.

with robust variance, single-type semiparametric model using the normal distribution for the random effects (Gaussian), and single-type semiparametric frailty gamma model (Gamma). Table 3 gives a summary of alternative model fittings to the skin cancer dataset. The AG1 model in Table 3 is the AG model according to the method described in pages 185-186 of Therneau and Grambsch (2000). In this model, there is only one covariate, which is the treatment group (treatment vs placebo). Thus, the model essentially is a time-varying Poisson process. In this special case, the AG model and the model in Lin et al. (2000) are the same. Because the independent increments model assumption may not be met, the robust variance is typically used for inference. Overall, the results from the single-type frailty models and the AG1 model reveal broadly comparable estimates for both the treatment effects. The results from the multi-type model, however, has a smaller p-value (0.340 vs 0.498) for the treatment effect in the SCC type, although this effect is not statistically significant. In addition, we fitted the AG model with the number of previous events as an additional covariate in the model. The results are shown as AG2 model in Table 3. Interestingly but as expected, the treatment effect is completely wiped out for both types of the events. The effect is picked up by the number of previous events. For subject with more number of previous events, it is more likely to have events in the future. We also examined a non-parametric approach to test the treatment effect within each type, using the robust test for treatment comparison in Cook, Lawless, and Nadeau (1996). The p-values were 0.124 and 0.527 for the treatment effects for types BCC and SCC, respectively. These results are

Table 3: Summary of alternative model fittings to the skin cancer dataset. Here,  $\beta_1$  and  $\beta_2$  are the treatment effects for BCC and SCC, respectively. The parameters  $\gamma_1$  and  $\gamma_2$  show the effect of the previous number of events on BCC and SCC, respectively.

Model	Para.	EST.	S.E.	p-value	95 % CI		RR EST.	95 % CI	
					lower	upper		lower	upper
BCC									
Gaussian	$\beta_1$	0.073	0.068	0.280	-0.060	0.206	1.076	0.942	1.229
Gamma	$\beta_1$	0.102	0.070	0.147	-0.036	0.240	1.108	0.965	1.272
AG1	$\beta_1$	0.114	0.074	0.122	-0.031	0.259	1.121	0.970	1.295
AG2	$\beta_1$	0.019	0.048	0.689	-0.075	0.114	1.019	0.927	1.121
	$\gamma_1$	0.148	0.006	0.000	0.136	0.160	1.160	1.146	1.174
SCC									
Gaussian	$\beta_2$	0.075	0.110	0.498	-0.141	0.291	1.078	0.868	1.337
Gamma	$\beta_2$	0.079	0.126	0.533	-0.168	0.325	1.082	0.845	1.385
AG1	$\beta_2$	0.089	0.140	0.526	-0.186	0.363	1.093	0.830	1.438
AG2	$\beta_2$	-0.002	0.126	0.984	-0.250	0.245	0.998	0.779	1.277
	$\gamma_2$	0.170	0.022	0.000	0.126	0.214	1.185	1.135	1.238

similar to the AG1 model fitting.

In summary, the treatment effects for both types of skin cancers are not statistically significant, although the point estimates show increased rates in the treatment group. Through the multi-type modeling, we found that there is more heterogeneity in the event process of the SCC type of cancer and there is a mild to medium correlation between the occurrence of the two cancer types.

## 6 Concluding Remarks and Areas for Future Research

This paper provides a semiparametric methodology to characterize the incidence rate of event types, estimate the impact of covariates, and understand the correlation structure among event types. Maximum likelihood estimates of the regression coefficients, variance-covariance components, and nonparametric baseline intensity functions are obtained based on an MCEM algorithm. A Metropolis-Hastings sampling was used to draw from the conditional distribution of the random effects. The proposed model performs well for simulated datasets with different censoring rates and number of events per subject. The method is illustrated via its application to the skin cancer prevention dataset collected to study the effect of selenium supplementation on the risk of developing two types of tumors.

The model is developed for data with multiple covariates and a multivariate Gaussian distribution is used to model the correlated random effects. The algorithm can be modified

for other multivariate distributions desired for the random effects. The model can also be adopted easily for data with time-dependent covariates. Furthermore, generalizations of the model are not considered in this paper but can be made in a future work by including multilevel frailty, dependent censoring, and assessing the goodness of fit and the influence analysis. Robust variance estimators can also be considered in the setting of multi-type events (e.g, Al-Khalidi et al. 2011). Automated MCEM algorithms (e.g., Booth and Hobert 1999, and Levine and Fan 2004) could be adopted for the current model and study the efficacy of the different routines. Finally, using the copula function to model the multivariate correlated frailties is an area of interest for our future work.

## Acknowledgments

We would like to thank the editor, an associate editor, and two referees, for their valuable comments that helped in improving the paper significantly. We also would like to thank Dr. Mary Reid, Roswell Park Cancer Institute, for providing us with permission to use dataset from the Nutritional Prevention of Cancer clinical trial. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources. The work by Bedair was partially funded by Tanta University (Egypt). The work by Hong was partially supported by the National Science Foundation under Grant BCS-1122876 to Virginia Tech.

## A Derivatives

In the appendix, we provide the technical details for the statistical inference in Section 3.3. Let  $\theta_a$  and  $\theta_b$  be two arbitrary elements in  $\boldsymbol{\theta} = \text{vech}(\Sigma)$ . The first and second partial derivatives of the loglikelihood function in (3) with respect to  $\theta_a$ 's are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \theta_a} &= -\frac{m}{2} \text{tr} \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_a} \right) + \frac{1}{2} \mathbf{w}' \Sigma_{\mathbf{w}}^{-1} \frac{\partial \Sigma_{\mathbf{w}}}{\partial \theta_a} \Sigma_{\mathbf{w}}^{-1} \mathbf{w} \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_a \partial \theta_b} &= \frac{m}{2} \text{tr} \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_b} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_a} \right) - \frac{1}{2} \mathbf{w}' \left( \Sigma_{\mathbf{w}}^{-1} \frac{\partial \Sigma_{\mathbf{w}}}{\partial \theta_b} \Sigma_{\mathbf{w}}^{-1} \frac{\partial \Sigma_{\mathbf{w}}}{\partial \theta_a} \Sigma_{\mathbf{w}}^{-1} \right. \\ &\quad \left. + \Sigma_{\mathbf{w}}^{-1} \frac{\partial \Sigma_{\mathbf{w}}}{\partial \theta_a} \Sigma_{\mathbf{w}}^{-1} \frac{\partial \Sigma_{\mathbf{w}}}{\partial \theta_b} \Sigma_{\mathbf{w}}^{-1} \right) \mathbf{w}, \end{aligned}$$

respectively. Here,  $\mathcal{L} = \mathcal{L}(\boldsymbol{\xi}|\mathbf{t}, \mathbf{w})$ . The first and second partial derivatives of the loglikelihood function in (3) with respect to  $\boldsymbol{\beta}_j$ 's and  $\lambda_{jl}$ 's are

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}_j} &= \sum_{i=1}^m N_{ij}(\tau_i) \mathbf{x}_i - \sum_{i=1}^m \Lambda_{0j}(\tau_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_j + w_{ij}) \mathbf{x}_i \\ \frac{\partial \mathcal{L}}{\partial \lambda_{jl}} &= \frac{d_{jl}}{\lambda_{jl}} - \sum_{i \in R_{jl}} \exp(\mathbf{x}'_i \boldsymbol{\beta}_j + w_{ij}) \\ \frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}'_j} &= - \sum_{i=1}^m \Lambda_{0j}(\tau_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_j + w_{ij}) \mathbf{x}_i \mathbf{x}'_i \\ \frac{\partial^2 \mathcal{L}}{\partial \lambda_{jl}^2} &= - \frac{d_{jl}}{\lambda_{jl}^2} \\ \frac{\partial^2 \mathcal{L}}{\partial \lambda_{jl} \partial \boldsymbol{\beta}'_j} &= - \sum_{i \in R_{jl}} \exp(\mathbf{x}'_i \boldsymbol{\beta}_j + w_{ij}) \mathbf{x}'_i,\end{aligned}$$

respectively. All other partial second derivatives that are not shown above are zero.

## References

- Abu-Libdeh, H., B. W. Turnbull, and L. C. Clark (1990). Analysis of multi-type recurrent events in longitudinal studies; application to a skin cancer prevention trial. *Biometrics* 16(4), 1017–1034.
- Al-Khalidi, H. R., Y. Hong, T. R. Fleming, and T. M. Therneau (2011). Insights on the robust variance estimator under recurrent-events model. *Biometrics* 67, 1564–1572.
- Booth, J. G. and J. P. Hobert (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61(1), 265–285.
- Cai, J. and D. E. Schaubel (2004). Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Analysis* 10(2), 121–138.
- Chen, B. E. and R. J. Cook (2009). The analysis of multivariate recurrent events with partially missing event types. *Lifetime Data Analysis* 15(1), 41–58.
- Chen, B. E., R. J. Cook, J. F. Lawless, and M. Zhan (2005). Statistical methods for multivariate interval-censored recurrent events. *Statistics in Medicine* 24(5), 671–691.
- Clark, L. C., G. F. Combs, B. W. Turnbull, E. H. Slate, D. K. Chalker, J. Chow, L. S. Davis, R. A. Glover, G. F. Graham, E. G. Gross, et al. (1996). Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin: a random-

- ized controlled trial. *JAMA: the Journal of the American Medical Association* 276(24), 1957–1963.
- Cook, R. J. and J. F. Lawless (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer.
- Cook, R. J., J. F. Lawless, and K.-A. Lee (2010). A copula-based mixed poisson model for bivariate recurrent events under event-dependent censoring. *Statistics in Medicine* 29(6), 694–707.
- Cook, R. J., J. F. Lawless, and C. Nadeau (1996). Robust tests for treatment comparisons based on recurrent event responses. *Biometrics* 52, 557–571.
- Cook, R. J., E. Ng, J. Mukherjee, and D. Vaughan (1999). Two-state mixed renewal processes for chronic disease. *Statistics in Medicine* 18(2), 175–188.
- Duchateau, L. and P. Janssen (2008). *The Frailty Model*. New York: Springer.
- Duchateau, L., P. Janssen, I. Kezic, and C. Fortpied (2003). Evolution of recurrent asthma event rate over time in frailty models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 52(3), 355–363.
- Duffield-Lillico, A. J., E. H. Slate, M. E. Reid, B. W. Turnbull, P. A. Wilkins, G. F. Combs, H. K. Park, E. G. Gross, G. F. Graham, M. S. Stratton, et al. (2003). Selenium supplementation and secondary prevention of nonmelanoma skin cancer in a randomized trial. *Journal of the National Cancer Institute* 95(19), 1477–1481.
- Friedman, M. (1982). Piecewise exponential models for survival data with covariates. *The Annals of Statistics* 10, 101–113.
- Levine, R. A. and G. Casella (2001). Implementations of the Monte Carlo EM algorithm. *Journal of Computational and Graphical Statistics* 10(3), 422–439.
- Levine, R. A. and J. Fan (2004). An automated (Markov chain) Monte Carlo EM algorithm. *Journal of Statistical Computation and Simulation* 74(5), 349–360.
- Lin, D. Y., L. J. Wei, I. Yang, and Z. Ying (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B* 62, 711–730.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 22(2), 226–233.
- Manda, S. O. and R. Meyer (2005). Bayesian inference for recurrent events data using time-dependent frailty. *Statistics in Medicine* 24(8), 1263–1274.



- Mazroui, Y., S. Mathoulin-Pélessier, G. MacGrogan, V. Brouste, and V. Rondeau (2013). Multivariate frailty models for two types of recurrent events with a dependent terminal event: Application to breast cancer data. *Biometrical Journal* 55(6), 866–884.
- Mazroui, Y., A. Mauguen, S. Mathoulin-Pélessier, G. MacGrogan, V. Brouste, and V. Rondeau (2015). Time-varying coefficients in a multivariate frailty model: Application to breast cancer recurrences of several types and death. *Lifetime Data Analysis*, 1–25.
- McGilchrist, C. and K. Yau (2008). Survival analysis with time dependent frailty using a longitudinal model. *Australian Journal of Statistics* 38(1), 53–60.
- McLachlan, G. J. and T. Krishnan (2007). *The EM Algorithm and Extensions*, Volume 382. New Jersey: Wiley-Interscience.
- Moreno, E. S. S. (2008). *Nonparametric Frailty Models for Clustered Survival Data*. Ph. D. thesis, Cornell University.
- Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics* 26, 183–214.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ripatti, S., K. Larsen, and J. Palmgren (2002). Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm. *Lifetime Data Analysis* 8(4), 349–360.
- Robert, C. and G. Casella (2009). *Introducing Monte Carlo Methods with R*. New York: Springer.
- Rondeau, V., Y. Mazroui, and J. R. Gonzalez (2012). FRAILTYPACK: an R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software* 47.
- Therneau, T. M. and P. M. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.
- Vaida, F. and R. Xu (2000). Proportional hazards model with random effects. *Statistics in Medicine* 19(24), 3309–3324.
- Wei, G. C. and M. A. Tanner (1990). A Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association* 85(411), 699–704.