Statistical Methods for Multi-type Recurrent Event Data Based on Monte Carlo EM Algorithms and Copula Frailties

Khaled F. Bedair

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Yili Hong, Chair
Jeffrey B. Birch
Pang Du
Inyoung Kim
Eric P. Smith

September 22, 2014
Blacksburg, Virginia

Statistical Methods for Multi-type Recurrent Event Data Based on Monte Carlo EM Algorithms and Copula Frailties

Khaled F. Bedair

Abstract

In this dissertation, we are interested in studying processes which generate events repeatedly over the follow-up time of a given subject. Such processes are called recurrent event processes and the data they provide are referred to as recurrent event data. Examples include the cancer recurrences, recurrent infections or disease episodes, hospital readmissions, the filing of warranty claims, and insurance claims for policy holders. In particular, we focus on the multi-type recurrent event times which usually arise when two or more different kinds of events may occur repeatedly over a period of observation. Our main objectives are to describe features of each marginal process simultaneously and study the dependence among different types of events. We present applications to a real dataset collected from the Nutritional Prevention of Cancer Trial. The objective of the clinical trial was to evaluate the efficacy of Selenium in preventing the recurrence of several types of skin cancer among 1312 residents of the Eastern United States.

Four chapters are involved in this dissertation. Chapter 1 introduces a brief background to the statistical techniques used to develop the proposed methodology. We cover some concepts and useful functions related to survival data analysis and present a short introduction to frailty distributions. The Monte Carlo expectation maximization (MCEM) algorithm and copula functions for the multivariate variables are also presented in this chapter.

Chapter 2 develops a multi-type recurrent events model with multivariate Gaussian random effects (frailties) for the intensity functions. In this chapter, we present nonparametric baseline intensity functions and a multivariate Gaussian distribution for the multivariate correlated random effects. An MCEM algorithm with MCMC routines in the E-step is adopted for the partial likelihood to estimate model parameters. Equations for the variances of the estimates are derived and variances of estimates are computed by Louis' formula. Predictions of the individual random effects are obtained because in some applications the magnitude of the random effects is of interest for a better understanding and interpretation of the variability in the data. The performance of the proposed methodology is evaluated by simulation studies, and the developed model is applied to the skin cancer dataset.

Chapter 3 presents copula-based semiparametric multivariate frailty models for multi-type recurrent event data with applications to the skin cancer data. In this chapter, we generalize the multivariate Gaussian assumption of the frailty terms and allow the frailty distributions to have more features than the symmetric, unimodal properties of the Gaussian density. More flexible approaches to modeling the correlated frailty, referred to as copula functions, are introduced. Copula functions provide tremendous flexibility especially in allowing taking the advantages of a variety of choices for the marginal distributions and correlation structures. Semiparametric intensity models for multi-type recurrent events based on a combination of the MCEM with MCMC sampling methods and copula functions are introduced. The combination of the MCEM approach and copula function is flexible and is a generally applicable approach for obtaining inferences of the unknown parameters for high dimension frailty models. Estimation procedures for fixed effects, nonparametric baseline intensity functions, copula parameters, and predictions for the subject-specific multivariate frailties and random effects are obtained. Louis' formula for variance estimates are derived and calculated. We investigate the impact of the specification of the frailty and random effect models on the inference of covariate effects, cumulative baseline intensity functions, prediction of random effects and frailties, and the estimation of the variance-covariance components. Performances of proposed models are evaluated by simulation studies. Applications are illustrated through the dataset collected from the clinical trial of patients with skin cancer. Conclusions and some remarks for future work are presented in Chapter 4.

Key Words: MCEM algorithm; cancer studies; multi-type recurrent events; multivariate frailty; semiparametric model; random effects; copula; survival analysis.

I dedicate this little work to
the spirit of my dear father.

Acknowledgments

First and foremost, I would like to thank God for his blessings and help that He has provided me throughout my life.

I express my deepest gratitude to my advisor Dr. Yili Hong for his invaluable support, guidance, constant encouragement, and insight throughout this research work. Dr. Hong took handouts of rough drafts and gave back the insight, suggestions, and kind of encouragement that only a wonderful advisor can truly provide. This research work would not have been possible without his valuable guidance and advice.

I thank my committee members Drs. Jeffrey B. Birch, Pang Du, Inyoung Kim, and Eric P. Smith for their valuable advice and time. Especial thanks to Dr. Eric P. Smith for taking the time out of his very busy schedule as the department head and always open his office to me. I wish to thank Dr. Jeffrey B. Birch for his wonderful job as the director of the graduate program and his continuous support not only to me but also to all graduate students that touches on almost all aspects of our life as graduate students. I thank Roswell Park Cancer Institute who has generously provided the cancer data for use in this dissertation.

Furthermore, my gratitude is due to my country, Egypt, for awarding me a PhD scholarship and the Department of Statistics at Virginia Tech for financially supporting me. I thank the faculty members, colleagues, and staffs in the Department of Statistics at Virginia Tech and Tanta University. I thank my research group members: Dr. Yuanyuan Duan, Caleb King, Yimeng Xie, Zhibing Xu, and Miao Yuan. It has been a pleasure working with such a competent group.

My parents, there is no way to measure how much I have gratefully appreciated your counsel over all my life. I thank my parents-in-law, my brother (Mamdouh), my sisters, and their families. The list would be very too long if I listed all great people who deserve my thanks. But I owe them a lot and will continue let them know that.

Last but not least, my extra appreciation and thankful to my wife, Basma, and my lovely children: Shahd, Ahmed, and Amgad for their love, patience, and support.

Khaled Farag Bedair
Blacksburg, VA, USA
September 22, 2014

Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1 General Introduction

1.1 Background

1.1.1 Time to Event Data

Survival data is a term used for describing data that measure the time to a certain event. In survival analysis, the data might be time to death, time to the occurrence of a disease (or complication), time to an epileptic seizure, time it takes for a patient to respond to a therapy, or time from response until disease relapse (e.g., disease returns) (see e.g., Wienke 2010, Cook and Lawless 2007). The event is a transition from one state to another. For instance, death is a transition from state alive to state dead, occurrence of disease is a transition from the state of being healthy to a state of presence of disease and for epileptic seizure, an event is a transition from the seizure-free state to the state of active seizure.

Time to an event is a random variable having a continuous distribution. It is necessary to define the starting time point, say 0, from which times are measured. When we measure age, the starting time point may be date of birth. For the drug trial, the starting time is the time of start of treatment.

Up to this point, we have assumed that the event of interest can occur only once for a given subject. However, in many research scenarios in which the event of interest is not death, a subject may experience an event several times over a follow-up period. The provided data is known as recurrent event data. Examples of recurrent event data include: multiple episodes of relapses from remission comparing different treatments for leukemia patients, recurrent heart attacks of coronary patients being treated for heart disease, recurrence of bladder cancer tumors in a cohort of patients randomized to one of two treatment groups, and multiple events of deteriorating visual acuity in patients with baseline macular degeneration, where each recurrent event is considered a more clinically advanced stage of a previous event.

For each of the above examples, the event of interest differs, but may occur more than once per subject. A logical objective for such data is to assess the relationship of relevant predictors to the rate in which events are occurring, allowing for multiple events per subject. (see e.g., Cook and Lawless 2007).

1.1.2 Censoring Schemes

Time to event data arises in a number of applied fields, such as medicine, biology, public health, epidemiology, engineering, economics, and demography. In biomedical applications, the data are collected over a period of time and consequently the "time to event" may not be observed for all the individuals in the study population (sample). This results in what is called "censored" data. That is, the "time to event" for those individuals who have not experienced the event under study is censored (by the end of study). It is also common that the amount of follow-up for the individuals in a sample vary from subject to subject. A common feature of these data sets is that they contain either censored or truncated observations.

Censored data arises when an individual's life length is known to occur only in a certain period of time. Well-known censoring schemes are right censoring, where all that is known is that the individual is still alive at a given time; left censoring, when all that is known is that the individual has experienced the event of interest prior to the start of the study; or interval censoring, where the only information is that the event occurs within some interval. Truncation schemes are right truncation, where only individuals who have experienced the event by a specified time are included in the sample; left truncation, where only individuals who survive a sufficient time are included in the sample.

1.2 Preliminaries

Usually, the event time T is assumed to follow a continuous distribution. All functions of the event time distribution are defined over the interval $[0, \infty)$.

1.2.1 Survival Function

In survival analysis, one is more interested in the probability of an individual to survive beyond time t, which is given by the survival function and is defined as $S(t) = \Pr[T > t]$, where T is the time variable. Since an observation either fails or survives and one of these two mutually exclusive alternatives must occur, we have the cumulative distribution function (cdf) , denoted by $F(t)$, that can be obtained from the survival function as $F(t) = 1 - S(t)$. The survival function is the integral of the probability density function (pdf), denoted by $f(t)$, $S(t) = \int_t^\infty f(x)\,dx$. (see e.g., Klein and Moeschberger 2003).

1.2.2 Hazard (Intensity) Function

Another major concept in survival analysis is the intensity (hazard) function. This function is also called (depending on the field of application) mortality rate, incidence rate, mortality curve, failure rate, or force of mortality. The intensity function is defined by

$$\lambda(t) = \lim_{\triangle t \to 0} \frac{\Pr\left[t \leq T < t + \triangle t |\ T \geq t\right]}{\triangle t} = \frac{f(t)}{S(t)}.$$

The cumulative intensity function can be expressed as

$$\Lambda(t) = \int_0^t \lambda(x)dx = \int_0^t f(x)/S(x)dx = -\ln\left[S(t)\right],$$

and consequently, it is straightforward to present $S(t) = \exp\left[-\Lambda(t)\right]$.

1.2.3 Likelihood for Survival Data

For random right censoring, survival data consist of a combination of event times t_i and right censoring indicators δ_i which is 0 if the observation is censored and 1 if event time is observed. The contribution of right censored survival data (t_i, δ_i), $i = 1, ..., n$, to the likelihood function can be given by

$$L_i = \left[f(t_i)\right]^{\delta_i} \left[S(t_i)\right]^{1-\delta_i}. \tag{1.1}$$

For complete observations, the likelihood is their density, as is the case in standard situations. Censored observations provide the information that the unknown survival time exceeds the observed censored time. If we consider a sample of independent lifetimes $\{(t_1, \delta_1), \ldots, (t_n, \delta_n)\}$, equation (1.1) can easily be expressed in terms of the hazard function

$$L = \prod_{i=1}^{n} \left[f\left(t_i\right)\right]^{\delta_i} \left[S\left(t_i\right)\right]^{1-\delta_i} = \prod_{i=1}^{n} \left[\lambda\left(t_i\right)\right]^{\delta_i} \exp\left[-\Lambda(t_i)\right].$$

Klein and Moeschberger (2003) and Duchateau and Janssen (2008) discuss in more detail the likelihood for right censoring and other censoring schemes with parametric and semiparametric models as well as the partial likelihood approach.

1.2.4 Proportional Intensity (Hazard) Models

The proportional intensity model is the most popular model for survival data. In the presence of covariates, the proportional intensity model can be written as

$$\lambda_{ij}(t) = \lambda_0(t)\, \boldsymbol{w}_i \exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta}), \tag{1.2}$$

where $\lambda_{ij}(t)$ is the conditional intensity function (given the frailty terms $\boldsymbol{w}_i$) of the j^{th} event time for the i^{th} subject and $\lambda_0(t)$ is the baseline intensity function. The baseline intensity (hazard) function $\lambda_0(t)$ can either be assumed to have a particular parametric form (parametric models) or can be left unspecified (semiparametric models). $\boldsymbol{\beta}=(\beta_1, \ldots, \beta_p)'$ is a vector of p fixed effects associated with a vector of covariates $\boldsymbol{x}_{ij}$ for subject i and $\exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta})$ is the corresponding relative risk for subject i.

1.2.5 Models for Baseline Intensity

In parametric models the baseline hazard is specified by a small number of one-dimensional parameters. Throughout the literature, certain parametric models have been used repeatedly, such as exponential, Weibull, and Gompertz. These distributions have closed-form expres-

sions for survival, density, and hazard functions. Gamma and lognormal distributions are computationally less convenient but applied frequently.

To avoid model validity issues, alternatively, we can leave the form of the baseline hazard $\lambda_0(t)$ in (1.2) unspecified. The model with nonparametric baseline intensity function is referred to as a semiparametric model. This model contains one parametric vector $\boldsymbol{\beta}$, one factor that is not specified in a parametric way, $\lambda_0(t)$, and the variance-covariance components of the frailty term $\boldsymbol{w}_i$. The properties of different parametric, semiparametric and nonparametric models and methodology for statistical inference of event time data can be found in Duchateau and Janssen (2008), Aalen et al. (2007), and Klein and Moeschberger (2003).

1.2.6 Frailty Distributions

Frailty models are used in survival analysis to account for unobserved heterogeneity in individual risks of disease and death. Some important frailty distributions are used more often and they have several applications. The main frailty distributions proposed in the literature are gamma, positive stable, power variance function, lognormal, Weibull, and compound Poisson distributions. Hanagal (2011), Duchateau and Janssen (2008), and Wienke (2010) discuss various frailty models in more detail. In practice, gamma and lognormal distributions are used to model the frailty term and most frailty models limit the choice of the frailty distribution to these cases.

1.3 Monte Carlo EM Algorithm

The Expectation-Maximization (EM) algorithm is a broadly applicable approach to the iterative computation of maximum likelihood (ML) estimates (Dempster et al. 1977). The situations where the EM algorithm can be applied include not only incomplete-data situations, where there are missing data, truncated distributions, or censored or grouped observations, but also a variety of situations where the incompleteness of the data is not all that natural or evident. These include statistical models such as random effects, mixtures, log linear models,

and latent class and latent variable structures.

On each iteration of the EM algorithm, there are two steps the Expectation step or E-step and the Maximization step or M-step. The situations where the EM algorithm is applied are where ML estimation is made difficult by the absence of some part of the data. The algorithm starts in the M-step where the parameters are estimated after filling in initial values for the missing data. In the E-step, the initial values are then updated by their conditionally expected values using the current parameter estimates. The parameters are then re-estimated and so on, proceeding iteratively until convergence. The EM algorithm has found applications in almost all statistical contexts and in almost all fields where statistical techniques have been applied.

1.4 Copula Functions

Copulas are methods for modeling multivariate distributions. A copula models the dependence between the marginals in a multivariate distribution and it can be combined with any set of univariate distributions for the marginal distributions. Consequently, the use of copulas allows us to take advantages of the wide variety of univariate models that are available. A copula is a function that links univariate marginals to their full multivariate distribution. All marginals are uniform over $(0, 1)$. For a P-dimensional random vector $\mathbf{U} = (U_1, ..., U_P)'$ on the unitary cube, a copula C is

$$C(u_1, ..., u_P) = \Pr\left(U_1 \leq u_1, ..., U_P \leq u_P\right). \tag{1.3}$$

Here the function $C(\cdot)$ is called a copula function with properties such that $U_1, ..., U_P \sim$ uniform $(0, 1)$, $C(u_1, \ldots, 0, \ldots u_P) = 0$, and $0 \leq C(u_1, ..., u_P) \leq 1$.

Let $F(y_1, ..., y_P)$ be a P-dimensional distribution function for a random vector $\boldsymbol{y} = (y_1, ..., y_P)'$ with margins cdf, $F_1, ..., F_P$. Sklar (1954) first showed that there exists a P-

dimensional copula C such that

$$F(y_1, ..., y_P) = C\left[F_1(y_1), ..., F_P(y_P)\right].$$

The multivariate distribution function $F(y_1, ..., y_P)$ can be described by the margins $F_1, ..., F_P$ and the copula C shown in (1.3). The copula density can be defined as

$$c\left(u_1, ..., u_P\right) = \frac{\partial^P}{\partial u_1, ..., \partial u_P} C\left(u_1, ...u_P\right). \tag{1.4}$$

By differentiating (1.3), the density of $\boldsymbol{y}$ is equal to

$$f(y_1, ..., y_P) = c\left(u_1, ..., u_P\right) \prod_{d=1}^{P} f_d\left(y_d\right). \tag{1.5}$$

The result in (1.5) shows that it is always possible to specify a multivariate density by specifying the marginal densities $f_d\left(y_d\right)$, $(d = 1, \ldots, P)$, and a copula density $c\left(\cdot\right)$.

Once the marginal distributions have been specified, an appropriate copula can be selected. Because copulas separate marginal distributions from dependence structures, the appropriate copula for a particular application is the one which best captures dependence features of the data. A large number of copulas have been proposed in the literature, and each of these imposes a different dependence structure on the data. Nelsen (1999), Joe (1997), and Jaworski et al. (2010) provide a thorough coverage of copulas and their properties. We briefly discuss copulas that have appeared frequently in applications, and explain dependence structures of each copula. The most common copulas are the elliptical and Archimedean copulas, which are described as follows.

1.4.1 Elliptical Copulas.

Let $F\left(\cdot\right)$ be the multivariate cdf of an elliptical distribution which is any member of a broad family of probability distributions that generalize the multivariate normal distribution and inherit some of its properties. Let $F_d\left(\cdot\right)$ be cdf of the d^{th} margin and $F_d^{-1}(\cdot)$ be its inverse

function (quantile function), $d = 1, ..., P$. The elliptical copula determined by $F(\cdot)$ is

$$C(u_1, ..., u_P) = F[F_1^{-1}(u_1), ..., F_P^{-1}(u_P)].$$

Gaussian copula and student's t-copula are the most known elliptical copulas.

1.4.1.1 Gaussian Copula

The Gaussian copula with correlation matrix $\boldsymbol{R}$ is defined by

$$C(u_1, ..., u_P) = \phi_{\boldsymbol{R}}\left[\phi^{-1}(u_1), \ldots, \phi^{-1}(u_P)\right],$$

where $\phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution, and $\phi_{\boldsymbol{R}}$ is the joint cumulative distribution function of a standard multivariate normal with the correlation matrix $\boldsymbol{R}$. The density of the normal copula is given by

$$c(u_1, \ldots, u_P) = \frac{1}{|\boldsymbol{R}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\boldsymbol{\omega}'\left(\boldsymbol{R}^{-1} - \boldsymbol{I}_P\right)\boldsymbol{\omega}\right],$$

where $\boldsymbol{\omega} = \left[\phi^{-1}(u_1), \ldots, \phi^{-1}(u_P)\right]'$, $\boldsymbol{R}$ is the correlation matrix between $u_1, \ldots, u_P$, and $\boldsymbol{I}_P$ is an identity matrix with dimensions P. The Gaussian copula is flexible in that it allows for equal degrees of positive and negative dependence.

1.4.1.2 Student's t-copula

The Student t-copula with correlation coefficients matrix $\boldsymbol{R}$ and degrees of freedom δ can be written as

$$C(u_1, \ldots, u_P) = t_{\boldsymbol{R},\delta}\left[t_\delta^{-1}(u_1), \ldots, t_\delta^{-1}(u_P)\right],$$

where the $t_{\delta}^{-1}(\cdot)$ is the quantile function of the univariate t-distribution with δ degrees of freedom and $t_{\boldsymbol{R},\delta}$ is the joint cumulative distribution function of the multivariate Student t-distribution with $\boldsymbol{R}$ and δ.

1.4.2 Archmedian Copulas

The P-dimensional Archimedean copula is the copula of P- uniform $(0,1)$ random variables. The Archimedean copula has the form

$$C(u_1,\ldots,u_P)=\varphi^{-1}\left[\varphi(u_1)+\ldots+\varphi(u_P)\right],$$

where $\varphi(t)$ is called the copula generator, and the inverse of the generator $\varphi^{-1}(s)$. A summary of the three common families of Archimedean copulas, their generator, their parameter space, and an expression of Kendall's tau is given as follows.

1.4.2.1 Clayton Copula

The Clayton copula takes the form

$$C(u_1,\ldots,u_P)=\left(\sum_{d=1}^{P}u_d^{-\alpha}-P+1\right)^{-\frac{1}{\alpha}},\ \alpha\geq 0$$

where $\varphi(t)=t^{-\alpha}-1$ and $\varphi^{-1}(s)=(1+s)^{-\frac{1}{\alpha}}$. Kendall's tau is given by $\tau=\frac{\alpha}{\alpha+2}$. It is used to study correlated variables exhibit strong left tail dependence and relatively weak right tail dependence.

1.4.2.2 Gumbel Copula

The Gumbel copula with parameter α is given by

$$C(u_1,\ldots,u_P)=\exp\left\{-\left[\Sigma_{d=1}^{P}\left(-\ln u_d\right)^{\alpha}\right]^{\frac{1}{\alpha}}\right\},\ \alpha\geq 1$$

where $\varphi(t) = (-\ln t)^{\alpha}$ and $\varphi^{-1}(s) = \exp\left(-s^{\frac{1}{\alpha}}\right)$. Kendall's tau is given by $\tau = 1 - \alpha^{-1}$. Similar to the Clayton copula, Gumbel does not allow negative dependence, but it contrast to Clayton, Gumbel exhibits strong right tail dependence and relatively weak left tail dependence. If outcomes are known to be strongly correlated at high values but less correlated at low values, then the Gumbel copula is an appropriate choice.

1.4.2.3 Franklin Copula

Franklin's copula with parameter α is given by

$$C(u_1, \ldots, u_P) = \frac{-1}{\alpha} \ln \left\{ 1 + \frac{\prod_{d=1}^{P} \left[\exp(-\alpha u_d) - 1\right]}{\left[\exp(-\alpha) - 1\right]^{P-1}} \right\},$$

with$-\infty \leq \alpha \leq \infty$ for $P = 2$ and $0 \leq \alpha$ for $P \geq 2$. Where

$$\varphi(t) = -\ln \left[\frac{\exp(-\alpha t) - 1}{\exp(-\alpha) - 1} \right],$$

and

$$\varphi^{-1}(s) = -\frac{\ln\left\{1 + \exp(-s)\left[\exp(-\alpha) - 1\right]\right\}}{\alpha}.$$

The Frank copula is popular for several reasons. First, unlike some other copulas, it permits negative dependence between the marginals. Second, dependence is symmetric in both tails, similar to the Gaussian copula and Student t-copula.

1.4.2.4 Copula Estimation

By taking log of (1.5), we find that the log-likelihood is

$$\mathcal{L}(\boldsymbol{\alpha}) = \sum_{n=1}^{N} \log c\left[F_1(y_{n1}), \ldots, F_P(y_{nP})\right] + \sum_{n=1}^{N} \sum_{d=1}^{P} \log f_d(y_{nd}),$$

where N is the number of observations. The maximum likelihood method finds the value of the parameter that maximizes $\mathcal{L}(\boldsymbol{\alpha})$ over the entire set of parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_P, \alpha)'$.

Methods for parameter estimation are discussed in detail by Jaworski et al. (2010, ch3).

1.5 Motivation for the Dissertation Work

The single-type recurrent event times in the presence or absence of the dependent censoring have been considered in many studies. For example, Ripatti and Palmgren (2000) and Rondeau et al. (2007) proposed frailty models using penalized partial likelihood (PPL). The Monte Carlo expectation maximization (MCEM) was described as an estimation procedure to present the maximum likelihood by Vaida et al. (2000) and Ripatti et al. (2002) for independent multivariate random effects models. In the presence of informative censoring, the MCEM algorithm was introduced for a joint model of recurrent events and a terminal event (see e.g., Liu et al. 2004, Liu and Huang 2008, Huang and Liu 2007, Huang and Wolfe 2002).

For two types of recurrent events, Mazroui et al. (2013) proposed a multivariate frailty parametric model that analyzes two types of recurrent events with a dependent terminal event. Two estimation methods are described using penalized likelihood estimation where baseline hazard functions are approximated by M-splines, and another one with piecewise constant baseline hazard functions. Cook et al. (2010) presented a copula-based mixed Poisson model for bivariate recurrent events with gamma marginals and Clayton copula. The EM algorithm is proposed using numerical integration to evaluate the conditional distribution of the frailty and existing software for the generalized linear models and survival analysis when parametric and semiparametric models, respectively to estimate model parameters. The standard error of the estimates is obtained by a bootstrap method.

However, there are a limited number of studies on modeling the multi-type recurrent event times especially semiparametic models. Consequently, we are motivated to consider the analysis of this type of data. The main interest is to summarize features of conditional intensity functions and associated fixed effects as well as study the association between different types of events. We introduce a partial likelihood approach based on semiparametric models for the multi-type recurrent events where parameter estimation is obtained by adopting a Monte

Carlo EM algorithm. The methodologies that we develop, however, are general and can be straightforward applied to other parametric models as well as multi-type recurrent event times with or without terminal events. The proposed methods are illustrated by the analysis of recurrences for different skin cancer types.

1.5.1 Multivariate Gaussian Random Effects Model

We present a semiparametric model for the multi-type recurrent events based on an MCEM algorithm and multivariate Gaussian distribution for the correlated random effects. The MCMC algorithm is adopted to evaluate the conditional expectation of the correlated random effects in the E-step. The Partial likelihood approach is applied to obtain the estimates of the fixed effects and the non-parametric intensity functions in addition variance-covariance components are obtained in the M-step. Forms of Louis' method for variance of estimates are derived and calculated.

1.5.2 Copula Multivariate Frailty Models

In the multivariate copula frailty models, we relax the assumption that the frailty terms have the Gaussian distribution. Moreover, allowing the frailty distribution to have more complex features than the symmetric, unimodal normal density may provide more understanding into the underlying heterogeneity and even suggest including important covariates in the model. We consider another approach which is a class of multivariate probability models for the multivariate frailty known as copula models have a number of useful features. Combinations of any univariate marginal distributions that are not necessary to be from the same distributional family are available. Elliptical copula have the property that they increase in complexity at a much slower rate than existing multivariate probability distribution as the number of dimensions increases. They are involving a number of existing multivariate models, and providing a framework for generating many more. These advantages make copula models a general option in empirical analysis and an alternative to existing multivariate probability models

used in modeling the multivariate frailty. We present semiparametric models based Clayton and Gaussian copulas with gamma marginals to model the multivariate frailty. MCEM and Louis' methods are adapted to estimate unknown parameters and their variances respectively for the proposed models.

1.6 Overview

The rest of this dissertation is organized as follows. In Chapter 2, a statistical methodology is proposed to model the multi-type recurrent event times with nonparametric baseline intensity based on an MCEM algorithm and multivariate Gaussian random effects. Chapter 2 is mainly based on Bedair and Hong (2014a). Chapter 3 introduces copula functions to model the correlated multivariate frailty. An MCEM algorithm is adapted to estimate the unknown parameters for the proposed models. The variances of the estimates and predictions of frailties are obtained. Chapter 3 is based on Bedair and Hong (2014b). Chapter 4 presents conclusions and areas for future work.

Bibliography

Aalen, O. O., Bjertness, E. and Sønju, T.: 2007, Analysis of dependent survival data applied to lifetimes of amalgam fillings, *Statistics in Medicine* **14**(16), 1819–1829.

Bedair, K. and Hong, Y.: 2014a, A multivariate frailty model for the multi-type recurrent event data using a Monte Carlo EM algorithms. To be submitted.

Bedair, K. and Hong, Y.: 2014b, Copula based semiparametric multivariate frailty models for multi-type recurrent event data: Application on cancer data. To be submitted.

Cook, R. J. and Lawless, J. F.: 2007, *The Statistical Analysis of Recurrent Events*, Springer, New York.

Cook, R. J., Lawless, J. F. and Lee, K.-A.: 2010, A copula-based mixed poisson model for bivariate recurrent events under event-dependent censoring, *Statistics in Medicine* **29**(6), 694–707.

Dempster, A. P., Laird, N. M. and Rubin, D. B.: 1977, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)* **39**(1), 1–38.

Duchateau, L. and Janssen, P.: 2008, *The Frailty Model*, Springer, New York.

Hanagal, D. D.: 2011, *Modeling Survival Data using Frailty Models*, Chapman & Hall/CRC, New York.

Huang, X. and Liu, L.: 2007, A joint frailty model for survival and gap times between recurrent events, *Biometrics* **63**(2), 389–397.

Huang, X. and Wolfe, R. A.: 2002, A frailty model for informative censoring, *Biometrics* **58**(3), 510–520.

Jaworski, P., Durante, F., Härdle, W. and Rychlik, T.: 2010, *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*, Springer, New York.

Joe, H.: 1997, *Multivariate Models and Dependence Concepts*, CRC Press, New York.

Klein, J. P. and Moeschberger, M. L.: 2003, *Survival Analysis*, 2nd edn, Springer, New York.

Liu, L. and Huang, X.: 2008, The use of Gaussian quadrature for estimation in frailty proportional hazards models, *Statistics in medicine* **27**(14), 2665–2683.

Liu, L., Wolfe, R. A. and Huang, X.: 2004, Shared frailty models for recurrent events and a terminal event, *Biometrics* **60**(3), 747–756.

Mazroui, Y., Mathoulin-Pélissier, S., MacGrogan, G., Brouste, V. and Rondeau, V.: 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.

Nelsen, R. B.: 1999, *An Introduction to Copulas*, Springer, New York.

Ripatti, S., Larsen, K. and Palmgren, J.: 2002, Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm, *Lifetime Data Analysis* **8**(4), 349–360.

Ripatti, S. and Palmgren, J.: 2000, Estimation of multivariate frailty models using penalized partial likelihood, *Biometrics* **56**(4), 1016–1022.

Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V. and Soubeyran, P.: 2007, Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events, *Biostatistics* **8**(4), 708–721.

Sklar, A.: 1954, Fonctions de répartition à n dimensions et leurs marges, *Publications de l'Institut de Statistique de l'Université de Paris* **8**, 229–231.

Vaida, F., Xu, R. et al.: 2000, Proportional hazards model with random effects, *Statistics in Medicine* **19**(24), 3309–3324.

Wienke, A.: 2010, *Frailty Models in Survival Analysis*, CRC Press, New York.

Chapter 2 A Multivariate Frailty Model for the Multi-type Recurrent Event Data using a Monte Carlo EM Algorithm

Abstract

There has been an increasing interest in analyzing the multi-type recurrent event data. This type of data arises in many situations when two or more different event types may occur repeatedly over an observation period. For example, subjects are at risk of different disease types, causes of production stoppages, financial transactions in commerce, and insurance claims filed by holders. The interest in this setting is to characterize the incidence rate of event types, estimate the impact of covariates, and understand the correlation structure among event types. We propose a multivariate frailty proportional intensity model with multivariate distributions for the random effects to model the data. The dependence among event types is taken into account as well as the effect of covariates. Maximum likelihood estimates of the regression coefficients, variance-covariance components, and the nonparametric baseline intensity function are obtained based on a Monte Carlo Expectation Maximization (MCEM) algorithm. The E-step of the algorithm involves the calculation of the conditional expectations of the random effects by using the Metropolis-Hastings sampling. Louis' formula is applied to obtain the variance of the estimator. Simulation studies are presented to illustrate the performance of the proposed method. An application is described to a randomized controlled clinical trial for the efficacy of nutritional supplements for the prevention of two types of skin cancer.

Key Words: MCEM algorithm; multi-type; recurrent events; multivariate frailty; semi-parametric model; random effects; survival analysis.

2.1 Introduction

2.1.1 Background

There is, in many research scenarios, a widely interest in studying processes which generate events repeatedly over the follow-up time for a given subject. Such processes are called to

as recurrent event processes and the data they provide are referred recurrent event data (Cook and Lawless 2007). These types of processes arise frequently in medical studies, where information is often available on many individuals, each of whom may experience transient clinical events repeatedly over a period of observation. Examples include the occurrence of heart attacks of coronary patients being treated for heart attack, epileptic seizures in neurology studies, fractures in osteoporosis studies, and recurrence of bladder cancer tumors. In business, examples include the filing of warranty claims on automobiles, or insurance claims for policy holders.

Multi-type recurrent event data arise when two or more different kinds of events may occur repeatedly over a period of observation. The scientific objectives in such settings are often to describe features of the marginal process and to study the association between the different types of events. For example, in a study of infections following bone marrow transplantation, it is of interest to study different types of recurrent infections simultaneously (e.g., bacterial, fungal and viral infections). Similarly, in a clinical trial regarding the efficacy of nutritional supplement of Selenium in relationship to preventing skin cancer (Abu-Libdeh et al. 1990), it is of interest to study the recurrence of several types of skin cancer.

2.1.2 Related Literature

One important field for application of frailty models is the analysis of recurrent event time data. Reviews of models for recurrence data appeared recently in Cook and Lawless (2007) and Finkelstein (2008). As introduced by McGilchrist and Aisbett (1991), a subject-specific frailty term can be used in the hazard function. For all observations from one subject, the same frailty term is used. The frailty variables of different subjects are independent realizations of a common frailty distribution. Parametric and nonparametric models, both with and without frailty, were applied in Duchateau et al. (2003) to recurrent asthma events from an asthma prevention trial in young children. Frailty models specially designed for recurrence data are considered in detail in McGilchrist and Yau (2008), Aalen et al. (2007), Manda and Meyer

(2005), and Yau and McGilchrist (1998). Reviews in the field are given Ezell et al. (2003), Lim et al. (2007), and Wienke (2010).

For single-type recurrent events, estimation of multivariate frailty models using penalized partial likelihood (PPL) was given in Ripatti and Palmgren (2000). A Monte Carlo Expectation Maximization (MCEM) to estimate a maximum likelihood estimation procedure has been presented by Vaida et al. (2000). Ripatti et al. (2002) has introduced an automated MCEM algorithm for multivariate frailty models.

Comparing with the single-type, methods for the analysis of the multi-type recurrent events have limited interest. The analysis of this kind of processes should consider the association between the processes occur for the same subject. Models with both the hazard function and frailty distribution described parametrically have been considered in. Abu-Libdeh et al. (1990) and Cook et al. (1999). In their models, the inference is based on the corresponding marginal likelihood.

A semiparametric model for multivariate survival data and recurrent events processes has been considered in Ng and Cook (1999). They take a marginal perspective and concentrate on the inference of the univariate rate and mean functions rather than the multivariate cumulative intensity functions. Their model requires the accurate specification for the covariance structure of the bivariate processes. Cai and Schaubel (2004) has considered a model for the clustered recurrent processes. They use the independence assumption between the processes and obtain a robust sandwich covariance estimates that allow for joint inferences.

A nonparametric bivariate model for the analysis of recurrent events with piecewise parametric baseline intensity function has been developed by Moreno (2008). The model employes the constant piecewise baseline intensity functions and the first two moments of the frailty distribution. The total parameter dimension has a significant effect on computing cost also, result in a numerical instability Moreno (2008, page 83).

For the multi-type interval-censored recurrent events, a method based on a mixed Poisson model with lognormal random effects and piecewise constant parametric baseline intensity

functions is reported by Chen et al. (2004). A multivariate recurrent event model for partially missing event types with piecewise constant baseline intensity functions is introduced by Chen and Cook (2009). Parameter estimation is obtained from an MCEM algorithm. They use the existing software of the Poisson regression for estimation with piecewise constant models in the M-step.

A bivariate mixed Poisson model using a copula function to describe the correlation between the two gamma distributed frailties has been described by Cook et al. (2010). Their developed model depends on the EM algorithm. In the M-step and by using the two gamma distributions for marginal frailties they use the existing software for the generalized linear models with the parametric baseline and survival analysis software for the semiparametric analysis. The numerical integration has been used in the E-step to calculate the conditional expectations, which may be limited when the number of event types is large. They apply the non-parametric bootstrap to obtain the stranded errors of the parameter estimates.

2.1.3 Overview

In this chapter, we generalize the independent frailty models described for the single-type event (Vaida et al. 2000, Ripatti et al. 2002) and develop a semiparamtric correlated frailty model for the multi-type event data. A multivariate Gaussian distribution is used to model the correlated random effects. The baseline intensity functions are modeled in a complete non-parametric way (different from piecewise models).

We propose an implementation of the MCEM algorithm to estimate the model parameters. In the E-step; a Metropolis-Hastings sampler (Hoff 2009) is used to obtain the expectation of the conditional distribution for the random effects. Different from the numerical integration, the Monte Carlo sampling is more flexible in dealing with multi-type events, even when the number of events is not small. In the M-step; maximum likelihood estimates of the regression coefficients, the variance components and a modified Nelson-Aalen estimator (Aalen 1978) of the baseline cumulative intensity functions are obtained. The observed information matrix

for the estimated parameters is derived and calculated by using Louis' formula (Louis 1982).

The remainder of this chapter is organized as follows. In Section 2.2 we describe the form of data to build the proposed model and introduce notations. At the same time, we define the intensity functions and set up the multivariate frailty model for the multi-type recurrent event data as well as a multivariate distribution for the random effects. In Section 2.3 the parameter estimation and the MCEM algorithm are described, and an estimator for the corresponding variance-covariance matrix is provided. Simulation studies and their results are reported in Section 2.4. The proposed model is applied to the skin cancer dataset in Section 2.5. Concluding remarks and topics for further research are given in Section 2.6.

2.2 Data and Model

2.2.1 Data

The k^{th} event time for subject i of type j is denoted by t_{ijk}, $i = 1, \ldots, m$, $j = 1, \ldots, J$, and $k = 1, \ldots, N_{ij}(\tau_i)$. Here m is the number of subjects under the study, J is the total number of event types, τ_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 occurring over the interval $[0, t]$for individual i. The censoring indicator δ_{ijk} equals to j if event type j is observed for subject i at time t_{ijk}, and $\delta_{ijk} = 0$ otherwise. We also have covariate information for subject i, denoted by $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$, where p is the number of covariates. Let $\mathbf{t}_i = \left(t_{i1}, \ldots, t_{iN_i(\tau_i)}\right)'$, and the corresponding censor indicator vector $\boldsymbol{\delta}_i = \left(\delta_{i1}, \ldots, \delta_{iN_i(\tau_i)}\right)'$. Here $N_i(t)$ is the total number of all event types observed for subject i over the interval $[0, t]$.

2.2.2 Notation and Intensity Function

The full vector of counting processes is denoted by $\mathbf{N}_i(t) = [N_{i1}(t), \ldots, N_{iJ}(t)]'$. We express $\Delta N_{ij}(t) = N_{ij}(t + \Delta t^-) - N_{ij}(t^-)$ to be the number of events occurring over the interval $[t,\, t+\Delta t]$. Let $dN_{ij}(t) = N_{ij}(t) - N_{ij}(t^-)$ indicate whether a type j event occurred for subject

i at time $t \geq 0$, and the full vector is $d\mathbf{N}_i(t) = [dN_{i1}(t), \ldots, dN_{iJ}(t)]'$. Let $[0, t]$ be the period of observation for subject i and let $\mathbf{Y}_i(t) = I(t \leq \tau_i)$. The event history for subject i is $\mathbf{H}_i(t) = \{\boldsymbol{x}_i, \mathbf{Y}_i(t), \mathbf{N}_i(s) : 0 \leq s < t\}$ and the conditional intensity function for the type j events of the form (Cook and Lawless 2007),

$$\lambda_j [t|\mathbf{H}_i(t)] = \lim_{\Delta t \downarrow 0} \frac{\Pr [\Delta N_{ij}(t) = 1|\mathbf{H}_i(t)]}{\Delta t}.$$

2.2.3 Model for Event Intensities

The proposed intensity model can be expressed as

$$\lambda_j(t_{ijk}) = \lambda_{0j}(t_{ijk}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij}), \tag{2.1}$$

where $i = 1, \ldots, m$, and $j = 1, \ldots, J$. Here, $\lambda_j(t_{ijk})$ is the conditional intensity function for the j^{th} event type of the i^{th} subject (conditional on w_{ij} and $\boldsymbol{x}_i$), observed times t_{ijk}, and the baseline intensity function for event type j is $\lambda_{0j}(t_{ijk})$. The vector of covariates of the i^{th} subject is $\boldsymbol{x}_i$, $\boldsymbol{\beta}_j$ is the fixed effect vector for the type j of dimension $p_j \times 1$, and w_{ij} is the random effect for the j^{th} type of the i^{th} subject. The cumulative intensity function can be written as

$$\Lambda_j(t) = \int_0^t \lambda_j(s)\, ds = \Lambda_{0j}(t) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij}), \tag{2.2}$$

where $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(s)\, ds$. The intensity function (2.1) can be rewritten as

$$\lambda_j(t_{ijk}) = \lambda_{0j}(t_{ijk})\, u_{ij} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j), \tag{2.3}$$

where $u_{ij} = \exp(w_{ij})$ is called the frailty for the j^{th} type of the i^{th} subject. Both models (2.1) and (2.3) are conditional intensity models (given w_{ij} and u_{ij}).

2.2.4 Model for Random Effects

When modeling the multivariate random effects the most commonly used distribution is the multivariate normal distribution. In this chapter, the random effects are distributed with a multivariate normal distribution. Let $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{iJ})'$ be the multivariate random effects, which are independent and identically distributed (i.i.d.) and follow multivariate normal distribution MVN $(\mathbf{0}, \Sigma)$. The variance-covariance matrix Σ for the i^{th} subject is

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1J} \\ \vdots & \ddots & \vdots \\ \sigma_{J1} & \cdots & \sigma_J^2 \end{bmatrix},$$

whereas the variances are σ_j^2 and the covariances are $\sigma_{jj\prime}$, $j \neq j'$. Let $\boldsymbol{w}$=$(\boldsymbol{w}_1', \ldots, \boldsymbol{w}_m')'$ to be a vector of random effects with the dimensions $Jm \times 1$. The multivariate normal distribution for $\boldsymbol{w}$ is MVN $(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}})$, where $\mathbf{0}$ is a vector of all elements equal to zero with the dimension $Jm \times 1$, and the variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{w}}$ can be defined as a black-diagonal matrix with all elements in the diagonal equal to Σ and with dimension $Jm \times Jm$. We write

$$\boldsymbol{\Sigma}_{\boldsymbol{w}} = \begin{bmatrix} \Sigma & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \Sigma \end{bmatrix}.$$

Here $\mathbf{0}$ is a zero matrix with dimension $J \times J$.

2.3 Parameter Estimation

2.3.1 Monte Carlo Expectation Maximization (MCEM) algorithm

The EM algorithm provides a tool for obtaining maximum likelihood estimates under models that yield analytically formidable likelihood equations. The EM algorithm is an iterative

routine requiring two primarily calculations at each iteration: computation of a particular conditional expectation of the log-likelihood (E-step) and M-step to maximize this expectation over the corresponding parameters. 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 and Metropolis-Hastings samplers (Robert and Casella 2009, McLachlan and Krishnan 2007). A key step is to treat the random effects $\boldsymbol{w}$ as missing data.

2.3.1.1 M-step

We use the complete data for the subject i including the event times $\boldsymbol{t_i}$ and types $\boldsymbol{\delta_i}$, the covariate vector $\boldsymbol{x_i}$ and random effects $\boldsymbol{w_i}$. Here we define $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_J')'$, $\boldsymbol{\theta}$ to be the parameter vector for the distribution of the random effects, and $\boldsymbol{\lambda}_{0j}(\cdot)$ to be the unspecified baseline hazard functions with $\boldsymbol{\lambda}_0(\cdot) = [\boldsymbol{\lambda}_{01}'(\cdot), \ldots, \boldsymbol{\lambda}_{0J}'(\cdot)]'$. We have the ordered distinct event times denoted by $\boldsymbol{t}_j = \left[t_{j(1)}, \ldots, t_{j(L_j)}\right]'$ and $N\left(t_{j(l_j)}\right)$ is the number of event type j for all subjects at the ordered distinct time $t_{j(l_j)}$, $l_j = 1, \ldots, L_j$, where L_j is the number of the distinct event times for event type j. The corresponding baseline intensity functions are $\boldsymbol{\lambda}_{0j}(\cdot) = \left[\lambda_{0j}(t_{j(1)}), \ldots, \lambda_{0j}(t_{j(L_j)})\right]'$. The vector of unknown parameters included in the model is $\boldsymbol{\xi} = [\boldsymbol{\beta}', \boldsymbol{\lambda}_0'(\cdot), \boldsymbol{\theta}']'$.

The likelihood function for the i^{th} subject based on the full data is

$$L_i(\boldsymbol{\xi}) = \prod_{j=1}^{J} \prod_{k=1}^{N_{ij}(\tau_i)} \left[\lambda_j(t_{ijk})\right]^{\delta_{ijk}} \exp\left[-\Lambda_j(\tau_i)\right] f(\boldsymbol{w}_i|\boldsymbol{\theta}), \tag{2.4}$$

where $f(\boldsymbol{w}_i|\boldsymbol{\theta})$ is the distribution of $\boldsymbol{w}_i$ depending on $\boldsymbol{\theta}$. Here $\lambda_j(t_{ik})$ and $\Lambda_j(t)$ are defined in (2.3) and (2.2), respectively. The log-likelihood for subject i^{th} is $\mathcal{L}_i(\boldsymbol{\xi}) = \log\left[L_i(\boldsymbol{\xi})\right]$. In particular,

$$\mathcal{L}_i\left(\boldsymbol{\xi}\right) = \sum_{j=1}^{J}\left\{\sum_{k=1}^{N_{ij}(\tau_i)}\delta_{ijk}\left[\log\,\lambda_{0j}(t_{ijk}) + (\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij})\right] - \Lambda_{0j}(\tau_i)u_{ij}\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\right\}$$
$$- \frac{J}{2}\log 2\pi - \frac{1}{2}\log(|\Sigma|) - \frac{1}{2}\left(\boldsymbol{w}_i'\Sigma^{-1}\boldsymbol{w}_i\right). \quad (2.5)$$

The corresponding log-likelihood is written as

$$\mathcal{L}_{full}\left(\boldsymbol{\xi}\right) = \mathcal{L}_{full_1}(\boldsymbol{\lambda}_0(\cdot), \boldsymbol{\beta}) + \mathcal{L}_{full_2}(\boldsymbol{\theta}), \quad (2.6)$$

where

$$\mathcal{L}_{full_1} = \sum_{i=1}^{m}\sum_{j=1}^{j}\left\{\sum_{k=1}^{N_{ij}(\tau_i)}\delta_{ijk}\log\left[\lambda_{0j}\left(t_{ijk}\right)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij})\right] - \Lambda_{0j}(\tau_i)u_{ij}\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\right\}, \quad (2.7)$$

and the log-likelihood for $\boldsymbol{\theta}$ conditional on random effect $\boldsymbol{w}$ is

$$\mathcal{L}_{full_2} = -\frac{Jm}{2}\log 2\pi - \frac{m}{2}\log(|\Sigma|) - \frac{1}{2}\left(\boldsymbol{w}'\Sigma_{\boldsymbol{w}}^{-1}\boldsymbol{w}\right). \quad (2.8)$$

The regression coefficient $\boldsymbol{\beta}$ can be estimated by profiling $\mathcal{L}_{full_1}$ to the partial log-likelihood (Johansen 1983) and replacing w_{ij} and u_{ij} by the expected values at the current iteration in the E-step of the algorithm. At each iteration of the MCEM, we obtain S' samples after burn-in and thinning from the conditional distribution of the random effects. The conditional expectations $\mathrm{E}(w_{ij})$ and $\mathrm{E}(u_{ij})$ can be estimated by

$$\mathrm{E}(w_{ij}) = \frac{1}{S'}\sum_{s=1}^{S'} w_{ij}, \text{ and } \mathrm{E}(u_{ij}) = \frac{1}{S'}\sum_{s=1}^{S'} u_{ij}.$$

We have

$$\mathrm{E}\left[\mathcal{L}_{partial}(\boldsymbol{\beta})\right] = \sum_{j=1}^{J}\sum_{l_j=1}^{L_j}\left[\boldsymbol{x}_i'\boldsymbol{\beta}_j + \mathrm{E}\left(w_{ij}\right) - \log\sum_{i\in R_{j\left(l_j\right)}}\mathrm{E}\left(u_{ij}\right)\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j\right)\right], \tag{2.9}$$

where $R_{j(l_j)}$ is the risk group for the event type j at the specific event time $t_{j(l_j)}$. Estimates of fixed effect parameters $\boldsymbol{\beta}$, can thus be obtained by maximizing $\mathrm{E}\left[\mathcal{L}_{partial}(\boldsymbol{\beta})\right]$.

For estimating the baseline and cumulative intensity, we use the profile likelihood approach at which we fix $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ in (2.7). The Nelson-Aalen estimator of $\Lambda_{0j}\left[t_{j(l_j)}\right]$ which maximizes the profile likelihood, can be obtained (Klein and Moeschberger 2003, p. 258). In particular,

$$L_{\boldsymbol{\beta}_j}\left[\lambda_{0j}(t_{j(l_j)})\right] = \left[\prod_{l_j=1}^{L_j}\lambda_{0j}(t_{l_j})\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij})\right]\exp\left[-\sum_{i=1}^{m}\Lambda_{0j}\left(\tau_i\right)\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij}\right)\right]. \tag{2.10}$$

Equation (2.10) can be written as

$$L_{\boldsymbol{\beta}_j}\left[\boldsymbol{\lambda}_{0j}(\cdot)\right] \propto \left[\prod_{l_j=1}^{L_j}\lambda_{0j}(t_{l_j})\right]\exp\left[-\lambda_{0j}(t_{l_j})\sum_{i\in R_{j\left(l_j\right)}}\mathrm{E}\left(u_{ij}\right)\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j\right)\right].$$

An estimate of the $\Lambda_{0j}(t)$ is given by

$$\Lambda_{0j}(t) = \sum_{t_{(l_j)}\leq t}\lambda_{0j}(t_{(l_j)}),$$

where

$$\lambda_{0j}(t_{l_j}) = \frac{N_{t_{j\left(l_j\right)}}}{\sum_{i\in R_{t_{(l_j)}}}\mathrm{E}(u_{ij})\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)},$$

and $N_{t_{j\left(l_j\right)}}$ is the number of events of type j at the specific ordered tine $t_{j(l_j)}$. By maximizing $\mathrm{E}(\mathcal{L}_{full2})$, the maximum likelihood estimate for the variance-covariance matrix can be given

in the form

$$\hat{\Sigma} \;=\; \frac{1}{m}\sum_{i=1}^{m}\mathrm{E}(\boldsymbol{w}_i\boldsymbol{w}_i') = \frac{1}{S'm}\sum_{s=1}^{S'}\sum_{i=1}^{m}\boldsymbol{w}_i\boldsymbol{w}_i', \tag{2.11}$$

where S' the number of samples after burn-in and thinning in the E-step and m is the number of subject.

2.3.1.2 E-step

In MCEM, the E-step is preceded by the M-step, where non-observed frailty terms are imputed from the conditional distribution of the random effects given the observed data and the parameter estimates obtained from the previous iteration. The estimates for the parameters at s^{th} iteration of the EM algorithm are represented by $\widehat{\boldsymbol{\xi}}^{(s)}$. Samples are drawn from the conditional distribution $g(\boldsymbol{w}_i|\boldsymbol{\xi})$ which can be shown using the MCMC approach as follows

$$g(\boldsymbol{w}_i|\boldsymbol{\xi}) = \frac{f(\boldsymbol{t},\boldsymbol{w}_i)}{\int_{-\infty}^{\infty} f(\boldsymbol{t},\boldsymbol{w}_i)d\boldsymbol{w}_i} = \frac{L_i\left(\boldsymbol{\xi}\right)}{L_{marg,i}(\boldsymbol{\xi})}$$

where $f(\boldsymbol{t},\boldsymbol{w}_i)$ is the joint density of the random effects and complete data, $L_i\left(\boldsymbol{\xi}\right)$ is defined in (2.4), and $L_{marg,i}(\boldsymbol{\xi}) = \int_{-\infty}^{\infty} L_i(\boldsymbol{\xi}|\boldsymbol{w}_i)\, d\boldsymbol{w}_i$ which can be dropped from the expression of the conditional density since it does not depend on $\boldsymbol{w}_i$. The conditional distribution for the subject random effect $\boldsymbol{w}_i$ is

$$g(\boldsymbol{w}_i|\boldsymbol{\xi}) \propto L_i\left(\boldsymbol{\xi}\right). \tag{2.12}$$

It is shown (see Appendix 2.A.1) that the conditional joint distribution for $\boldsymbol{w}$ is proportional to

$$g(\boldsymbol{w}|\boldsymbol{\xi}) \propto \exp\left[\sum_{i=1}^{m}\sum_{j=1}^{J} N_{ij}(\tau_i)w_{ij} - \Lambda_{0j}(\tau_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\exp(w_{ij}) - \frac{1}{2}\left(\boldsymbol{w}_i'\Sigma^{-1}\boldsymbol{w}_i\right)\right]. \tag{2.13}$$

The required conditional expectations $\mathrm{E}(w_{ij})$, $\mathrm{E}(u_{ij})$, $\mathrm{E}(w_{ij}w_{ij'})$, and $\mathrm{E}(w_{ij}^2)$ in (2.9) and

Algorithm: 2.1 MCEM algorithm steps with Metropolis-Hastings in E-step.

1. Choose starting values $\hat{\boldsymbol{\xi}}^{(0)}$.
2. (E-step), Using the current parameter values, generate S samples $\left(\boldsymbol{w}^{(1)'}, \cdots, \boldsymbol{w}^{(S)'}\right)'$ from the conditional distribution of $g(\boldsymbol{w}|\boldsymbol{\xi})$ using a Metropolis-Hastings algorithm like the one described in Appendix 2.A.2 and obtain Monte Carlo approximations to the required expectations $\mathrm{E}(w_{ij})$, $\mathrm{E}(u_{ij})$, $\mathrm{E}(w_{ij}w_{ij'})$, and $\mathrm{E}(w_{ij}^2)$.
3. (M-step), Given the conditional expectation values in Step 2, then
   (a) Find estimates for $\boldsymbol{\beta}$ that maximize $\mathrm{E}\left[\mathcal{L}_{partial}(\boldsymbol{\beta})\right]$.
   (b) Given $\widehat{\boldsymbol{\beta}}$ in step (3.a), maximize the expected profile likelihood to estimate the baseline cumulative intensity $\Lambda_{0j}(t)$.
   (c) Maximize $\mathrm{E}(\mathcal{L}_{full(2)})$ to find estimates for $\boldsymbol{\theta}$.
4. If convergence is achieved, declare the current values to be the maximum likelihood parameter estimates; otherwise let $s = s + 1$ and return to Step 2.

(2.11) are not available in closed form. 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 et al. (2010). This solution is feasible, however, only for low-dimensional random vectors $\boldsymbol{w}_i$ (Vaida et al. 2000).

We propose computing for the E-step expectations based on a simulated sample from $g(\boldsymbol{w}_i|\boldsymbol{\xi})$. There are many ways to perform the sampling including a Gibbs sampler (McCulloch 1994), Metropolis-Hastings algorithm (McCulloch 1997) and independent samples from either the distribution of interest (rejection sampling) or importance weighted random samples from a candidate distribution close to that distribution (importance sampling) (Booth and Hobert 1999). A Metropolis-Hastings sampler is used here and the implementation is illustrated in Appendix 2.A.2. Incorporating the Metropolis step into the EM algorithm gives an algorithm as shown in Algorithm 2.1.

2.3.2 Stopping Rule

When the relative change in the parameter values from three successive iterations is small considers as a standard stopping rule or convergence criterion for deterministic EM algorithms is to stop and declare convergence. Here, we apply the stopping rule suggested by (Booth and Hobert 1999), which is

$$\max_d \left(\left| \frac{\hat{\boldsymbol{\xi}}_d^{(s)} - \hat{\boldsymbol{\xi}}_d^{(s-1)}}{\hat{\boldsymbol{\xi}}_d^{(s-1)} - \delta_1} \right| \right) < \delta_2, \tag{2.14}$$

where $(d = 1, \ldots, D)$ is the number of parameters in the parameter vector $\boldsymbol{\xi}$, and δ_1 and δ_2 are predetermined values.

2.3.3 Variance Estimation

The EM algorithm provides only the parameter estimates. Louis' formula (Louis 1982) is used for finding estimates corresponding variance-covariance matrix $\mathbf{I}_{\boldsymbol{\xi}}^{-1}$ as follows

$$\begin{aligned} \mathbf{I}_{\boldsymbol{\xi}} &= -\mathrm{E}\left(\frac{\partial^2 \mathcal{L}_{full}}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}} | \boldsymbol{w}, \hat{\boldsymbol{\xi}} \right) - \mathrm{Var}\left(\frac{\partial \mathcal{L}_{full}}{\partial \boldsymbol{\xi}} | \boldsymbol{w}, \hat{\boldsymbol{\xi}} \right) \\ &= -\mathrm{E}\left(\frac{\partial^2 \mathcal{L}_{full}}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'} | \boldsymbol{w}, \hat{\boldsymbol{\xi}} \right) - \mathrm{E}\left[\left(\frac{\partial \mathcal{L}_{full}}{\partial \boldsymbol{\xi}} \right) \left(\frac{\partial \mathcal{L}_{full}}{\partial \boldsymbol{\xi}'} \right) | \boldsymbol{w}, \hat{\boldsymbol{\xi}} \right]. \end{aligned} \tag{2.15}$$

Note that $\frac{\partial \mathcal{L}_{full}}{\partial \boldsymbol{\xi}} = 0$ when $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}$. The technique requires computation of a complete-data gradient vector and a second derivative matrix of the full data log-likelihood (see Appendix 2.A.3). We can construct the terms in (2.15) using samples of the last iteration.

2.3.4 Prediction of the Frailties

In applications the magnitude of the random effects or frailties is of interest. An important advantage of the frailty models is that they allow inference of the subject-specific random effects, and therefore a better understanding and interpretation of the variability in the data.

Prediction of the random effects $\boldsymbol{w}_i$ is based on the conditional expectation $\mathrm{E}(\boldsymbol{w}_i|\boldsymbol{\xi})$, and the variance $\mathrm{Var}(\boldsymbol{w}_i|\boldsymbol{\xi})$. These quantities are easily computed as by-products of the MCEM algorithm at convergence. The corresponding 95 % credible intervals using the normal approximation is $\mathrm{E}(\boldsymbol{w}_i|\boldsymbol{\xi}) \pm z_{(1-\alpha/2)} [\mathrm{Var}(\boldsymbol{w}_i|\boldsymbol{\xi})]^{\frac{1}{2}}$, or using the middle 95 % of the Monte Carlo sample from the conditional distribution $g(\boldsymbol{w}_i|\boldsymbol{\xi})$. We can also, build 95 % credible intervals for the frailties $\boldsymbol{u_i}$ based on the conditional expectation $\mathrm{E}(\boldsymbol{u}_i|\boldsymbol{\xi})$, and variance $\mathrm{Var}(\boldsymbol{u}_i|\boldsymbol{\xi})$, using the normal approximation as $\mathrm{E}(\boldsymbol{u}_i|\boldsymbol{\xi}) \pm z_{(1-\alpha/2)} [\mathrm{Var}(\boldsymbol{u}_i|\boldsymbol{\xi})]^{\frac{1}{2}}$, or using the the middle 95 % of the Monte Carlo sample from the conditional distribution.

2.4 Simulation Studies

Here we simulate data to examine the empirical performance for finite samples of the developed method.

2.4.1 Simulation Settings

To investigate the effect of sample size increasing on estimation performance, suppose that two types of events may occur. One treatment effect with two levels is considered in simulation. For every case, 500 simulated datasets were generated. All scenario combinations of the following settings are reported.

1. We considered two sample sizes with a variable number of subjects (100 and 200) and a variable number of recurrent events by subject.

2. Regression coefficients were set to be $\beta_1 = -1, \beta_2 = -.5$.

3. The multivariate distribution was simulated with variance-covariance matrix set to be

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

corresponding to independent, moderate, and high correlation cases. For each subject i, the data were generated by the following steps.

4. Generated i.i.d. bivariate normal random variables $\boldsymbol{w}_i = (w_{i1}, w_{i2})'$with mean $\boldsymbol{u} = (0, 0)'$and variance-covariance
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}.$$

5. The treatment effect variable x_i is generated from a Bernoulli distribution with parameter $(p = 0.5)$.

6. The fixed right-censored time was set at, $C = 1$, and 2, $i = 1, \cdots, m$.

7. The random censoring time C_i^* is assumed to be exponentially distributed with rate $\alpha_c = 0.5$.

8. C_i is the censored time for subject i^{th} where $C_i = \min\{C_i^*, C\}$.

9. For every event type ($j = 1$, 2), generate gap times z_{ijk} from the exponential distribution with $\lambda_{0j}(t) = 1$ and the rate parameter $\alpha_j = \left[\lambda_{0j}(t) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij})\right]^{-1}$, and let $Y_{ijk} = \sum_{k=1}^{k} z_{ijk}$ the observed event times $t_{ijk} = \min(C_i, Y_{ijk})$ with the first start time set to zero.

10. By straightforward way, the censoring indicator δ_{ijk} was set at 1 if $t_{ijk} = Y_{ijk}$ and $\delta_{ijk} = 0$ otherwise.

2.4.2 Results

Table 2.1 reports the distribution of subjects according to the expected number of events of each event type based on 10,000 samples with the parameters $\sigma_1^2 = \sigma_2^2 = 1$, $\beta_1 = -1$, $\beta_2 = -.5$, $\lambda_{0j}(t) = 1$, $\alpha_c = .5$, and $C = 1$ for every subject with m=100. In the conducted simulation studies, the average number of observed recurrent events per subject ranges from 0.40 to 1.00 for the event type I and between 0.84 to 1.80 for the event type II for the treatment and

placebo group, respectively. Between 29% and 33 % of the subjects did not have a recurrent event type I and from 24% to 30 % in the event type II for respectively the placebo and treatment group.

Table 2.1 – Expected number of recurrences for every 100 subject according to $\sigma_1^2 = \sigma_2^2 = 1$, $\beta_1 = -1$, $\beta_2 = -0.5$, $\lambda_{0j}(t) = 1$, $\alpha_c = 0.5$, and $C = 1$.

		Expected number of recurrences with m=100 subject.						
		0	1	2	3	4	≥ 5	No. event per subject
Type I	Total	62	22	8	4	2	2	0.70
	Placebo	29	8	5	4	2	2	1.00
	Treatment	33	12	3	0	0	0	0.40
Type II	Total	54	24	6	5	4	7	1.32
	Placebo	24	13	4	3	3	1	1.80
	Treatment	30	11	2	2	1	6	0.84

We used 500 simulation runs. All calculations for the simulation studies were performed using parallel computing by *snowfall* and *multicore* packages in R-project. The results are obtained based on the algorithm of Section 2.3. The mean square errors (MSEs) of the estimates are reported in Table 2.2.

Table 2.2 – Empirical mean square error of estimates for treatment effects and variance-covariance components with $\sigma_1^2 = \sigma_2^2 = 1$, $\beta_1 = -1$, $\beta_2 = -0.5$, $\lambda_{0j}(t) = 1$, and 500 repeats for every setting.

Setting			MSE				
C	m	σ_{12}	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_{12}$
1	100	0.0	0.090	0.079	0.092	0.090	0.064
		0.4	0.090	0.081	0.085	0.075	0.072
		0.8	0.083	0.073	0.082	0.042	0.091
	200	0.0	0.061	0.056	0.068	0.044	0.014
		0.4	0.042	0.037	0.047	0.017	0.021
		0.8	0.045	0.040	0.029	0.040	0.025
2	100	0.0	0.090	0.079	0.080	0.051	0.025
		0.4	0.055	0.060	0.059	0.071	0.035
		0.8	0.076	0.074	0.048	0.052	0.026
	200	0.0	0.031	0.040	0.049	0.017	0.013
		0.4	0.040	0.022	0.039	0.028	0.021
		0.8	0.039	0.029	0.016	0.013	0.023

As expected, in all simulations the estimates for the mean square errors were smaller for $m = 200$ than for $m = 100$. The simulation study shows that the regression coefficients and variance-covariance parameters from the proposed method were well estimated with small MSE values at all settings which implies the performance of the estimates is good with moderate samples.

2.5 Application to a Skin Cancer Dataset

Here we apply our proposed methodology to the data collected as a part of the Nutritional Prevention of Cancer trial conducted by Arizona Cancer Center (e.g., Clark et al. 1996, Duffield-Lillico et al. 2003) to study the efficacy of a nutritional supplement of Selenium in the prevention of skin cancer in high-risk patients. A total of 1312 patients were enrolled in the study. 120 patients were 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 1192 patients with 606 and 586 randomly assigned to treatment (Selenium) group ($x_i = 1$) and placebo group

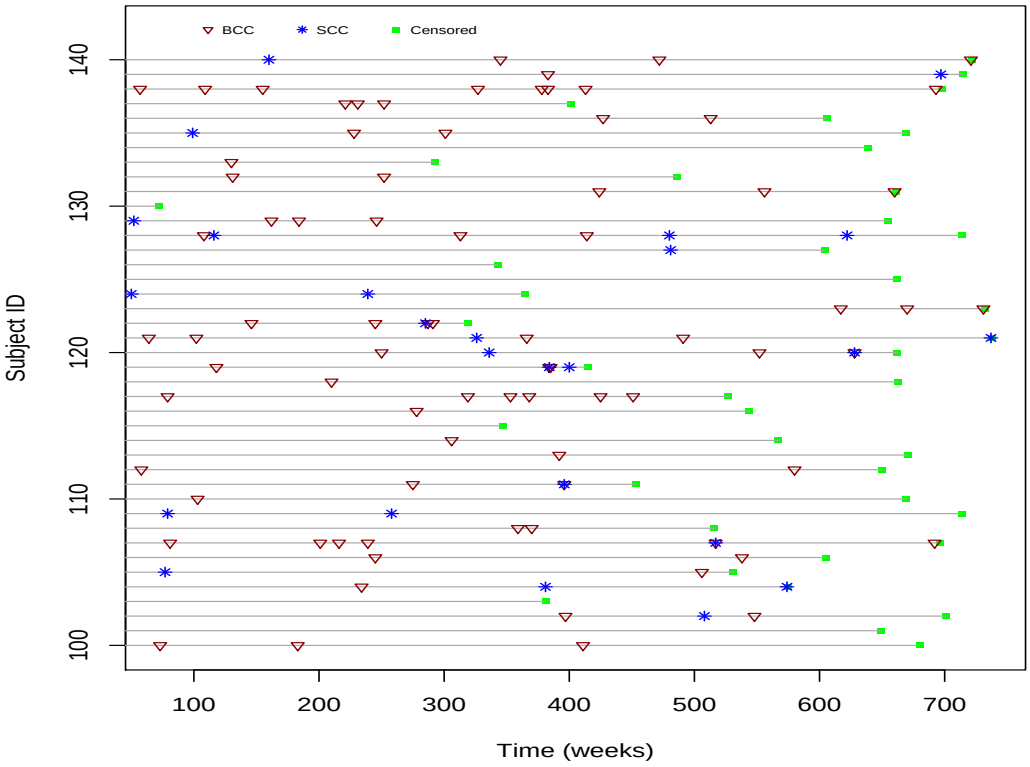


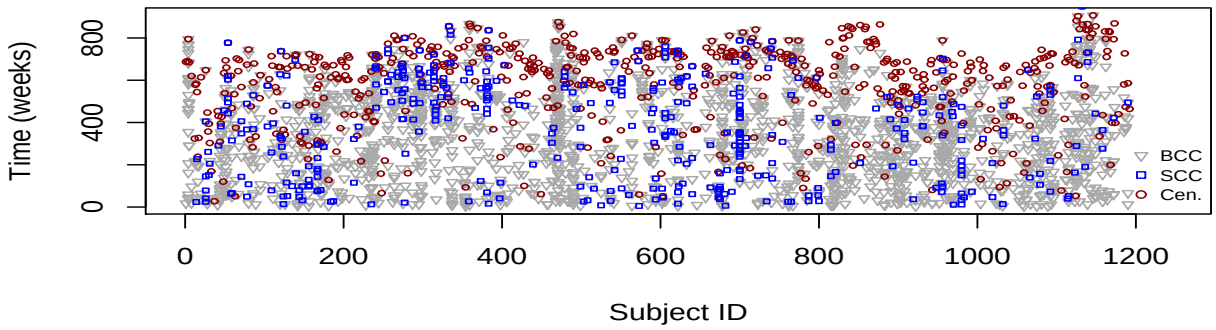
Figure 2.1 – Multi-type event plot for tumor occurrence times and types for a sample of 40 patients of the skin cancer dataset.

($x_i = 0$), respectively. The patients received a nutritional supplement of 200 micro-gram of Selenium per day or a placebo. Patients with previous history of either multiple basal cell carcinoma (BCC) or a squamous cell carcinoma (SCC) or otherwise at high risk were recruited into the trial in one of seven clinics and assigned at random to either treatment or placebo. The 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.

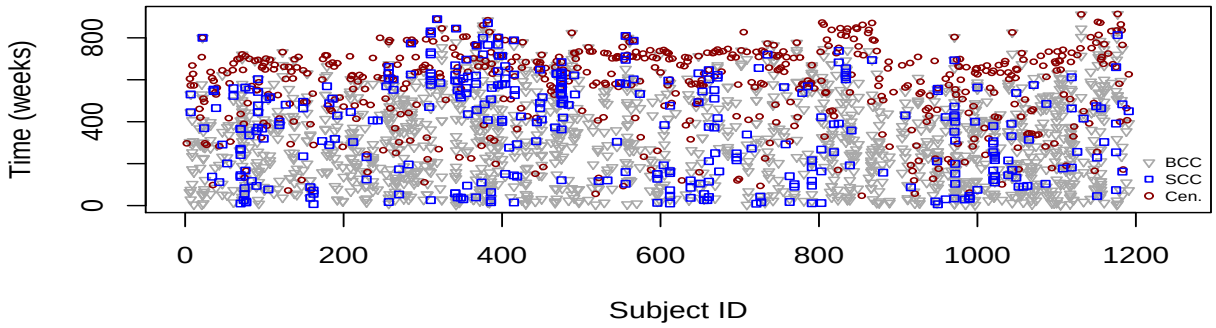
The events of interest are the successive incidence times of the two types of tumors, where we define an event of the first type to be the detection of one or more basal cell carcinoma (BCC) at a clinic visit. An event of the second type is similarly the medical detection of one or more squamous cell carcinoma (SCC) at a clinic visit.

Figure 2.1 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 observed to develop only new events of a particular type (e.g., subjects 100, 110, and 124), and some were experienced a multiple types (e.g., subjects 102, 120, and 140). The occurrence times and types (BCC, SCC, and censored) for 1192 patients in the clinical trial according to the treatment group are shown in Figure 2.2. It is clear that the occurrence of the two event types have higher intensity in the Selenium group than the placebo group.



(a) Selenium group



(b) Placebo group

Figure 2.2 – The occurrence times and types (BCC, SCC, and censored) for 1192 patients in the clinical trial for treatment groups.

Descriptive statistics for the total number of observed events and number of subjects affected by the events according to the cancer type and treatment group and the frequency distribution of the total number of patients according to the number of events from each type are given in Tables 2.3, and 2.4, respectively.

Table 2.3 – Frequency distribution of the number of subjects according to the number of events from each cancer type.

	BCC or SCC		BCC		SCC	
No. event	Placebo No.(%)	Selenium No.(%)	Placebo No.(%)	Selenium No.(%)	Placebo No.(%)	Selenium No.(%)
0 (Censored)	152 (0.259)	116 (0.191)	179 (0.305)	157 (0.259)	414 (0.706)	404 (0.667)
1	105 (0.179)	125 (0.206)	120 (0.205)	133 (0.219)	96 (0.164)	120 (0.198)
2	71 (0.121)	81 (0.134)	73 (0.125)	80 (0.132)	37 (0.063)	41 (0.068)
3	57 (0.097)	66 (0.109)	42 (0.072)	72 (0.119)	16 (0.027)	19 (0.031)
4	48 (0.082)	51 (0.084)	47 (0.080)	43 (0.071)	9 (0.015)	6 (0.010)
5	31 (0.053)	28 (0.046)	30 (0.007)	17 (0.008)	4 (0.007)	5 (0.008)
≥ 6	122 (0.208)	139 (0.229)	95 (0.017)	104 (0.018)	10 (0.017)	11 (0.018)
Max. no. events	28	34	22	33	18	28

Table 2.4 – Number of observed events and patients affected by cancer types according to the cancer type and treatment group.

	Total (1192 subject)			Placebo (586 subject)			Selenium (606 subject)		
Cancer type	No. subject	No. events	No. event per subject	No. subject	No. events	No. event per subject	No. subject	No. events	No. event per subject
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

A variety of analysis were carried out for the two types of events (BCC type I; SCC type II) including single-type Andersen-Gill (AG) analysis with the cluster (subject ID) and obtain robust standard errors for the parameter estimates, single-type semiparametric frailty gamma model (Gamma), and single-type semiparametric model using the normal distribution for the random effects (Gaussian). The proposed joint multi-type recurrent events model using the MCEM proposed algorithm (Multi-type). The developed algorithm converged after

59 iterations when we used $\delta_1 = 0.001$, and $\delta_2 = 0.002$ for the stopping rule in (2.14). The initials values for $\boldsymbol{\xi}$ were the estimated values from the corresponding Cox model with Gaussian frailties.

For the proposed model the estimates for the treatment effects and variance-covariance component parameters are obtained and respectively reported in Tables 2.6, 2.6, and 2.7. Standard errors for the estimates are calculated using Louis' formula. An approximate 95% confidence interval for $\boldsymbol{\xi}$ is $\hat{\boldsymbol{\xi}} \pm z_{(1-\alpha/2)}\text{S.E.}\left(\hat{\boldsymbol{\xi}}\right)$.

Table 2.5 – Estimates of the treatment effect of BCC event type for single-type models and the proposed multi-type model.

Parameter	β_1				RR (β_1)	
Method	EST.	S.E.	p-value	95 % CI	EST.	95 % CI
Multi-type	0.091	0.034	0.007	0.024 : 0.158	1.095	1.025 : 1.171
Gaussian	0.072	0.034	0.034	0.005 : 0.139	1.075	1.005 : 1.149
Gamma	0.101	0.035	0.004	0.032 : 0.170	1.106	1.033 : 1.185
AG	0.114	0.074	0.123	0.032 : 0.170	1.121	0.969 : 1.296

The results reveal broadly comparable estimates for both the treatment effect ($\hat{\beta}_1$ and $\hat{\beta}_2$) and the variance-covariance components ($\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, and $\hat{\sigma}_{12}$). For the multi-type model, there is an estimated 9.5 % increasing in the rate of BCC in the Selenium group more than placebo group with estimated relative risk RR(β_1) = 1.095 (95 % CI : 1.025, 1.171; p-value =.007).

Table 2.6 – Estimates of the treatment effect of BCC event type for single-type models and the proposed multi-type model.

Parameter	β_2				RR (β_2)	
Method	EST.	S.E.	p-value	95 % CI	EST.	95 % CI
Multi-type	0.143	0.070	0.041	0.006 : 0.280	1.154	1.006 : 1.323
Gaussian	0.075	0.083	0.366	-0.088 : 0.238	1.078	0.916 : 1.268
Gamma	0.079	0.076	0.299	-0.070 : 0.228	1.082	0.932 : 1.256
AG	0.089	0.140	0.525	-0.185 : 0.363	1.093	0.831 : 1.438

The rate of SCC is significantly higher in the Selenium group with estimated RR(β_2) = 1.154 (95 % CI : 1.006, 1.323; p-value =0.041) in the multi-type model by considering the

correlation between tumor types. While results in table 2.6 show that (β_2) is not significant in the single-type models with frailty (Gaussian and Gamma) or without frailties (AG).

Table 2.7 – Estimates of the variance-covariance component parameters of single-type models and the multi-type model.

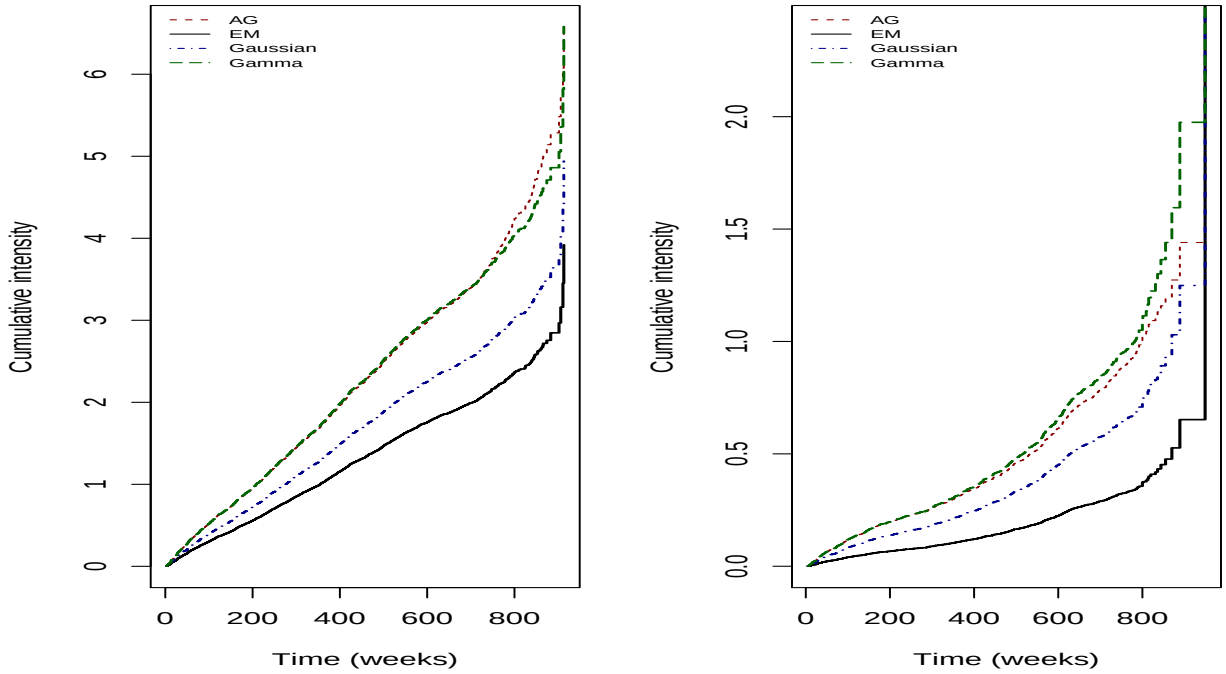
Parameter	σ_1^2			σ_2^2			σ_{12}		
Method	EST.	S.E.	95 % CI	EST.	S.E.	95 % CI	EST.	S.E.	95 % CI
MCEM	1.024	0.019	0.987 : 1.061	2.094	0.024	2.047 : 2.141	0.464	0.022	0.421 : 0.507
Gaussian	0.819			1.29					
Gamma	0.782			1.82					
AG									

There are significant variances for the random effects of both event types whereas $\hat{\sigma}_1^2$= 1.024 (95 % CI : 0.987, 1.061) and $\hat{\sigma}_2^2$= 2.094 (95 % CI : 2.047, 2.141). The point estimate of the covariance between random effects for the two event types equals 0.464 with (95 % CI : 0.421, 0.507). The results from the mylti-type model indicate that the treatment effects associated with a significant increase in both BCC and SCC incidence. The correlation coefficient (ρ_{12}) is estimated at 0.31 that indicates the patients with higher event rates of one cancer type tend to have higher event rates of the other.

Plots of the estimated baseline cumulative intensity functions for the BCC and the SCC are given in Figure 2.3. The plots contain the Nelson-Aalen estimates of the cumulative baseline intensity functions obtained from single-type models and multi-type event model. For both tumor types, the plots show decreasing in the estimates of the cumulative functions for the multi-type event model comparing with single-type models through all the distinct event times.

The predicted frailties are plotted in Figure 2.4 with the number of events. It is shown that the patients with higher number of observed event types have the higher expected frailty for both tumor types BCC and SCC. It is clear also to see the correlation between the expected frailties of the two tumor types.

Figure 2.5 shows the 95 % credible intervals for the random effects of patients who had

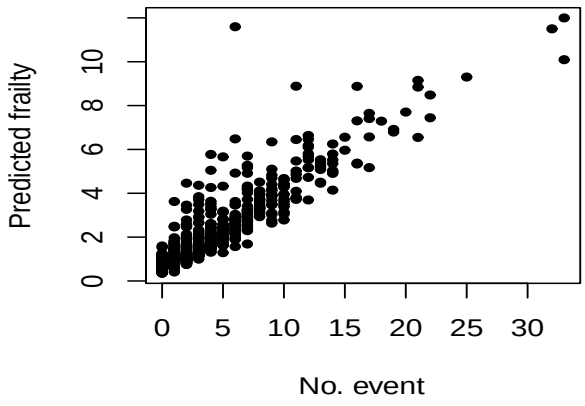


(a) BCC type **(b)** SCC type

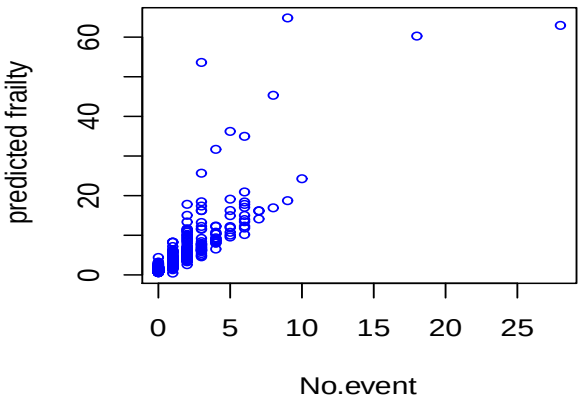
Figure 2.3 – Cumulative intensity functions for BCC and SCC tumor types.

two events according to the tumor types (BCC and SCC) and the treatment group (Placebo or Selenium). They become narrower with increasing group size. It is also clear to see that the length of the credible intervals (vertical bars) for all groups are close to each other since all patients in this group have the same number of events.

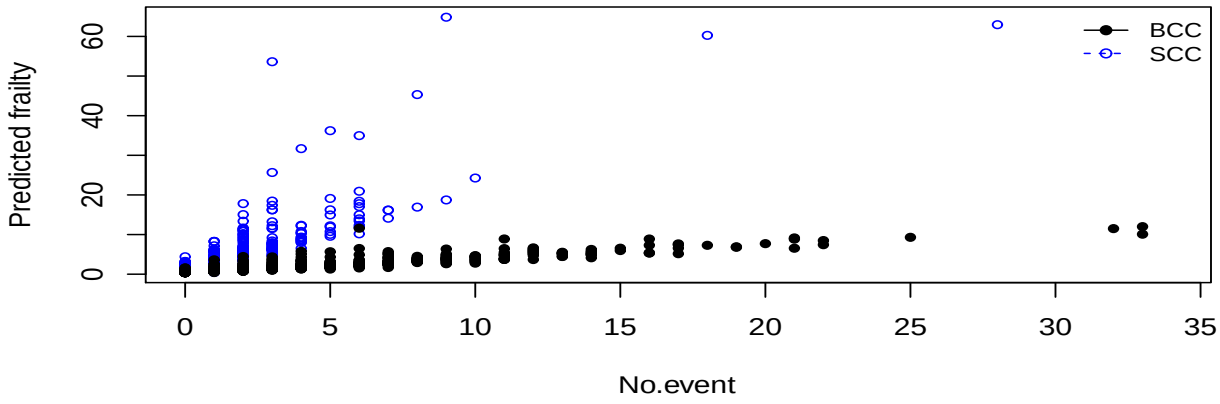
In the E-step of the MCEM algorithm, we used the Metropolis-Hastings algorithm with a total sample of 7000 iterations with a burn-in sample of 1000. We checked the convergence of the MCMC rounds through visual inspection. Figure 2.6 shows the MCMC trace plot, a running mean plot which is a plot of the iterations against the mean of the draws up to each iteration, and the ACF plot to assess the autocorrelation between the draws of our Markov chain. The plots show the good convergence of the MCMC chain within the chosen initial values. Figure 2.6 also includes the trace plot for the average of the rejection rate (r) to be around 0.55.



(a) BCC

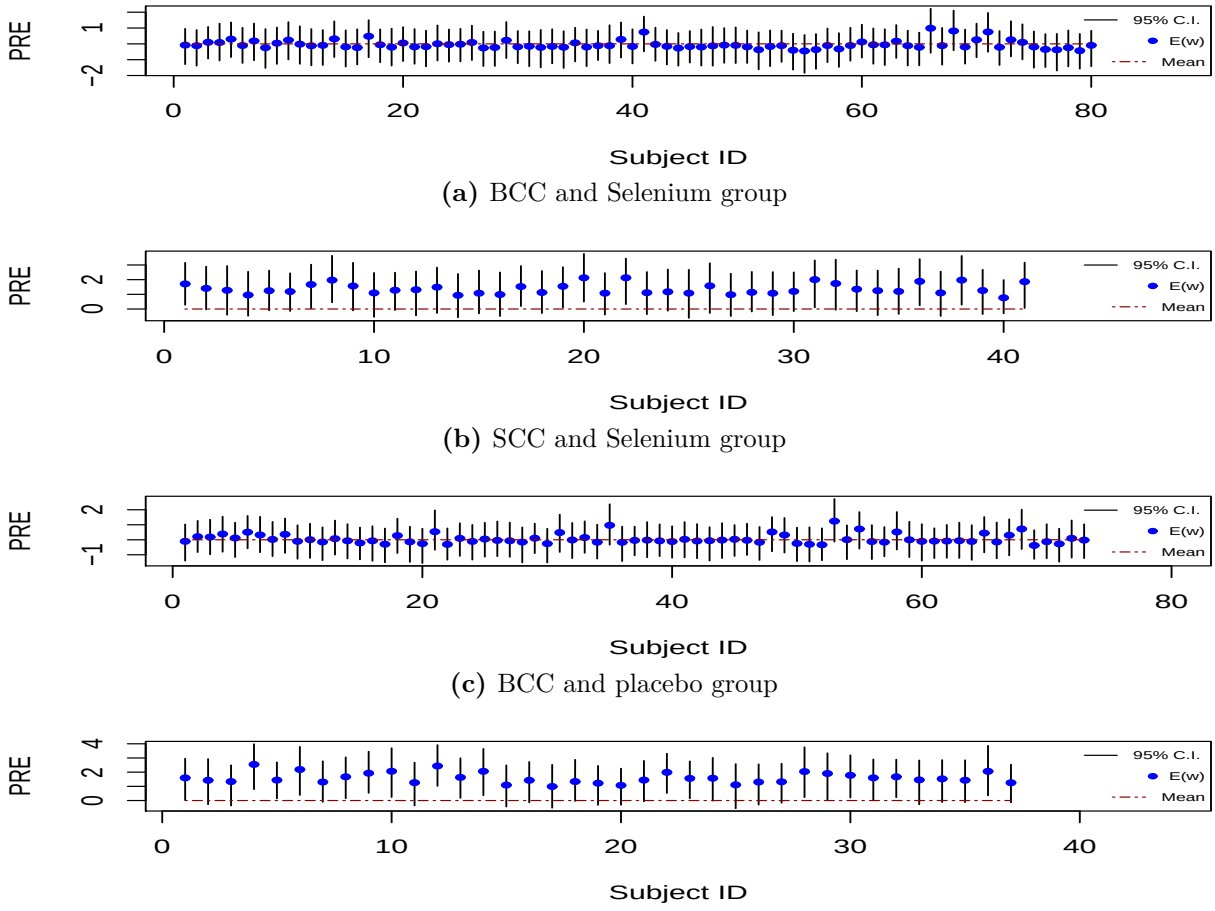


(b) SCC



(c) BCC and SCC

Figure 2.4 – Predicted frailties with the number of events for BCC and SCC tumor types.



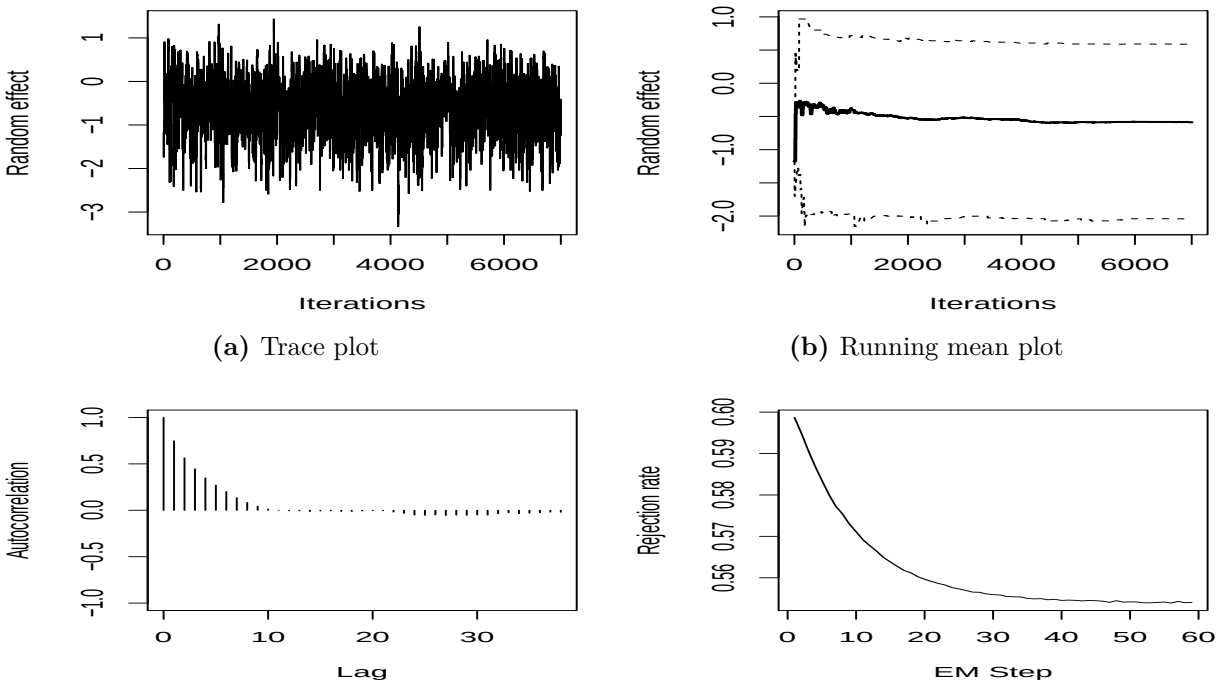
(a) BCC and Selenium group

(b) SCC and Selenium group

(c) BCC and placebo group

(d) SCC and placebo group

Figure 2.5 – Predicted random effects (PRE) for the patients with two event according to the tumor type and treatment group with a ref. line at a mean equals to zero.



(a) Trace plot

(b) Running mean plot

(c) ACF plot

(d) MCMC rejection rate

Figure 2.6 – (a) Trace plot of simulated draws of $\boldsymbol{w_i}$ for an MCMC chain, (b) Running mean plot, (c) ACF, running mean plot, and (d) Trace plot for the MCMC rejection rate.

2.6 Concluding Remarks and Areas for Future Research

This chapter provides a semiparametric methodology to characterize the incidence rate of tumor types, estimate the impact of covariates, and understand the correlation structure among tumor types. Maximum likelihood estimates of the regression coefficients, variance-covariance components, nonparametric baseline intensity functions, and credible intervals for the subject-specific random effects are obtained based on an MCEM algorithm. Where we use Metropolis-Hastings sampling to draw from the conditional distribution of the random effects, the convergence of the algorithm is assessed by the visual inspection of the estimates. A convergence criterion to stop and declare convergence of the algorithm has been used. The score vector and the information matrix for the complete data are derived here to adopt Louis' formula for variance estimations. The proposed model performs well for simulated datasets with different censoring rates and number of events per subject. The method is illustrated using dataset collected from the study on 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 to model the correlated random effects. The algorithm can be modified for other multivariate distributions desired for the random effects. it is also straightforward to adapt the model for data with time dependent covariates, multiple-event types with more than two event processes, and multivariate frailty for each type without restrictions to the frailty dimension. Further generalizations of the model are not considered in this chapter but can be made in a future work include multilevel frailty, dependent censoring, and assessing the goodness of fit and the influence analysis. Automated MCEM algorithms (e.g., Booth and Hobert 1999, Levine and Fan 2004) could be adapted for the current model and study the efficacy of the different routines to reduce the cost of computing. Finally, using the copula function to model the multivariate correlated frailties is an area of our future work.

Acknowledgments

We thank Dr. Mary Reid, Roswell Park Cancer Institute, for the permission to use the dataset from the Nutritional Prevention of Cancer clinical trial. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources and technical support that have contributed to the results reported within this chapter (http://www.arc.vt.edu). We thank Yuanyuan Duan and Zhibing Xu, Department of Statistics (Virginia Tech), for their comments. The work by Bedair was partially funded by Tanta University (Egypt). The work by Hong was partially supported by the National since Foundation under Grant BCS-1122876 to Virginia Tech.

2.A Appendices

2.A.1 The Conditional Distribution for the Random Effects

Here we provide the conditional distribution for the random effects when the marginal distribution is $\boldsymbol{w} \sim \text{MVN}\left[\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}}\right]$. We define

$$L_{i(full)}\left(\boldsymbol{\xi} \mid \boldsymbol{w}_i\right) = \prod_{j=1}^{J} \prod_{k=1}^{N_{ij}(\tau_i)} \left[\lambda_{ij}(t)\right]^{\delta_{ijk}} \exp\left[-\Lambda_{ij}(\tau_i)\right] f(\boldsymbol{w}_i|\boldsymbol{\theta}).$$

Let

$$\begin{aligned}
\mathcal{L}_{i(full1)} &= \sum_{j=1}^{J} \left\{ \sum_{k=1}^{N_{ij}(\tau_i)} \left[\delta_{ijk} \log\left(\lambda_{ij}\left(t\right)\right)\right] - \Lambda_{ij}\left(\tau_i\right) \right\} \\
&= \sum_{j=1}^{J} \left\{ \sum_{k=1}^{N_{ij}(\tau_i)} \delta_{ijk} \log\left[\lambda_{0j}\left(t\right) \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij}\right)\right] - \Lambda_{0j}\left(\tau_i\right) \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij}\right) \right\} \\
&= \sum_{j=1}^{J} \left\{ N_{ij}(\tau_i) \left[\log\left(\lambda_{0j}\left(t\right)\right) + \boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij}\right] - \Lambda_{0j}\left(\tau_i\right) \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij}\right) \right\}.
\end{aligned}$$

We write

$$\mathcal{L}_{i(full2)} = -\frac{Jm}{2}\log 2\pi - \frac{1}{2}\log(|\Sigma|) - \frac{1}{2}\left(\boldsymbol{w}_i'\Sigma^{-1}\boldsymbol{w}_i\right).$$

By the Bayesian formula, the logarithm of the distribution of $\boldsymbol{w}$ given data with parameters $(\boldsymbol{\xi})$ can be given by

$$L_{full}(\boldsymbol{w}|\boldsymbol{\xi}) \propto \exp\left[\sum_{i=1}^{m}\sum_{j=1}^{J} N_{ij}(\tau_i)w_{ij} - \Lambda_{0j}(\tau_i)u_{ij}\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j) - \frac{1}{2}\left(\boldsymbol{w}_i'\Sigma^{-1}\boldsymbol{w}_i\right)\right].$$

2.A.2 Metropolis-Hastings Algorithm

Updating $\boldsymbol{w}_i$ in a Markov chain can be proceed by proposing a new $\boldsymbol{w}_i^\star$ based on the current parameter values and then accepting or rejecting it with the appropriate probability. A standard proposal distribution in this case would be the multivariate normal distribution with mean equal to the current value $\boldsymbol{w}_i^{(s-1)}$ and with some proposal variances. The Metropolis-Hastings algorithm steps are shown in Algorithm 2.2.

Algorithm: 2.2 A Metropolis-Hastings algorithm for the E-step in the MCEM algorithm.

---

1. Choose a start value $\boldsymbol{w}^{(0)} = (\boldsymbol{w}_1^{(0)}, \ldots, \boldsymbol{w}_m^{(0)})'$. At iteration s,
2. Draw a candidate value $\boldsymbol{w}_i^\star$, for the i^{th} component of $\boldsymbol{w}$ from a proposal distribution $q_k(\boldsymbol{w}_i^\star|\boldsymbol{w}_i^{(s-1)})$.
3. Compute an acceptance ratio (probability) $r^{(s)} = \dfrac{g\left(\boldsymbol{w}_i^\star|\boldsymbol{\xi}\right)q_K\left(\boldsymbol{w}_i^{(s-1)}|\boldsymbol{w}_i^\star\right)}{g\left(\boldsymbol{w}_i^{(s-1)}|\boldsymbol{\xi}\right)q_K\left(\boldsymbol{w}_i^\star|\boldsymbol{w}_i^{(s-1)}\right)}$..
4. Accept $\boldsymbol{w}_i^\star$ as the new value with probability $\min(1,\, r^{(s)})$; otherwise, $\boldsymbol{w}_i^{(s)} = \boldsymbol{w}_i^{(s-1)}$.
5. Repeat steps 2 to 4 for generating S samples $\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(S)}$.

In case we propose from a symmetric distribution it is obviously to note that $q_K\left(\boldsymbol{w}_i^{(s-1)}|\boldsymbol{w}_i^\star\right) = q_K\left(\boldsymbol{w}_i^\star|\boldsymbol{w}_i^{(s-1)}\right)$.

2.A.3 Louis' Formula Components

In this section, we provide the first and second partial derivatives for $\mathcal{L}_{full}$ which are required in Louis' s formula for the variance of the estimates.

2.A.3.1 First Derivatives

The first derivatives can be derived as

$$\begin{aligned}
\frac{\partial \mathcal{L}_{full1}}{\partial \boldsymbol{\beta}_j} &= \sum_{i=1}^{m}\left[\left(\sum_{k=1}^{N_{ij}(\tau_i)}\delta_{ijk}\boldsymbol{x}_i\right) - \boldsymbol{x}_i\Lambda_{0j}(\tau_i)u_{ij}\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\right] \\
&= \sum_{i=1}^{m}\left[N_{ij}(\tau_i)\boldsymbol{x}_i - \boldsymbol{x}_i\Lambda_{0j}(\tau_i)u_{ij}\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\right]. \\
\frac{\partial \mathcal{L}_{partial}}{\partial \boldsymbol{\beta}_j} &= \sum_{l_j=1}^{L_j}\left[\boldsymbol{x}_i - \frac{\sum_{i\in R(l_j)}\boldsymbol{x}_i\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij})N\left(t_{j(l_j)}\right)}{\sum_{i\in R(l_j)}\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij})}\right].
\end{aligned}$$

See, for example, (Klein and Moeschberger 2003, p. 254).

$$\frac{\partial \mathcal{L}_{full1}}{\partial \lambda_{l_j}} = \frac{1}{\lambda_{l_j}} - \sum_{i\in R_{t_{(l_j)}}}\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij}),\ l_j = 1,\ldots,L_j.$$

Let $\mathrm{A} = (\sigma_1^2\sigma_2^2 - \sigma_{12}^2)$, and $\mathrm{B}_i = (w_{1i}^2\sigma_2^2 + w_{2i}^2\sigma_1^2 - 2\sigma_{12}w_{1i}w_{2i})$, we can write $\mathcal{L}_{full2}$ in the proportional form as

$$\mathcal{L}_{(full2)} \propto -1/2\left[m\log\mathrm{A}\right] + \left[1/\mathrm{A}\right]\sum_{i=1}^{m}\mathrm{B}_i.$$

The first derivatives with respect to the variance-covariance components are

$$\begin{aligned}
\frac{\partial \mathcal{L}_{full2}}{\partial \sigma_1^2} &= \frac{1}{2}\left[\frac{\sigma_2^2\sum_{i=1}^{m}\mathrm{B}_i}{\mathrm{A}^2} - \frac{(m\sigma_2^2 + \sum_{i=1}^{m}w_{2i}^2)}{\mathrm{A}}\right], \\
\frac{\partial \mathcal{L}_{full2}}{\partial \sigma_2^2} &= \frac{1}{2}\left[\frac{\sigma_1^2\sum_{i=1}^{m}\mathrm{B}_i}{\mathrm{A}^2} - \frac{(m\sigma_1^2 + \sum_{i=1}^{m}w_{1i}^2)}{\mathrm{A}}\right],
\end{aligned}$$

$$\frac{\partial \mathcal{L}_{full2}}{\partial \sigma_{12}} = \frac{1}{2}\left[-\frac{2\sigma_{12}\sum_{i=1}^{m}\mathrm{B}_i}{\mathrm{A}^2} + \frac{2\left(m\sigma_{12} + \sum_{i=1}^{m} w_{1i}w_{2i}\right)}{\mathrm{A}}\right].$$

2.A.3.2 Second Derivatives

The second derivatives of the full log-likelihood are

$$\frac{\partial^{(2)}\mathcal{L}_{(full1)}}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}_{j'}} = \sum_{i=1}^{m}\left[-\boldsymbol{x}_i\boldsymbol{x}_i'\Lambda_{0j}(\tau_i)u_{ij}\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\right]$$

.

$$\frac{\partial^{(2)}\mathcal{L}_{(full1)}}{\partial \lambda_{l_j}^2} = \left\{-\frac{N\left[t_{j(l_j)}\right]}{\lambda_{l_j}^2}\right\},$$

$$\frac{\partial^{(2)}\mathcal{L}_{full1}}{\partial \lambda_{l_j}\partial \boldsymbol{\beta}_j} = -\sum_{i\in R_{t(l_j)}} \boldsymbol{x}_i' \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j + w_{ij}\right),$$

$$\frac{\partial^{(2)}\mathcal{L}_{full2}}{\partial\left(\sigma_1^2\right)^2} = \frac{1}{2}\left[-\frac{2\sigma_2^4\sum_{i=1}^{m}\mathrm{B}_i}{\mathrm{A}^3} + \frac{m\sigma_2^4 + 2\sigma_2^2\sum_{i=1}^{m} w_{2i}^2}{\mathrm{A}^2}\right],$$

$$\frac{\partial^{2}\mathcal{L}_{full2}}{\partial\left(\sigma_2^2\right)^2} = \frac{1}{2}\left[-\frac{2\sigma_1^4\sum_{i=1}^{m}\mathrm{B}_i}{\mathrm{A}^3} + \frac{m\sigma_1^4 + 2\sigma_1^2\sum_{i=1}^{m} w_{1i}^2}{\mathrm{A}^2}\right],$$

$$\frac{\partial^{(2)}\mathcal{L}_{full2}}{\partial\left(\sigma_{12}\right)^2} = \frac{1}{2}\left[\frac{8\sigma_{12}\sum_{i=1}^{m} w_{1i}w_{2i}}{\mathrm{A}^2} - \sum_{i=1}^{m}\mathrm{B}_i\left(\frac{8\sigma_{12}^2}{\mathrm{A}^3} + \frac{2}{\mathrm{A}^2}\right) - \left(-\frac{4\sigma_{12}^2}{\mathrm{A}^2} - \frac{2}{\mathrm{A}}\right)\right],$$

$$\frac{\partial^{(2)}\mathcal{L}_{full2}}{\partial\sigma_1^2\partial\sigma_2^2} = \frac{1}{2}\left[-\frac{2\sigma_1^2\sigma_2^2\sum_{i=1}^{m}\mathrm{B}_i}{\mathrm{A}^3} + \frac{\mathrm{B}_i + m\sigma_1^2\sigma_2^2 + \sum_{i=1}^{m} w_{2i}^2\sigma_1^2 + w_{1i}^2\sigma_2^2}{\mathrm{A}^2} - \frac{m}{\mathrm{A}}\right],$$

$$\frac{\partial^{(2)}\mathcal{L}_{full2}}{\partial\sigma_1^2\partial\sigma_{12}} = \frac{1}{2}\left[\frac{4\sigma_2^2\sigma_{12}\sum_{i=1}^{m}\mathrm{B}_i}{\mathrm{A}^3} - \frac{2\left(m\sigma_2^2\sigma_{12} + \sum_{i=1}^{m} w_{1i}w_{2i}\sigma_2^2 + w_{2i}^2\sigma_{12}\right)}{\mathrm{A}^2}\right],$$

$$\frac{\partial^{(2)}\mathcal{L}_{full2}}{\partial\sigma_2^2\partial\sigma_{12}} = \frac{1}{2}\left[\frac{4\sigma_1^2\sigma_{12}\sum_{i=1}^{m}\mathrm{B}_i}{\mathrm{A}^3} - \frac{2\left(m\sigma_1^2\sigma_{12} + \sum_{i=1}^{m} w_{1i}w_{2i}\sigma_1^2 + w_{1i}^2\sigma_{12}\right)}{\mathrm{A}^2}\right].$$

Bibliography

Aalen, O.: 1978, Nonparametric inference for a family of counting processes, *The Annals of Statistics* **6**(4), 701–726.

Aalen, O. O., Bjertness, E. and Sønju, T.: 2007, Analysis of dependent survival data applied to lifetimes of amalgam fillings, *Statistics in Medicine* **14**(16), 1819–1829.

Abu-Libdeh, H., Turnbull, B. W. and Clark, L. C.: 1990, Analysis of multi-type recurrent events in longitudinal studies; application to a skin cancer prevention trial, *Biometrics* **16**(4), 1017–1034.

Booth, J. G. and Hobert, J. P.: 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 Schaubel, D. E.: 2004, Marginal means/rates models for multiple type recurrent event data, *Lifetime Data Analysis* **10**(2), 121–138.

Chen, B. E. and Cook, R. J.: 2009, The analysis of multivariate recurrent events with partially missing event types, *Lifetime Data Analysis* **15**(1), 41–58.

Chen, B. E., Cook, R. J., Lawless, J. F. and Zhan, M.: 2004, Statistical methods for multivariate interval-censored recurrent events, *Statistics in Medicine* **24**(5), 671–691.

Clark, L. C., Combs, G. F., Turnbull, B. W., Slate, E. H., Chalker, D. K., Chow, J., Davis, L. S., Glover, R. A., Graham, G. F., Gross, E. G. et al.: 1996, Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skina randomized controlled trial, *JAMA: the Journal of the American Medical Association* **276**(24), 1957–1963.

Cook, R. J. and Lawless, J. F.: 2007, *The Statistical Analysis of Recurrent Events*, Springer, New York.

Cook, R. J., Lawless, J. F. and Lee, K.-A.: 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., Ng, E., Mukherjee, J. and Vaughan, D.: 1999, Two-state mixed renewal processes for chronic disease, *Statistics in Medicine* **18**(2), 175–188.

Duchateau, L., Janssen, P., Kezic, I. and Fortpied, C.: 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., Slate, E. H., Reid, M. E., Turnbull, B. W., Wilkins, P. A., Combs, G. F., Park, H. K., Gross, E. G., Graham, G. F., Stratton, M. S. 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.

Ezell, M. E., Land, K. C. and Cohen, L. E.: 2003, Modeling multiple failure time data: A survey of variance–corrected proportional hazards models with empirical applications to arrest data, *Sociological Methodology* **33**(1), 111–167.

Finkelstein, M.: 2008, *Failure Rate Modelling for Reliability and Risk*, Springer, New York.

Hoff, P. D.: 2009, *A First Course in Bayesian Statistical Methods*, Springer, New York.

Johansen, S.: 1983, An extension of Cox's regression model, *International Statistical Review* **51**(2), 165–174.

Klein, J. P. and Moeschberger, M. L.: 2003, *Survival Analysis*, 2nd edn, Springer, New York.

Levine, R. A. and Fan, J.: 2004, An automated (Markov chain) Monte Carlo EM algorithm, *Journal of Statistical Computation and Simulation* **74**(5), 349–360.

Lim, H. J., Liu, J. and Melzer-Lange, M.: 2007, Comparison of methods for analyzing recurrent events data: application to the emergency department visits of pediatric firearm victims, *Accident Analysis & Prevention* **39**(2), 290–299.

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 Meyer, R.: 2005, Bayesian inference for recurrent events data using time-dependent frailty, *Statistics in Medicine* **24**(8), 1263–1274.

McCulloch, C. E.: 1994, Maximum likelihood variance components estimation for binary data, *Journal of the American Statistical Association* **89**(425), 330–335.

McCulloch, C. E.: 1997, Maximum likelihood algorithms for generalized linear mixed models, *Journal of the American Statistical Association* **92**(437), 162–170.

McGilchrist, C. and Aisbett, C.: 1991, Regression with frailty in survival analysis, *Biometrics* **47**(2), 461–466.

McGilchrist, C. and Yau, K.: 2008, Survival analysis with time dependent frailty using a longitudinal model, *Australian Journal of Statistics* **38**(1), 53–60.

McLachlan, G. J. and Krishnan, T.: 2007, *The EM Algorithm and Extensions*, Vol. 382, Wiley-Interscience, New Jersey.

Moreno, E. S. S.: 2008, *Nonparametric Frailty Models for Clustered Survival Data*, PhD thesis, Cornell University.

Ng, E. and Cook, R. J.: 1999, Robust inference for bivariate point processes, *Canadian Journal of Statistics* **27**(3), 509–524.

Ripatti, S., Larsen, K. and Palmgren, J.: 2002, Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm, *Lifetime Data Analysis* **8**(4), 349–360.

Ripatti, S. and Palmgren, J.: 2000, Estimation of multivariate frailty models using penalized partial likelihood, *Biometrics* **56**(4), 1016–1022.

Robert, C. and Casella, G.: 2009, *Introducing Monte Carlo Methods with R*, Springer, New York.

Vaida, F., Xu, R. et al.: 2000, Proportional hazards model with random effects, *Statistics in Medicine* **19**(24), 3309–3324.

Wei, G. C. and Tanner, M. A.: 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.

Wienke, A.: 2010, *Frailty Models in Survival Analysis*, CRC Press, New York.

Yau, K. and McGilchrist, C.: 1998, ML and REML estimation in survival analysis with time dependent correlated frailty, *Statistics in Medicine* **17**(11), 1201–1213.

Chapter 3 Copula Based Semiparametric Multivariate Frailty Models for Multi-type Recurrent Event Data: Application to Cancer Data

3.1 Introduction

3.1.1 Background

Recurrent event data arise in a wide variety of situations including medical studies, public health, economics, demography, transportation, and engineering. Multivariate (or correlated) survival data are frequently encountered in biomedical researches where clustered or multiple event times are observed. Medical studies involving recurrent asthma events and patients with a chronic disease may experience repeated hospitalizations. In other situations the recurrent events could be stopped by a terminal event. For example, recurrence of serious events (e.g., heart attacks) increases the risk of death.

Here, we are primarily interested in the analysis of the multi-type recurrent event data which arise when two or more different kinds of events may occur repeatedly over an observation period. The motivation comes from clinical trial data collected to study the efficacy of a nutritional supplement of Selenium in help to prevent of different types of skin cancer (Clark et al. 1996, Duffield-Lillico et al. 2003). Our aim in such settings is often to describe features of the marginal process and to study the association between the different types of events. Similarly, in a study of infections following bone marrow transplantation, it is of interest to study different types of recurrent infections simultaneously (e.g., bacterial, fungal and viral infections).

The concept of frailties or random effects provides a good tool to incorporate the association between survival times/recurrent events into the hazard (intensity) function. Likelihood

estimation in frailty models remains to be a challenge because the marginal likelihood involves analytically intractable integrals whose dimension depends on the structure of the random effects. A numerical integration approach is used by Cook et al. (2010), but is generally not recommended when the dimensionality of the integral is not small (e.g., the dimension is larger than two) (An and Bentler 2012). The Monte Carlo simulation is more efficient and stable in this regard.

A common assumption underlying many approaches developed for frailty models is that the frailty terms (random effects) have a distribution belonging to some parametric family, almost always the normal distribution. However, this assumption may be unrealistic, raising concerns over the validity of inferences both on fixed and random effects if it is violated. Moreover, allowing the random effects distribution to have more complex features than the symmetric, unimodal normal density may provide insight into underlying heterogeneity and even suggest failure to include important covariates in the model. Accordingly, considerable interest has focused on approaches that allow more variety in selecting the multivariate frailty models.

Many alternatives to parametric modeling assume only that the random effects distribution has a smooth density and represent the density in different ways. In particular, Chen et al. (2002) and Moreno (2008) proposed smooth and nonparametric random effects in case of generalized linear mixed models and clustered survival data, respectively.

Here, we consider an another approach consisting of a class of multivariate probability models for the multivariate frailties known as "copula functions" which have a number of attractive features. First, they permit the combination of any univariate marginal distributions not necessary to be from the same distribution family. Second, a particular class of copula models, called elliptical copula, have the property that they increase in complexity at a much slower rate than existing multivariate probability distributions as the number of dimensions increase. Third, they are very general, involving a number of existing multivariate models, and provide a framework for generating many more. These advantages give copula models

a greater potential for use in empirical analysis than many existing multivariate probability models.

3.1.2 Related Literature

There has been a widely growing research interest in modeling correlated failure times where clustered or multiple recurrent event times are observed. The association between survival times has been usually modeled by introducing a frailty component (random effect) into the hazard function. These generalizations of semi-parametric model (Cox 1972) have now been widely used.

For a single-type recurrent event in the absence of informative censoring, a variety of statistical methods (e.g., Therneau and Grambsch 2000) have been proposed. In-subject correlation is considered through stratification of the baseline hazard (Prentice et al. 1981, Wei et al. 1989). In the model introduced by Andersen and Gill (1982), independent increments are assumed.

Parametric and nonparametric frailty models have been widely adapted for the analysis of survival data (Hougaard 2000, Duchateau and Janssen 2008). For all observations of one subject, the same frailty term is used. The frailty variables of different subjects are treated as independent realizations of a common frailty distribution (Duchateau et al. 2003). Frailty models specially designed for recurrence data are considered in detail by McGilchrist and Yau (2008), Aalen et al. (2007), Manda and Meyer (2005), and Yau and McGilchrist (1998). A detailed reviews of models for recurrence data can be found in Cook and Lawless (2007) and Finkelstein (2008) .

For the frailty models without informative censoring, there are two estimation methods commonly used in practice. The first approach is the EM algorithm which treats the unobserved frailty terms as missing values. In some papers such as Vaida et al. (2000) and Ripatti et al. (2002), a Monte Carlo EM algorithm for a non-parametric baseline hazard and normal random effects model is used to estimate the hazard function and model parameters.

The second approach is the partial penalized likelihood (PPL), which was first proposed by McGilchrist and Aisbett (1991) . They treated the density of the frailty as a penalty term in the likelihood. Ripatti and Palmgren (2000) extended the PPL to estimate the parameters of multivariate normal random effects models. Duchateau and Janssen (2008) and Therneau and Grambsch (2000) showed that the solution from the PPL method is equivalent to the EM for the gamma frailty model for any fixed value of the frailty variance. Ha et al. (2001) implemented alternative hierarchical-likelihood procedures for frailty models (Ha et al. 2002, 2011). The lognormal or gamma distribution can be adapted as the frailty distribution, corresponding to the normal or log-gamma distribution for the random effects.

In the presence of informative censoring, the MCEM algorithm is usually an option to estimate the "joint frailty models" parameters even though it is more complicated (Liu et al. 2004, Liu and Huang 2008, Huang and Liu 2007, Huang and Wolfe 2002). Rondeau et al. (2007) have proposed a semiparametric joint frailty models with cubic splines for the baseline hazard by using PPL. Gaussian quadrature is another choice for estimation in simple frailty models with a piecewise constant hazard function. The resulting likelihood can be conveniently estimated by using current software tools (Liu and Huang 2008).

The copula function is extensively used for bivariate survival data (see e.g., Duchateau and Janssen 2008). It is regularly claimed that there is equivalence between a copula model and a particular frailty model, but they demonstrate that this claim is inaccurate using as an example the Clayton copula and the shared gamma frailty model.

For the bivariate and multi-type recurrent events there are a limited number of publications. Particularly, bivariate parametric models for the baseline hazard and frailty have been specified (Abu-Libdeh et al. 1990). A nonparametric random effect bivariate model with parametric pairwise baseline hazard has been introduced by Moreno (2008). Chen et al. (2004) and Chen and Cook (2009) have introduced a mixed Poisson model with lognormal random effects and piecewise constant parametric intensity functions. They use the existing software of the Poisson regression for estimation of the model parameters in the M-step.

A bivariate mixed Poisson model using a copula function for positive correlated gamma frailties has been described by Cook et al. (2010). In the M-step, the parameters are estimated by using existing software for the generalized linear models and survival analysis techniques for the parametric baseline model and semiparametric analysis, respectively. In the E-step the conditional expectations is approximated by a numerical integration method which might be convenient to use in this case for the specific properties of the Clayton copula and gamma density function. This solution is feasible, however, only for low-dimensional random effect vectors. A bootstrap procedure is used to obtain the standard errors of the parameter estimates.

3.1.3 Overview

In this chapter, we generalize the single-type event models with multivariate independent random effects (covariance matrix is diagonal) developed by Vaida et al. (2000) and Ripatti et al. (2002) to model the random baseline intensity and treatment effects. These two random effects are assumed to be independent. However, independence may be unrealistic for multi-type recurrent events. Here, we develop multi-type event models with correlated frailty terms or random effects. We present a different approach to model both multivariate correlated frailties and random effects using copula functions with gamma and Gaussian marginals instead of the normal distribution in the single-type case. Copula function allow us to avoid the restrictions for the random effect and frailty distributions. An MCEM algorithm is adapted to fit models using MCMC routines in the E-step. we provide estimates of fixed effects, and variance components as well as evaluate their standard errors using the Louis' formula (Louis 1982). The cumulative intensity functions are completely nonparametric and are estimated for each type. We consider predictions for the random effects, frailty terms, and investigate the heterogeneity between subjects.

The rest of this chapter is organized as follows. In Section 3.1 we describe the data used to build the proposed model as well as introduce notations and present the intensity functions.

The Gaussian and Clayton copula functions with gamma marginals are briefly introduced in Section 3.3. Section 3.4 covers the estimation procedures and the MCEM algorithms. Section 3.5 includes an evaluation of the variance for the parameter estimates along with a derivation of its forms and predictions of the frailty terms. Simulation studies are conducted to evaluate the performance of the proposed methods in Section 3.6. We illustrate the introduced methods using the skin cancer data in Section 3.7. Finally, we discuss the conclusions in Section 3.8.

3.2 The Model

3.2.1 Notation

We consider a study involving N independent subjects or individuals, each with a $p \times 1$ vector of covariates $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$. Subject i is observed over time period $[0, \tau_i]$, where time is measured from a defined starting point for that subject. Individuals experience repeated "events" or "failures" each of which can be any of J different types. Suppose that K_{ij} number of events of type j $(j = 1, \ldots, J)$ are observed on individual i $(i = 1, \ldots, N)$ at times $y_{ij1} < \ldots < y_{ijK_{ij}}$. For event type j in subject i we observe $t_{ijk_{ij}} = \min\left(y_{ijk_{ij}}, C_i\right)$, where C_i is a random censoring time independent of $y_{ijk_{ij}}$ $(k_{ij} = 1, \ldots, K_{ij})$. Additionally, a censoring indicator indicator $\delta_{ijk_{ij}}$ is observed, with $\delta_{ijk_{ij}}$ equal to 1 if $t_{ijk_{ij}} = y_{ijk_{ij}}$ and $\delta_{ijk_{ij}} = 0$ otherwise. We let $\boldsymbol{\delta}_{ij} = \left(\delta_{ij1}, \ldots, \delta_{ijK_{ij}}\right)'$ denote the corresponding $K_{ij} \times 1$ indicator vector. The total number of occurrence of type j for subject i is $K_{ij} - 1$. In summary, we observe $\mathbf{d}_{ijk_{ij}} = (\boldsymbol{x}_i', t_{ijk_{ij}}, \delta_{ijk_{ij}})'$. Let $\mathbf{D}_{ij} = \left(\mathbf{d}_{ij1}, \ldots, \mathbf{d}_{ijK_{ij}}\right)$ denote the observed $K_{ij} \times (p + J)$ data matrix of event type j for subject i and $\mathbf{D}_i = (\mathbf{D}_{i1}', \ldots, \mathbf{D}_{iJ}')'$ is used to represent the observed data for the subject i over all J recurrent event types. Let $\mathbf{D}$ denote the data for all subjects in the dataset.

3.2.2 Multi-type Intensity Model

In this chapter we present a multivariate frailty model for the intensity function $\lambda_{ij}(\cdot)$ as

$$\lambda_{ij}(t_{ijk_{ij}}|w_{ij}) = \lambda_{0j}(t_{ijk_{ij}})\, w_{ij}\, \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j), \tag{3.1}$$

where $\lambda_{0j}(\cdot)$ is a nonparametric baseline intensity function for the event type j. Here, $\boldsymbol{\beta}_j$=$(\beta_1, \ldots, \beta_p)'$ is a vector of the fixed effects associated with a vector of covariates $\boldsymbol{x}_i$ for event type j and $\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)$ is the corresponding relative risk. We denote the subject-specific frailty term for the j^{th} event type by w_{ij} and the subject-specific multivariate frailty vector by $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{iJ})'$. We assume that the event type frailties within subject i are correlated, and $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$ are independent. We also use the baseline cumulative intensity function defined as $\Lambda_{0j}(t) = \int_0^t \lambda_{0j}(s)\, ds$. Consequently, the cumulative intensity function can be presented by $\Lambda_{ij}(t) = \Lambda_{0ij}(t)\, w_{ij}\, \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)$.

The intensity function in (3.1) can be rewritten as

$$\lambda_{ij}(t_{ijk_{ij}}|w_{ij}) = \lambda_{0j}(t_{ijk_{ij}})\, \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j + b_{ij}), \tag{3.2}$$

where $b_{ij} = \log(w_{ij})$ is considered to be the random effect for the j^{th} recurrent event type of the i^{th} subject. The baseline intensity function $\lambda_{0j}(t_{ijk_{ij}})$ is either assumed to be completely unspecified (semiparametric model), follow a distribution depending on a low-dimensional unknown parameter vector (e.g., Weibull, Gompertz), or be modeled by a piecewise constant baseline intensity function. In this chapter, we focus mainly on semiparametric frailty models, which means that traditional maximum likelihood estimations procedures are not appropriate for parameter estimation. The expectation-maximization (EM) algorithm, the penalized partial likelihood (PPL) approach and Markov chain Monte Carlo (MCMC) methods can be used to fit semiparametric frailty models. These estimation techniques are discussed in detail for the single-type models in Therneau and Grambsch (2000) and Duchateau and Janssen (2008).

3.3 Multivariate Frailty Models

We use copula functions to model the association between the multivariate frailty. Copula modeling is an approach to formulating different multivariate distributions. The idea is that a simple transformation can be made of each marginal variable in such a way that each transformed marginal variable has a uniform $(0, 1)$ distribution. Once this is done, the dependence structure can be expressed as a multivariate distribution on the obtained uniform random variables. A copula is precisely a multivariate distribution on marginally uniform random variables. A copula model is often constructed from marginal survival functions or marginal distribution functions $F_j(\cdot)$. The copula function $C\left[F_1(w_{i1}), \ldots, F_J(w_{iJ})\right]$ is a joint distribution function such that

$$F\left(w_{i1}, \ldots, w_{iJ}\right) = C\left[F_1(w_{i1}), \ldots, F_J(w_{iJ})\right],$$

with marginal distribution functions $F_j(\cdot)$. For simplicity, we denote $F_j(w_{ij})$ by u_{ij}. Consequently, the copula density is defined by

$$\mathrm{c}\left(u_{i1}, \ldots, u_{iJ}\right) = \frac{\partial^J}{\partial u_{i1}, \ldots, \partial u_{iJ}} C\left(u_{i1}, \ldots, u_{iJ}\right). \tag{3.3}$$

The multivariate density function of frailty terms $\boldsymbol{w}_i$ is equal to

$$g(w_{i1}, \ldots, w_{iJ}) = c\left(u_{i1}, \ldots, u_{iJ}\right) \prod_{j=1}^{J} g_j\left(w_{ij}\right). \tag{3.4}$$

The result in (3.4) shows that it is always possible to specify a multivariate density $g(w_{i1}, \ldots, w_{iJ})$ by specifying the marginal densities $g_j\left(w_{ij}\right)$ and a copula density function $c\left(\cdot\right)$.

All information concerning dependence between the marginals is contained in the association parameter. The wo most frequently used copula families are the elliptical and Archimedean copulas which can be conveniently used for modeling multivariate frailty models with marginal distributions such as gamma, lognormal distributions, or even mixtures of

different types of frailty distributions. Details for various frailty distributions can be found in Duchateau and Janssen (2008). A range of distributions are supported and several types of variance structures for the random effects or frailty terms can be fitted. In this chapter, we present the Gaussian and Clayton copula functions to fit models for the frailty terms.

3.3.1 Elliptical Copula

Gaussian and the Student t-copula are the most commonly used elliptical copulas. The elliptical copula can be written as

$$C\left(u_{i1},...,\,u_{iJ}\right)=F\left[F_1^{-1}(u_{i1}),...,\,F_J^{-1}(u_{iJ})\right], \tag{3.5}$$

where $F\left(\cdot\right)$ is the multivariate cumulative distribution function (cdf) for an elliptical distribution, and $F_j^{-1}\left(\cdot\right)$ is the quantile function of the marginal distribution j.

Here, we use the Gaussian copula (Song 2000) to model the multivariate frailty. The multivariate Gaussian copula is given by

$$C\left(u_{i1},...,\,u_{iJ}\right)=\Phi_J\left[\Phi^{-1}(u_{i1}),...,\,\Phi^{-1}(u_{iJ})\right],$$

where $\Phi_J(\cdot)$ and $\Phi\ (\cdot)$ are the respective cdf of a multivariate normal MVN($\mathbf{0}$, $\boldsymbol{R}_J$) with a correlation matrix $\boldsymbol{R}_J$ and standard univariate normal $\mathrm{N}(0,1)$ marginals. The density of the normal copula is given by

$$c\left(u_{i1},\ldots,\,u_{iJ}\right) \;=\; \mid \boldsymbol{R}_J \mid^{-\frac{1}{2}}\, \exp\left[-\frac{1}{2}\,\boldsymbol{q}_i'\left(\boldsymbol{R}_J^{-1}-\boldsymbol{I}_J\right)\boldsymbol{q}_i\right], \tag{3.6}$$

where matrix $\boldsymbol{R}_J$ with unique elements equal to $J(J-1)/2$ is the correlation matrix of $\boldsymbol{u}_i=(u_{i1},\ldots,\,u_{iJ})'$ and therefore among the elements of the vector $\boldsymbol{w}_i=(w_{i1},\ldots,\,w_{iJ})'$ as well. $\boldsymbol{q}_i=(q_{i1},\ldots,\,q_{iJ})'$ is a vector of normal scores $q_{ij}=\Phi^{-1}(u_{ij})$, and $\boldsymbol{I}_J$ is the J-dimensional identity matrix. Obviously, $\boldsymbol{R}_J=\boldsymbol{I}_J$ implies independence of the J components.

The multivariate density function for the subject-specific frailty $\boldsymbol{w}_i$ can be then obtained as

$$g(w_{i1}, \ldots, w_{iJ}) \;=\; |\, \boldsymbol{R}_J \,|^{-\frac{1}{2}} \, \exp\left[-\frac{1}{2}\, \boldsymbol{q}_i' \left(\boldsymbol{R}_J^{-1} - \boldsymbol{I}_J\right) \boldsymbol{q}_i\right] \prod_{j=1}^{J} g_j\left(w_{ij}\right). \tag{3.7}$$

For the Gaussian copula, the mosty commonly used correlation structures are exchangeable (ex), auto regressive of order 1 (ar1), Toeplitz (toep), and unstructured (un). For example, in the case of dimension $J = 3$, the corresponding correlation matrices are

$$\begin{pmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_1 \\ \rho_1 & \rho_1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 & \rho_1^2 \\ \rho_1 & 1 & \rho_1 \\ \rho_1^2 & \rho_1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{pmatrix}, \text{ and } \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix}, \text{ for ex, ar1, toep, and un, respectively.}$$

3.3.2 Archimedean Copula

The J-dimensional Archimedean copula is the copula of J uniform $(0, 1)$ random variables. The Archimedean copula with a strict generator has the form

$$C\left(u_{i1}, ..., u_{iJ}\right) = \psi^{-1}\left[\psi\left(u_{i1}\right) + \ldots + \psi\left(u_{iJ}\right)\right],$$

where ψ^{-1} is the inverse of the generator ψ of the copula which satisfies the following assumptions

- ψ is a continuous, strictly decreasing, and convex function mapping $[0, 1]$ onto $[0, \infty]$,
- $\psi\left(0\right) = \infty$, and
- $\psi\left(1\right) = 0$.

In this chapter, we adopt the Clayton copula to model the subject-specific multivariate frailty. Clayton copula with $\psi\left(u_{ij}\right) = u_{ij}^{-\alpha} - 1$ and $\psi^{-1}\left(s\right) = \left(1 + s\right)^{-\frac{1}{\alpha}}$ has the following form

$$C\left(u_{i1}, ..., u_{iJ}\right) = \left(u_{i1}^{-\alpha} + \ldots + u_{iJ}^{-\alpha} - J + 1\right)^{-\frac{1}{\alpha}},\ \alpha \geq 0. \tag{3.8}$$

Here α is the copula parameter which controls the strength of dependence. When $\alpha = 0$ there is no dependence; when $\alpha = +\infty$ there is perfect dependence. Kendall's tau can be used as a measurement for the association by $\tau = {}^{\alpha}/_{(\alpha+2)}$, which takes values over the interval [0, 1]. In general, the multivariate density for J-dimensional frailty $\boldsymbol{w}_i$ is

$$g(w_{i1}, \ldots, w_{iJ}) = (-\alpha)^J \left[\prod_{j=0}^{J-1}(-\frac{1}{\alpha} - j)\right] \left[\left(\sum_{j=1}^{J} u_{ij}^{-\alpha} - J + 1\right)^{-1/\alpha - J}\right] \left[\prod_{j=1}^{J}\left(u_{ij}^{-\alpha-1}\right) g_j\left(w_{ij}\right)\right].$$

For example, when $J = 2$, the joint bivariate density function can be obtained as

$$g(w_{i1},\, w_{i2}) = (1+\alpha)\left(u_{i1}^{-\alpha} + u_{i2}^{-\alpha} - 1\right)^{-1/\alpha - 2} u_{i1}^{-\alpha-1} u_{i2}^{-\alpha-1} \prod_{j=1}^{J} g_j\left(w_{ij}\right). \tag{3.9}$$

3.3.3 Marginal (Frailty) Distributions

A comprehensive set of statistical distributions such as the lognormal, gamma, log-t, inverse Gaussian, positive stable, and compound Poisson distributions can be used as different choices of frailty and specify the copula marginal distributions. In practice, the gamma and lognormal distribution are primarily used to model the frailty terms. In the present chapter we restrict our considerations to the gamma frailty model and the lognormal frailty model, respectively. For the gamma frailty model, we assume the frailty terms w_{ij} are distributed gamma $\left(\frac{1}{\alpha_j},\, \alpha_j\right)$ with variances α_j and have means equal to 1 to avoid the unidentifiable problem, which might otherwise occur if we multiply or divide the frailty term and the baseline intensity by the same constant. The probability density function is then

$$g_j(w_{ij}) = \frac{w_{ij}^{(1/\alpha_j - 1)} \exp\left(-w_{ij}/\alpha_j\right)}{\Gamma\left(1/\alpha_j\right) \alpha_j^{1/\alpha_j}}.$$

Another important frailty distribution is the lognormal distribution. The popularity of the lognormal frailty model is due to the link with mixed models, where the standard assumption is that the random effects follow a normal distribution. When $b_{ij} \sim \mathrm{N}(0,\, \sigma_j^2)$ is a normally dis-

tributed random effect and the frailty is given by $w_{ij} = \exp(b_{ij})$ the expectation and variance of the frailty are functions of σ_j^2 with $\mathrm{E}(w_{ij}) = \exp(\sigma_j^2/2)$ and $\mathrm{Var}(w_{ij}) = \exp(\sigma_j^2)\exp(\sigma_j^2 - 1)$. Consequently, in the gamma frailty model the parameter α_j denotes the variance of the frailty w_{ij} whereas in the lognormal model σ_j^2 denotes the variance of the random effect $b_{ij} = \log(w_{ij})$. Both expressions cannot be directly compared. Furthermore, in the lognormal model the expectation of the frailty variable is usually not equal to one despite the fact that the expectation of the random effect b_{ij} is zero. The frailty lognormal model has the form shown in (3.2). It is important to note that the subject-specific frailty terms $\boldsymbol{w}_i$ are mutually independent.

3.4 The Estimation Method

We use the complete data $\mathbf{D}_i$ and frailty terms $\boldsymbol{w}_i$ for the subject i. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_J')'$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J, \alpha)'$ to be the parameter vector for the multivariate density function $g(w_{i1}, \ldots, w_{iJ})$ of the frailty terms $\boldsymbol{w}_i$. Define $\boldsymbol{\lambda}_{0j}(\cdot)$ as the unspecified baseline intensity functions with $\boldsymbol{\lambda}_0(\cdot) = [\boldsymbol{\lambda}_{01}'(\cdot), \ldots, \boldsymbol{\lambda}_{0J}'(\cdot)]$. The ordered distinct event times are denoted by $\boldsymbol{t}_j = \left[t_{j(1)}, \ldots, t_{j(K_j)}\right]'$ where K_j is the number of event type j for all subjects at the ordered distinct time $t_{j(K_j)}$. $\boldsymbol{\lambda}_{0j}(\cdot) = \left\{\lambda_{0j}\left[t_{j(1)}\right], \ldots, \lambda_{0j}\left[t_{j(K_j)}\right]\right\}'$ are the corresponding baseline intensity functions. Consequently, The vector of unknown parameters in the model to estimate is $\boldsymbol{\xi} = [\boldsymbol{\beta}', \boldsymbol{\lambda}_0'(\cdot), \boldsymbol{\alpha}']'$.

3.4.1 Log-likelihood Function

Given data and $\boldsymbol{w}_i$, we can write the i^{th} subject's contribution to the likelihood function as

$$L_i(\boldsymbol{\xi}, \boldsymbol{w}_i) = \prod_{j=1}^{J}\prod_{k=1}^{K_{ij}}\left[\lambda_j(t_{ijk_{ij}})\right]^{\delta_{ijk_{ij}}}\exp\left[-\Lambda_j(\tau_i)\right]g(\boldsymbol{w}_i|\boldsymbol{\alpha}). \tag{3.10}$$

The complete log-likelihood is $\mathcal{L}(\boldsymbol{\xi}) = \sum_{i=1}^{N}\log\left[L_i(\boldsymbol{\xi}, \boldsymbol{w}_i)\right]$. In particular,

$$\mathcal{L}(\boldsymbol{\xi}, \boldsymbol{W}) = \sum_{i=1}^{N}\sum_{j=1}^{J}\left\{\sum_{k=1}^{K_{ij}}\delta_{ijk_{ij}}\left[\log\ \lambda_{0j}\left(t_{ijk_{ij}}\right) + \log\left(w_{ij}\right) + \boldsymbol{x}_i'\boldsymbol{\beta}_j\right] - \Lambda_{0j}(\tau_i)w_{ij}\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j\right)\right\} + \sum_{i=1}^{N}\log\left[g\left(\boldsymbol{w}_i|\boldsymbol{\alpha}\right)\right], \quad (3.11)$$

where $\boldsymbol{W} = (\boldsymbol{w}_i', \ldots, \boldsymbol{w}_N')'$ is the frailty vector all over $\boldsymbol{w}_i$. The marginal complete likelihood of the observed data over all subjects ($i = 1, \ldots, N$) is

$$L(\boldsymbol{\xi}) = \prod_{i=1}^{N}\int_{-\infty}^{\infty} L_i(\boldsymbol{\xi}, \boldsymbol{w}_i)\, d\boldsymbol{w}_i. \quad (3.12)$$

In most situations, this integration is not in a closed form. The likelihood function in (3.12) has two problems to be used in the inference of $\boldsymbol{\xi}$. First, it depends on the high dimensional nonparametric baseline intensity function. Second, it is usually has a multi-dimensional integration.

This feature makes the EM algorithm a good methodology to fit the model. The use of the EM algorithm to estimate the parameters of model has been proposed for a single-type model in the presence or absence of the terminal event (Vaida et al. 2000, Liu et al. 2004, Huang and Liu 2007, Huang and Wolfe 2002). In the following section we adopt an MCEM algorithm with copula multivariate frailty for the multi-type model in (3.1).

3.4.2 Mote Carlo EM Algorithm

The EM algorithm is a general approach to iterative computation of maximum-likelihood estimates when the observations can be viewed as incomplete data (Dempster et al. 1977). Monte Carlo EM (MCEM) algorithm is a modification of the EM algorithm where the expectation in the E-step is computed numerically through Monte Carlo simulations (Johansen 1983, Wei and Tanner 1990). The most flexible and generally applicable approach to obtaining a Monte Carlo sample in each iteration of an MCEM algorithm is through Markov

chain Monte Carlo (MCMC) methods such as the Metropolis-Hastings and Gibbs samplers (Robert and Casella 2009, McLachlan and Krishnan 2007). MCMC estimation presents a useful solution to problems where the E-step is not available in closed forms. The MCEM algorithm includes two steps: computation of particular conditional expectations for the log of the complete likelihood (E-step) and maximization this expectation with respect to all parameter components (M-step). Starting from initial values of parameters, the algorithm iterates between the two steps until convergence is achieved. It is important to notice that, under regularity conditions, the algorithm is guaranteed to converge to a stationary point.

3.4.2.1 E-step

In the E-step the conditional density of $\boldsymbol{w}_i$ conditional on observed data is

$$g_{\boldsymbol{w}_i|\mathbf{D}_i}\left(\boldsymbol{w}_i|\boldsymbol{\xi}\right) = g(\boldsymbol{w}_i|\boldsymbol{\xi}) = \frac{f(\boldsymbol{D}_i, \boldsymbol{w}_i)}{\int_{-\infty}^{\infty} f(\boldsymbol{D}_i, \boldsymbol{w}_i) d\boldsymbol{w}_i} = \frac{L_i\left(\boldsymbol{\xi}, \boldsymbol{w}_i\right)}{L_i\left(\boldsymbol{\xi}\right)}, \tag{3.13}$$

where $f(\boldsymbol{D}_i, \boldsymbol{w}_i)$ is the joint density of the complete data and frailty, $L_i\left(\boldsymbol{\xi}, \boldsymbol{w}_i\right)$ is defined in (3.10), and $L_i\left(\boldsymbol{\xi}\right)$ is the marginal likelihood for the i^{th} subject. In the E-step, since there is no closed form for the density of $g_{\boldsymbol{w}_i|\mathbf{D}_i}\left(\boldsymbol{w}_i|\boldsymbol{\xi}\right)$, Metropolis-Hastings algorithm can be used to generate M random samples of $\boldsymbol{w}_i$ with the conditional distribution in (3.13). A brief introduction to the Metropolis-Hastings algorithm is given in Appendix 3.A.1**.** For each subject i and after burn in samples, we generate random samples $\boldsymbol{w}_i^{(m)}$, $m = 1, \ldots, M$. Then the expectation of functions of $\boldsymbol{w}_i$ conditional on the observed data are estimated by the sample means for M'samples after the burn-in and thining. For example,

$$\mathrm{E}(w_{ij}) = \frac{1}{M'} \sum_{m=1}^{M'} w_{ij}^{(m)}, \text{ and } \mathrm{E}(\log w_{ij}) = \frac{1}{M'} \sum_{m=1}^{M'} \left[\log\left(w_{ij}^{(m)}\right)\right].$$

This is the procedure of the E-steps in the EM algorithm.

3.4.2.2 M-step

The EM algorithm requires $Q(\boldsymbol{\xi})$, which is the expectation of the log-likelihood (3.11) over $i = 1, \ldots, N$ conditional on the data $\mathbf{D}$ and the current parameter estimates. In particular, $Q(\boldsymbol{\xi})$ can be expressed as

$$Q(\boldsymbol{\xi}) = Q_1[\boldsymbol{\beta}, \boldsymbol{\lambda}_0(\cdot)] + Q_2(\boldsymbol{\alpha}),$$

where

$$Q_1(\boldsymbol{\beta}, \boldsymbol{\lambda}_0(\cdot)) = \sum_{i=1}^{N}\sum_{j=1}^{J}\left\{\sum_{k=1}^{K_{ij}}\delta_{ijk_{ij}}\left[\log\ \lambda_{0j}\left(t_{ijk_{ij}}\right) + \mathrm{E}(\log w_{ij}) + \left(\boldsymbol{x}_i'\boldsymbol{\beta}_j\right)\right]\right\} - \sum_{i=1}^{N}\sum_{j=1}^{J}\left[\Lambda_{0j}(\tau_i)\mathrm{E}(w_{ij})\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j\right)\right]. \quad (3.14)$$

The regression parameters are updated by the profile partial likelihood via arguments of Johansen (1983) for $\boldsymbol{\lambda}_0(\cdot)$. In particular, (see e.g., Klein and Moeschberger 2003, p. 258)

$$\mathcal{L}_{partial}(\boldsymbol{\beta}) = \sum_{j=1}^{J}\sum_{k_j=1}^{K_j}\left[\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j\right) + \mathrm{E}(\log w_{ij}) - \log\sum_{i\in R_{t_{j(k_j)}}}\mathrm{E}(w_{ij})\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j\right)\right].$$

The estimating equation for $\boldsymbol{\beta}_j$ is given as

$$\frac{\partial\mathcal{L}_{partial}}{\partial\boldsymbol{\beta}_j} = \sum_{k_j=1}^{K_j}\left[\boldsymbol{x}_i - \frac{\sum_{i\in R_{t_{j(k_j)}}}\boldsymbol{x}_i\mathrm{E}(w_{ij})\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j\right)N_j\left(t_{(k_j)}\right)}{\sum_{i\in R_{t_{j(k_j)}}}\mathrm{E}(w_{ij})\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j\right)}\right],$$

where $R_{t_{j(k_j)}}$ and $N_j\left[t_{(k_j)}\right]$ are the risk group and number of events of type j at the specific ordered time $t_{(k_j)}$, respectively. Setting theses equations equal to zero and solving them yields the update parameter estimates for $\boldsymbol{\beta}_j$. The Nelson-Aalen formula to estimate the baseline intensity $\lambda_{0j}(\cdot)$ is

$$\lambda_{0j}(t_{k_j}) = \frac{N_j\left[t_{(k_j)}\right]}{\sum\limits_{i \in R_{t_{j(k_j)}}} \mathrm{E}\left(w_{ij}\right) \exp\left(\boldsymbol{x}_i' \boldsymbol{\beta}_j\right)}. \tag{3.15}$$

Consequently, the cumulative intensity functions $\Lambda_{0j}(\cdot)$ for the recurrent events can be updated by

$$\Lambda_{0j}(t_{k_j}) = \sum_{t_{(k_j)} \leq t} \lambda_{0j}(t_{(k_j)}).$$

In such situation the estimation of copula vector parameter $\boldsymbol{\alpha}$ can be obtained by maximizing of $Q_2\left(\boldsymbol{\alpha}\right)$ where

$$\begin{aligned} Q_2\left(\boldsymbol{\alpha}\right) &= \mathrm{E}\left\{\log\left[g(w_{i1}, \ldots, w_{iJ})\right]\right\} = \mathrm{E}\left\{\log\left[c\left(u_{i1}, \ldots, u_{iJ}\right) \prod_{j=1}^{J} g_j\left(w_{ij}\right)\right]\right\} \\ &= Q_2\left(\alpha\right) + \sum_{j=1}^{J} Q_2(\alpha_j), \end{aligned}$$

where $Q_2(\alpha)$ and $Q_2(\alpha_j)$ are the expectation of the log-likelihood of the copula and marginal distributions, respectively. The estimation of $\boldsymbol{\alpha}$ can be achieved by maximizing $Q_2\left(\boldsymbol{\alpha}\right)$. In particular, the forms of expected log-likelihood for Gaussian and Clayton Copula $Q_2\left(\boldsymbol{\alpha}\right)$ with gamma marginals and their score equations are derived in Appendix 3.A.2**.** The Metropolis-Hastings algorithm is adapted and outlined in Algorithm 3.1

3.4.3 Stopping Rule

We consider using a convergence criterion based on the change in the parameter estimates (Booth and Hobert 1999). Particularly, the algorithm is stopped if

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

where δ_2 and δ_2 are predetermined small values (e.g. $\delta_1 = 0.01$, $\delta_2 = 0.003$).

Algorithm: 3.1 Monte Carlo EM algorithm with MCMC in the E-step with copula frailty.

1. Initialize $\boldsymbol{\xi}^{(0)}$. At iteration $s+1$ $(s = 1, \ldots, S)$,
2. E-step:
   (a) Generate $\boldsymbol{W} = (\boldsymbol{w}_1', \ldots, \boldsymbol{w}_N')' \sim g_{\boldsymbol{w}_i|\mathbf{D}_i}(\boldsymbol{w}_i|\boldsymbol{\xi}^{(\mathbf{s})})$ via an MCMC algorithm.
   (b) Estimate the required conditional expectations $\mathrm{E}\left\{g_{\boldsymbol{w}_i|\mathbf{D}_i}\left[\boldsymbol{w}_i|\boldsymbol{\xi}^{(\mathbf{s})}\right]\right\}$ of the frailty terms .
3. M-step: Maximize the expected complete log-likelihood $Q(\boldsymbol{\xi})$ to obtain $\hat{\boldsymbol{\xi}}^{(s+1)}$.
4. Repeat Steps 2 and 3 until the convergence is declared.

3.5 Variance Estimation

Since the EM algorithm does not provide the information matrix for the observed data likelihood directly, Louis' formula (Louis 1982) is used. The observed information matrix $\boldsymbol{I}\left(\hat{\boldsymbol{\xi}}\right)$ is given by

$$\boldsymbol{I}\left(\hat{\boldsymbol{\xi}}\right) = -\mathrm{E}\left(\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'} | \boldsymbol{W}, \hat{\boldsymbol{\xi}}\right) - \mathrm{E}\left[\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\xi}}\right)\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\xi}'}\right) | \boldsymbol{W}, \hat{\boldsymbol{\xi}}\right] + \mathrm{E}\left[\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\xi}} | \boldsymbol{W}, \hat{\boldsymbol{\xi}}\right) \mathrm{E}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\xi}'} | \boldsymbol{W}, \hat{\boldsymbol{\xi}}\right)\right]. \tag{3.16}$$

All of these terms are evaluated at the last iteration of the EM algorithm, when the last term becomes zero for the MLE $\hat{\boldsymbol{\xi}}$. The first two expectations can be calculated by averaging over the corresponding terms involving Metropolis-Hastings. The components of $\boldsymbol{I}\left(\hat{\boldsymbol{\xi}}\right)$ are given in Appendix 3.A.3. The inverse of $\boldsymbol{I}\left(\hat{\boldsymbol{\xi}}\right)$ gives the covariance matrix for $\hat{\boldsymbol{\xi}}$. The bootstrap re-sampling method has been used to obtain it (see e.g., Cook et al. 2010).

3.6 Simulation Studies

In this section, we perform separate simulation studies for models in Section 3.2.2 to assess finite sample properties of the proposed algorithm in Section 3.4. We carry out simulation

studies for sample size 50 and 100 subjects. The baseline hazards were assumed constant, with $\lambda_{0j}(t) = 1$. The effects of covariate x_{ij} were assumed to be equal, that is $\beta_1 = 1$, and $\beta_2 = 1$. The variances associated with the frailty terms w_{i1} and w_{i2} were also assumed to be equal, $\alpha_1 = \alpha_2 = 1$. The large value of α_j corresponds to a strong correlation between the recurrent event times of same type while zero variance means that the recurrent times are independent for the same subject. Three values 0.1, 1.333, and 8 for the copula parameter α were considered. This is equivalent to assuming the correlation coefficient ρ to be zero, moderate (0.4), and high (0.8). Between 29% and 33% of observations for each of the two types were censored. The average number of observed recurrent events per subject ranges from 0.40 to 1.00 for both event types for the placebo and treatment, respectively. For each setting of the parameters, 300 simulated datasets were generated.

3.6.1 Data Generation

1. For each subject i $(i = 1, \ldots, N)$, we generated $\mathbf{b}_i$ bivariate random effects or bivariate frailty terms $\mathbf{w}_i$ from the Gaussian or Clayton Copula with a copula parameter α and Gaussian marginals $b_{ij} \sim \mathrm{N}(0, \alpha_j)$ and Gamma marginals $w_{ij} \sim \Gamma\left(\frac{1}{\alpha_j}, \alpha_j\right)$.

2. The treatment effect variable x_i is generated from a Bernoulli distribution with parameter $(p = 0.5)$.

3. The fixed right-censored time was set at, $C = 1$, $i = 1, \cdots, N$.

4. C_i is the censored time for subject i^{th} where $C_i = \min\{C_i^*, C\}$ and the random censoring time C_i^* is assumed to be exponentially distributed with rate $\alpha_c = 0.5$.

5. For every event type $(j = 1, 2)$, generate gap times z_{ijk} from the exponential distribution with $\lambda_{0j}(t) = 1$ and the rate parameter $\alpha_j = \left[\lambda_{0j}(t)\, w_{ij} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\right]^{-1}$ and $\left[\lambda_{0j}(t)\, w_{ij} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j + b_{ij})\right]^{-1}$for the models in 3.3 and 3.6, respectively.

6. let $y_{ijk_{ij}} = \sum_{k_{ij}=1}^{K_{ij}} z_{ijk_{ij}} (k_{ij} = 1, \ldots, K_{ij})$ the observed event times $t_{ijk_{ij}} = \min(C_i, y_{ijk_{ij}})$ with the first start time set to zero.

7. Finally, the censoring indicator $\delta_{ijk_{ij}}$ was set at 1 if $t_{ijk_{ij}} = y_{ijk}$ and $\delta_{ijk} = 0$ otherwise.

3.6.2 Simulation Results

In the simulation studies, we assume both Gaussian random effects and gamma frailties with Gaussian and Clayton copulas. The results are listed in the corresponding subsections as follows. Gamma frailties and Clayton Copula, gamma frailties and Gaussian Copula, Gaussian random effects and Clayton Copula, and Gaussian random effects and Gaussian Copula are assumed in Sections 3.6.2.1, 3.6.2.2, 3.6.2.3, and 3.6.2.4, respectively. The summary of statistics of the estimated regression coefficients, marginal frailty or random effects distribution, and the copula parameter(s) are reported in Tables 3.1, 3.2, 3.3, and 3.4. In the simulation results, the bias is the mean of the parameter estimates (based on 300 replicates) minus the true value and Var is the sampling variance of the parameter estimates. The MSE denotes to the empirical mean square errors of the the corresponding parameter over all 300 replicates. We also compute the empirical coverage probabilities (CP) of the estimates of corresponding 95% confidence interval.

First, it can be seen that, in the three settings of the Clayton copula and gamma frailty model in Table 3.1, parameter estimates of our proposed models are well estimated. The magnitudes of the empirical biases of the estimates are negligible. Small biases (around 10%) are observed for the copula parameter α in settings I, II, and III. The CPs are close to the nominal level 95%. We observe only minor biases, regarding the coverage probabilities, for the copula parameter α with some improvements with the increasing in the sample size. It can be seen that as the value of copula parameter increases, MSE of the estimates decrease. As the sample size N increases, both bias and MSE decrease over all parameter estimates.

We constucted additional scenarios with Gaussian copula and gamma frailty in Section 3.6.2.2, Clayton copula and Gaussian random effects in Section 3.6.2.3, and Gaussian copula and Gaussian random effects in Section 3.6.2.4. However, we reached the same conclusion for these three settings with negligible bias for the parameter estimate of the copula parameter

and coverage probability of the Gaussian copula.

3.6.2.1 Gamma Frailties and Clayton Copula

Table 3.1 gives the simulation results for the gamma frailty and Clayton copula model.

Table 3.1 – Empirical results from simulation studies examining the frequency properties of estimates of treatment effects, variances, and copula parameter for the gamma frailty and Clayton copula model with $\alpha_1 = \alpha_2 = 1$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{0j}(t) = 1$, and 300 repeats for every setting.

Param.	Value	Mean		Bias		Var		MSE		CP	
# of subjects		50	100	50	100	50	100	50	100	50	100
					Setting I						
β_1	1	1.053	1.051	0.0523	0.051	0.095	0.074	0.097	0.077	0.952	0.954
β_1	1	1.044	1.044	0.044	0.044	0.094	0.092	0.097	0.094	0.930	0.942
α_1	1	0.936	1.005	-0.063	0.005	0.063	0.033	0.067	0.033	0.960	0.962
α_2	1	0.957	0.979	-0.042	-0.021	0.058	0.020	0.060	0.021	0.940	0.952
α	0.1	0.279	0.158	0.179	0.058	0.091	0.027	0.123	0.030	0.938	0.936
					Setting II						
β_1	1	1.074	1.068	0.074	0.068	0.093	0.074	0.099	0.078	0.951	0.964
β_1	1	1.073	1.065	0.073	0.065	0.090	0.093	0.095	0.097	0.936	0.943
α_1	1	0.972	1.018	-0.028	0.018	0.071	0.044	0.072	0.044	0.933	0.949
α_2	1	1.006	1.007	0.006	0.007	0.076	0.037	0.076	0.037	0.950	0.946
α	1.333	1.774	1.512	0.441	0.179	0.573	0.099	0.768	0.131	0.940	0.942
					Setting III						
β_1	1	1.1082	1.084	0.108	0.084	0.062	0.077	0.074	0.084	0.945	0.949
β_1	1	1.088	1.101	0.088	0.101	0.051	0.101	0.0591	0.112	0.935	0.974
α_1	1	0.992	1.035	-0.008	0.035	0.071	0.050	0.071	0.051	0.950	0.951
α_2	1	0.999	1.028	-0.0001	0.028	0.079	0.061	0.079	0.061	0.961	0.943
α	8	6.790	7.018	-1.210	-0.982	2.002	1.402	2.466	2.368	0.930	0.933

3.6.2.2 Gamma Frailties and Gaussian Copula

Results for the gamma frailty and Gaussian Copula are presented in Table 3.2.

Table 3.2 – Empirical results from simulation studies examining the frequency properties of estimates of treatment effects, variances, and copula parameter for the gamma frailty and Gaussian copula model with $\alpha_1 = \alpha_2 = 1$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{0j}(t) = 1$, and 300 repeats for every setting.

Param.	Value	Mean		Bias		Var		MSE		CP	
# of subjects		50	100	50	100	50	100	50	100	50	100
					Setting I						
β_1	1	0.991	1.070	-0.009	0.070	0.097	0.010	0.097	0.015	0.954	0.956
β_2	1	1.141	1.071	0.141	0.071	0.091	0.008	0.111	0.013	0.962	0.952
α_1	1	0.915	0.900	-0.085	-0.100	0.015	0.000	0.022	0.010	0.942	0.941
α_2	1	0.929	1.065	-0.071	0.065	0.049	0.024	0.055	0.028	0.940	0.948
α	0.1	0.134	0.155	0.034	0.055	0.055	0.007	0.056	0.010	0.920	0.934
					Setting II						
β_1	1	1.065	1.053	0.065	0.053	0.104	0.040	0.099	0.043	0.970	0.976
β_2	1	1.070	1.071	0.070	0.071	0.088	0.039	0.093	0.044	0.960	0.962
α_1	1	0.841	0.878	-0.159	-0.122	0.012	0.001	0.037	0.016	0.942	0.974
α_2	1	0.937	1.054	-0.063	0.054	0.068	0.069	0.072	0.072	0.950	0.963
α	0.4	0.429	0.402	0.029	0.002	0.029	0.015	0.030	0.015	0.934	0.935
					Setting III						
β_1	1	1.076	1.062	0.076	0.062	0.091	0.038	0.097	0.042	0.948	0.950
β_2	1	1.052	1.064	0.052	0.064	0.072	0.053	0.074	0.057	0.952	0.962
α_1	1	0.958	0.945	-0.042	-0.055	0.023	0.005	0.024	0.008	0.944	0.962
α_2	1	1.021	1.006	0.021	0.006	0.026	0.073	0.026	0.073	0.940	0.952
α	0.8	0.0826	0.806	0.026	0.006	0.013	0.003	0.013	0.003	0.922	0.932

3.6.2.3 Gaussian Random Effects and Clayton Copula

Summary of simulation studies for Gaussian random effects and Clayton copula is shown in Table 3.3.

Table 3.3 – Empirical results from simulation studies examining the frequency properties of estimates of treatment effects, variances, and copula parameter for the Gaussian random effects and Clayton copula model with $\alpha_1 = \alpha_2 = 1$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{0j}(t) = 1$, and 300 repeats for every setting.

Param.	Value	Mean		Bias		Var		MSE		CP	
# of subjects		50	100	50	100	50	100	50	100	50	100
					Setting I						
β_1	1	1.075	1.014	0.075	0.014	0.103	0.035	0.109	0.035	0.942	0.944
β_2	1	1.013	1.037	0.013	0.037	0.080	0.056	0.080	0.058	0.950	0.958
α_1	1	0.938	0.971	-0.062	-0.029	0.041	0.012	0.045	0.013	0.960	0.958
α_2	1	0.915	0.960	-0.085	-0.040	0.039	0.014	0.046	0.015	0.942	0.944
α	0.1	0.013	0.082	-0.087	0.082	0.000	0.106	0.008	0.113	0.926	0.928
					Setting II						
β_1	1	1.053	1.059	0.053	0.059	0.023	0.037	0.026	0.040	0.932	0.948
β_2	1	1.011	1.069	0.011	0.069	0.080	0.027	0.080	0.032	0.949	0.958
α_1	1	0.935	0.962	-0.065	-0.038	0.041	0.012	0.045	0.014	0.940	0.954
α_2	1	0.928	0.950	-0.072	-0.050	0.039	0.011	0.044	0.013	0.966	0.968
α	1.333	0.014	0.409	-1.319	0.009	0.000	0.104	1.740	0.104	0.912	0.922
					Setting III						
β_1	1	1.062	1.052	0.062	0.052	0.115	0.085	0.119	0.088	0.932	0.944
β_2	1	1.103	1.093	0.103	0.093	0.074	0.044	0.084	0.052	0.944	0.948
α_1	1	0.951	0.942	-0.049	-0.058	0.042	0.032	0.044	0.035	0.946	0.956
α_2	1	0.963	0.947	-0.044	-0.053	0.042	0.030	0.044	0.033	0.956	0.960
α	8	6.641	6.941	-1.359	-1.059	0.272	0.072	2.118	1.193	0.920	0.922

3.6.2.4 Gaussian Random Effects and Gaussian Copula

Results of Gaussian random effects and Gaussian copula are summarized in Table 3.4.

Table 3.4 – Empirical results from simulation studies examining the frequency properties of estimates of treatment effects, variances, and copula parameter for the Gaussian random effects and Gaussian copula model with $\alpha_1 = \alpha_2 = 1$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{0j}(t) = 1$, and 300 repeats for every setting.

Param.	Value	Mean		Bias		Var		MSE		CP	
# of subjects		50	100	50	100	50	100	50	100	50	100
					Setting I						
β_1	1	1.034	1.041	0.034	0.041	0.028	0.069	0.029	0.070	0.943	0.967
β_2	1	0.957	0.967	-0.043	-0.033	0.038	0.053	0.040	0.054	0.946	0.954
α_1	1	0.973	0.984	-0.027	-0.016	0.027	0.014	0.028	0.014	0.934	0.973
α_2	1	0.973	0.967	-0.027	-0.033	0.035	0.019	0.035	0.020	0.953	0.957
α	0.1	0.124	0.095	0.024	-0.005	0.025	0.010	0.025	0.010	0930	0.948
					Setting II						
β_1	1	0.971	0.932	-0.029	-0.068	0.065	0.041	0.066	0.046	0.960	0.971
β_2	1	1.000	1.049	0.000	0.049	0.074	0.048	0.074	0.051	0.950	0.962
α_1	1	0.982	0.922	-0.018	-0.078	0.025	0.016	0.025	0.022	0.942	0.962
α_2	1	0.999	0.965	-0.001	-0.035	0.030	0.011	0.030	0.012	0.947	0.953
α	0.4	0.332	0.332	0.068	-0.016	0.042	0.026	0.047	0.026	0.935	0.938
					Setting III						
β_1	1	1.048	1.051	0.048	0.051	0.051	0.073	0.054	0.076	0.934	0.942
β_2	1	1.067	0.991	0.067	-0.009	0.033	0.064	0.037	0.064	0.942	0.958
α_1	1	0.946	0.976	-0.054	-0.024	0.023	0.014	0.026	0.015	0.945	0.946
α_2	1	0.967	0.946	-0.043	-0.054	0.031	0.018	0.033	0.021	0.951	0.956
α	0.8	0.863	0.809	0.063	0.009	0.012	0.005	0.016	0.005	0.920	0.937

In summary, the proposed estimates for the regression coefficient, marginal distribution parameters, and copula parameter(s) have a small bias. The confidence intervals constructed based on the empirical stranded error have appropriate coverage probabilities. In contrast, the estimates based on a separate model for the recurrent event type will result in larger biases and worse empirical coverage probabilities for all of the parameter estimates in the presence of dependence between the recurrent event types. See for example, Cook et al. (2010) and Mazroui et al. (2013). Although, we investigate settings with smaller number of subject, higher censoring rates, and smaller average number of events per subject than the studies of Cook et al. (2010) and Mazroui et al. (2013), our results are comparable with some improvements.

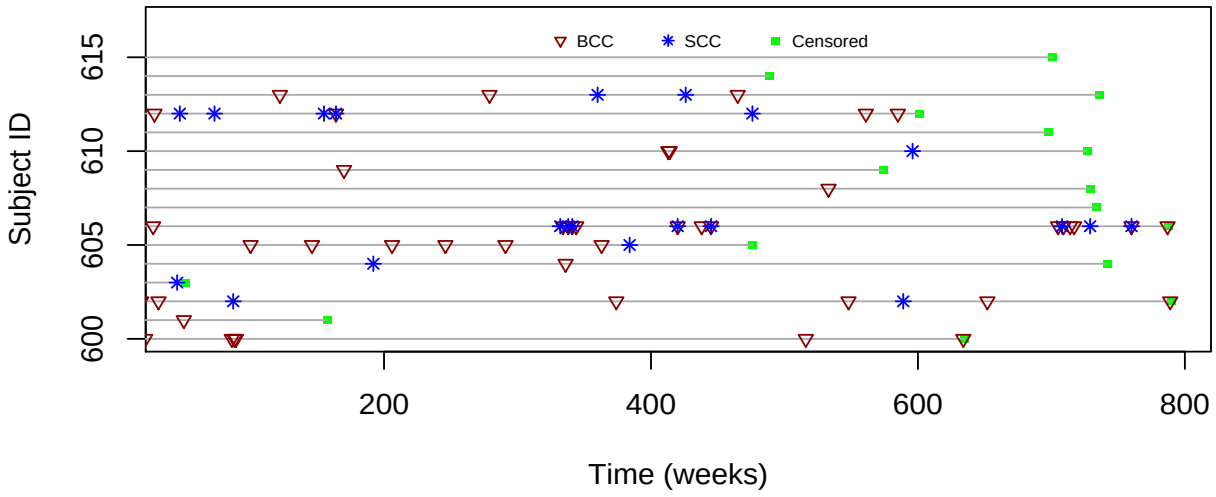


Figure 3.1 – The profile plot for subjects 600 to 615.

3.7 Application to a Skin Cancer Dataset

3.7.1 Summary of the Skin Cancer Dataset

The Nutritional Prevention of Cancer (NPC) study was a randomized, double-blinded, placebo-controlled clinical trial conducted by Arizona Cancer Center which started on 1983. The NPC was designed to evaluate the efficacy of Selenium as selenized yeast (200 mg daily), for up to five years, in preventing the recurrence of nonmelanoma skin cancer (NMSC) among 1312 residents of the Eastern United States. The study was originally designed to test the efficacy of Selenium supplementation in preventing NMSC recurrence in men (%75) and women (%25) with a history of two or more basal cell carcinoma (BCCs) or one squamous cell carcinoma (SCCs) of the skin. The hypothesis for this trial was supported by the observation that populations in the southeastern United States, a region with soil Selenium concentrations lower than those of the rest of the country, showed elevated NMSC rates. Thus, a randomized clinical trial of Selenium supplementation is initiated for preventing the recurrence of NMSC in this high-risk population. The length of the follow-up and the recurrent times of the BCC and SCC for sample of 15 patients are represented in Figure 3.1.

Sixty-two participants (including two cancer types in each treatment group) whose initial

Table 3.5 – Total cancer occurrences by treatment group.

	BCC or SCC		BCC		SCC	
No.events	Placebo No.(%)	Selenium No.(%)	Placebo No.(%)	Selenium No.(%)	Placebo No.(%)	Selenium No.(%)
Censored	152 (0.259)	116 (0.191)	179 (0.305)	157 (0.259)	414 (0.706)	404 (0.667)
1	105 (0.179)	125 (0.206)	120 (0.205)	133 (0.219)	96 (0.164)	120 (0.198)
2	71 (0.121)	81(0.134)	73 (0.125)	80 (0.132)	37 (0.063)	41 (0.068)
3	57 (0.097)	66 (0.109)	42 (0.072)	72 (0.119)	16 (0.027)	19 (0.031)
4	48 (0.082)	51 (0.084)	47 (0.080)	43 (0.071)	9 (0.015)	6 (0.010)
5	31 (0.053)	28 (0.046)	30 (0.007)	17 (0.008)	4 (0.007)	5 (0.008)
≥ 6	122(0.208)	139 (0.229)	95 (0.017)	104 (0.018)	10 (0.017)	11 (0.018)
Uncensored	434(0.741)	490(0.809)	407(0.695)	449(0.471)	172(0.294)	202 (0.333)
No.event.	1933(3.298)	2279(3.760)	1582(2.699)	1871(3.087)	351(0.599)	408(0.673)
Max.	28	34	22	33	18	28

blood draws were drawn 4 days after the randomization date and 58 patients were found to have incomplete follow-up data for the tumor types are excluded from the analysis. The statistical analysis is based on data from those 1192 participants with initial blood draws within 4 days of randomization and has complete follow-up data. The mean follow-up time was 6.4 years. The total number of both BCC and SCC cancer types were 1582 and 351 for the placebo group, and 1871 and 408 for the Selenium group, respectively. The average numbers of events per subject by the treatment group were between 2.699 and 3.087 for the BCC and from 0.599 to 0.673 for the SCC cancer types. The maximum numbers of BCC and SCC incidences were 33 and 28 in the Selenium group. The censoring rates were ranging from 0.259 to 0.667 for both the BCC and SCC types, respectively. Results were obtained from the total cohort of 1312 participants and of 1192 participants with valid baseline Selenium values (606 participants in the Selenium group and 589 participants in the placebo group). Selected characteristics of subjects by treatment group are displayed in Table 3.5 and Figures 3.2, 3.3, and 3.4.

A summary of the follow-up times according to the treatment group is given in Table 3.6. We can conclude there is no significant difference between the follow-up times in the placebo

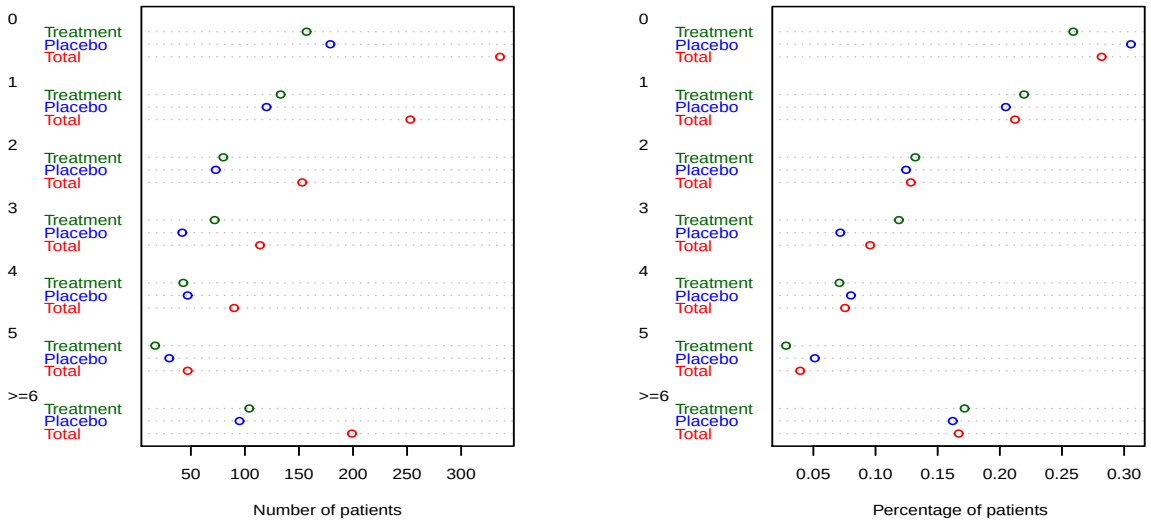


Figure 3.2 – Frequency and percentage summary for treatment (Selenium) and placebo groups of BCC cancer type. The y-axis shows the number of events.

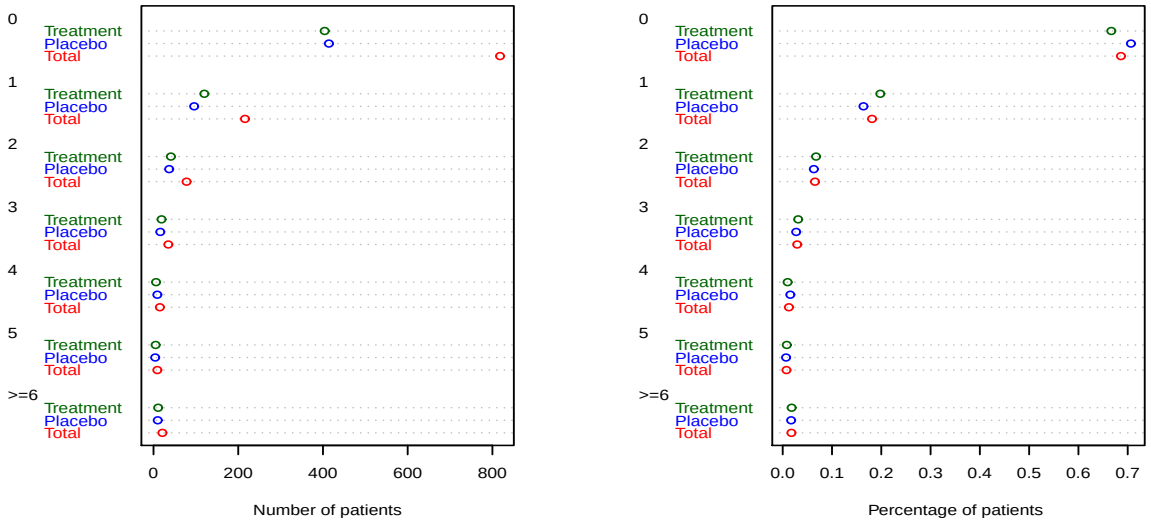


Figure 3.3 – Frequency and percentage summary for treatment (Selenium) and placebo groups of SCC cancer type. The y-axis shows the number of events.

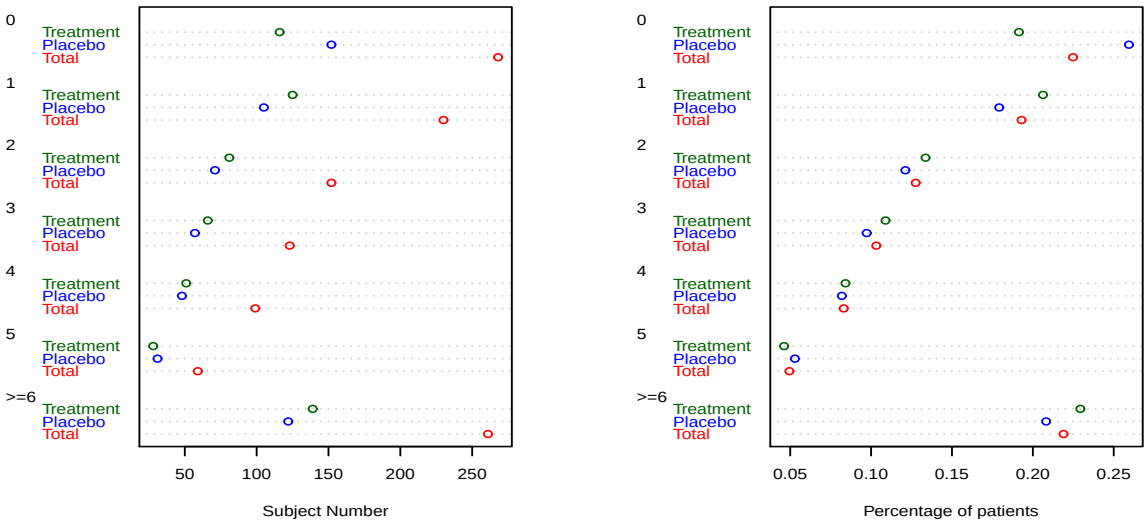


Figure 3.4 – Frequency and percentage summary for treatment (Selenium) and placebo groups of either BCC or SCC cancer types. The y-axis shows the number of events.

Table 3.6 – Summary of the follow-up times for the treatment and placebo.

Group	Min.	Max.	Mean	s. d.
Treatment	27	950	587	188.157
Placebo	47	914	577	187.778

and Selenium group. Figure 3.5 recodes the event times for BCC and SCC cancer types as well as the censoring or the follow-up times all over 1192 patients are included in our analysis. The event and censoring times by the treatment group and cancer type are depicted in Figure 3.6.

3.7.2 Data Analysis Results

This study intends to estimate the influence of the Selenium treatment to the occurrence rates of BCC and SCC cancer types. Because, we consider the dependence between the recurrent of BCC and SCC cancer types, it is necessary to use a multi-type frailty model to make valid inferences. Another important point of interest is to study whether the subjects who are at higher risk of BCC tend to have a higher or a lower risk of SCC. This approach allows us to assess firstly the association between BCC and SCC and secondly the dependence of the

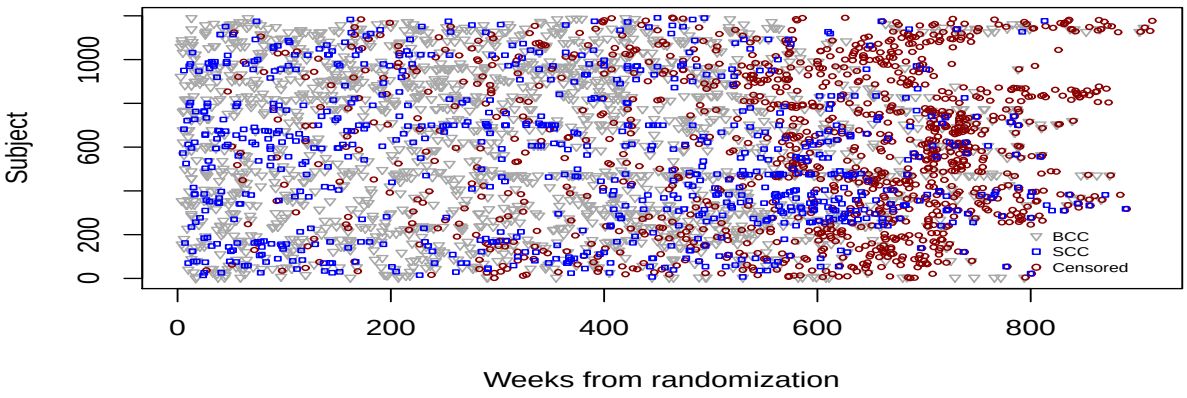


Figure 3.5 – Event times for BCC and SCC cancer types and censoring times.

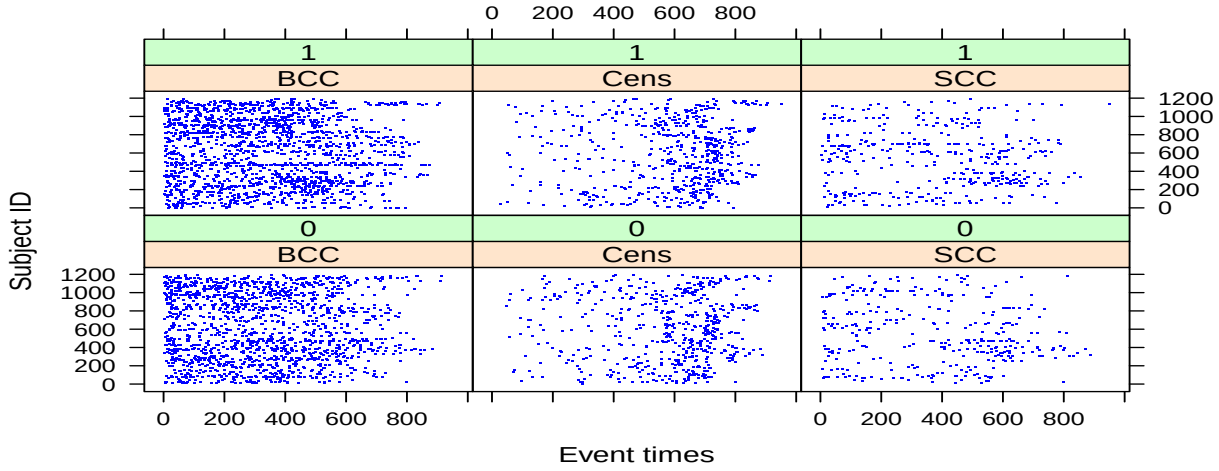


Figure 3.6 – Event times for BCC and SCC cancer types and censoring times (Cens) of the treatment (1) and placebo (0) groups.

Table 3.7 – Parameter estimates for the multi-type Clayton copula and gamma frailty model.

Parameter	EST.	S.E.	Relative risk	p-value
β_1	0.085	0.034	1.088	0.013
β_2	0.079	0.073	1.082	0.279
α_1	1.045	0.037		<0.0001
α_2	2.686	0.088		<0.0001
α	0.856	0.068		<0.0001
Kendall's tau	0.300			

recurrences within subjects. Here, we apply the proposed multi-type models in Section 3.2 to the skin cancer dataset. In the proposed models, we present both Gaussian and Clayton copula functions with two different marginal distributions (Gamma, and Gaussian) to model the multivariate frailty and random effects.

Figure 3.7 shows the estimates of the regression coefficients, variances and copula parameter(s) at each of the MCEM iteration for all the multi-type fitted models. It can be concluded that the MCEM algorithms converged after 30 iterations. The regression coefficient parameters are initialized with values obtained from the Cox proportional hazard models for Gaussian random effects and initialized to one for gamma frailty models. The variances of the frailties and random effects are initialized to zero as well as the correlation parameter.

Conditional on the Clayton copula and gamma frailty model (Model 1) results in Table 3.7 suggest that the intensity of BCC cancer type is significantly (p-value= 0.013) reduced in the placebo group compared to the Selenium group with relative risk (RR) of 1.088. Selenium group is associated with higher risk of the SCC cancer type recurrence (RR=1.082), but the relation was not statistically significant (p-value= 0.279). The estimates of the variances (α_1 and α_2) are 1.045 and 2.686, respectively.

We can conclude that there is a significant association between the event times within the same subject for SCC type (p-value<0.0001) higher than the event times within the same subject for BCC type (p-value<0.0001). A moderate correlation between the event times for BCC and SCC cancer types. The point estimate of Kendall's tau is 0.3 indicates to a moderate

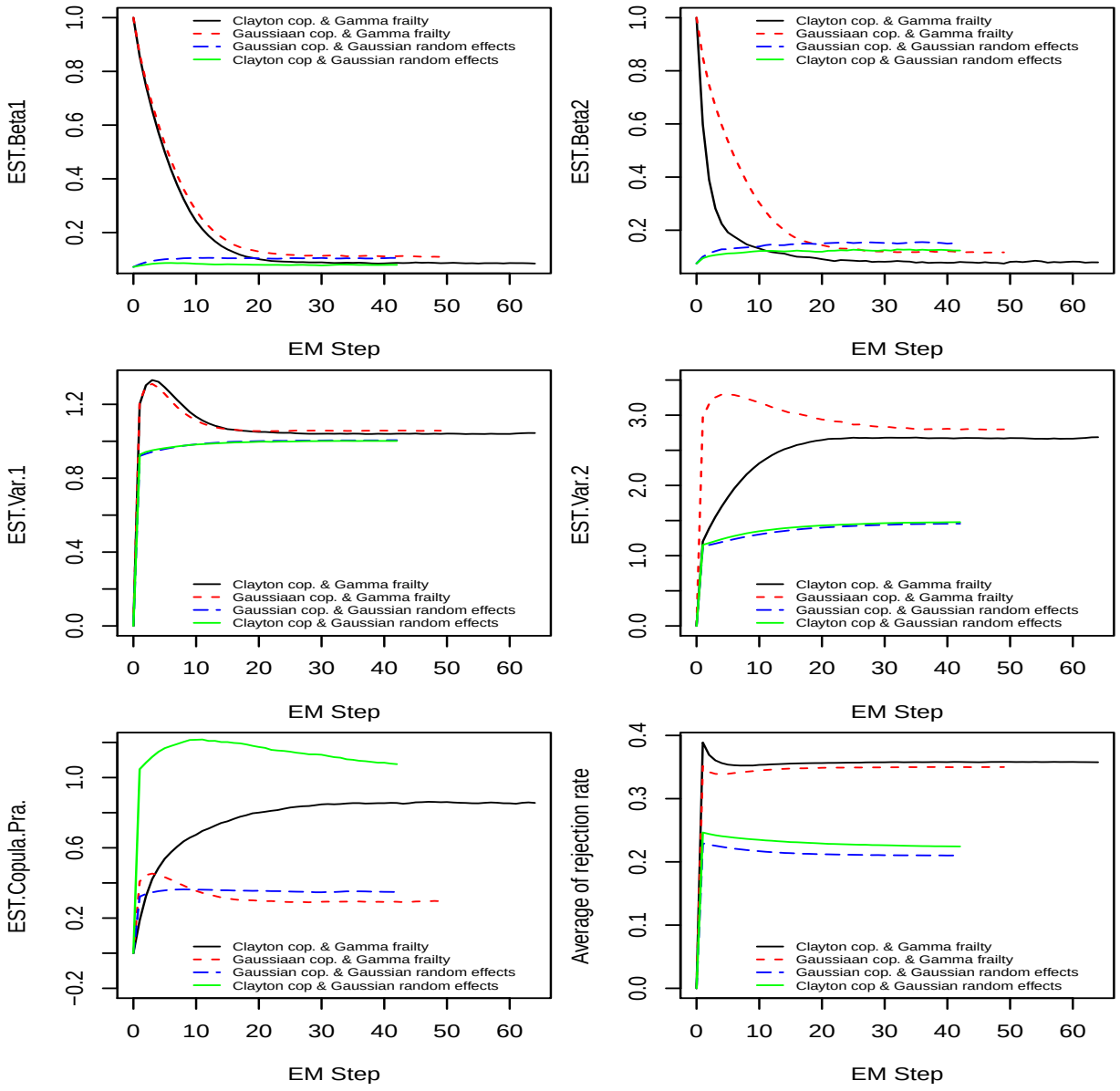


Figure 3.7 – Trace plots for convergence parameter estimates and the average of rejection rates.

Table 3.8 – Parameter estimates for the multi-type Gaussian copula and gamma frailty model.

Parameter	EST.	S.E.	Relative risk	p-value
β_1	0.109	0.034	1.116	0.0013
β_2	0.117	0.073	1.124	0.1099
α_1	1.057	0.031		<0.0001
α_2	2.798	0.091		<0.0001
α	0.296	0.026		<0.0001

Table 3.9 – Parameter estimates for the multi-type Clayton copula and Gaussian random effects model.

Parameter	EST.	S.E.	Relative risk	p-value
β_1	0.0798	0.034	1.083	0.0195
β_2	0.1236	0.072	1.131	0.090
α_1	1.001	0.021		<0.0001
α_2	1.478	0.030		<0.0001
α	1.076	0.077		<0.0001
Kendall's tau	0.349			

association between the frailties of BCC and SCC. The positive estimates of the correlation between the two processes; that is, higher event rate of BCC cancer type be subject to occur with higher event rate of the SCC cancer type.

The results are reported in Table 3.8 for the Gaussian copula and gamma frailties model (Model 2). The rate of the BCC is significantly (p-value= 0.0013) higher in the Selenium group than the placebo group (RR=1.116), while incidence rate is higher (RR=1.124) in the Selenium group than the placebo group but not significant (p-value=0.1099) for the SCC type. There is a significant heterogeneity for both the BCC and SCC event types with α_1=1.057 and α_2=2.798. The correlation coefficient between two event types is estimated at 0.296.

According to the fitted model with the Clayton copula and Gaussian random effects (Model 3), Table 3.9 shows that the Selenium group significantly increases the occurrence rate of BCC (p-value=0.0195) and insignificant increases for SCC (p-value=0.090) cancer types with relative risks (RR) 1.083 and 1.131, respectively. The estimates of the variances and Kendall's tau are 1.001, 1.478, and 0.349, respectively.

Table 3.10 presents the results of using the Gaussian copula and Gaussian random effects

Table 3.10 – Parameter estimates for the multi-type Gaussian copula and Gaussian random effects model.

Parameter	EST.	S.E.	Relative risk	p-value
β_1	0.105	0.034	1.111	0.002
β_2	0.151	0.073	1.163	0.038
α_1	1.005	0.020		<0.0001
α_2	1.455	0.030		<0.0001
α	0.348	0.025		<0.0001

model (Model 4). The rate of recurrence of BCC (p-value=0.002) and SCC (p-value=0.038) is significantly increased with Selenium group (RR=1.111) relative to placebo group. The positive values of α_1=1.005, α_1=1.455, and α=0.348 indicate dependence within the subject event times and between the two event times.

For all fitted models, the intensity rates of BCC are significantly different for patients in the Selenium group versus patients in the placebo group. The intensity rates of SCC are not significantly different for patients in the Selenium group versus patients in the placebo group. However, Model 4 suggests that the Selenium group is significantly increases the rate of SCC. All results for separate models in Tables 2.5, 2.6, and 2.7 and for multi-type models present significant heterogeneity for both the BCC and SCC event types. The event times within the subject of SCC type have more dependence than in the BCC cancer type. The correlation coefficient α and the Kendall's tau are also significantly different from zero, indicating a moderate association between the risk of BCC and SCC recurrence.

The Nelsson-Aalen estimates of the cumulative baseline intensity function for both BCC and SCC cancer types obtained from the separate models and multi-type models are shown in Figures 3.8 and 3.9, respectively. It can be seen that the cumulative baseline intensity for BCC is higher than the cumulative baseline incidence rate of the SCC. We notice that the estimates of the baseline intensity functions using gamma frailty are higher than using Gaussian random effects for both the Clayton and Gaussian copula. This means the choice of the marginal distributions affects the inference of the cumulative baseline intensity functions.

For subjects with two event times for each cancer type, plots based on the credible interval

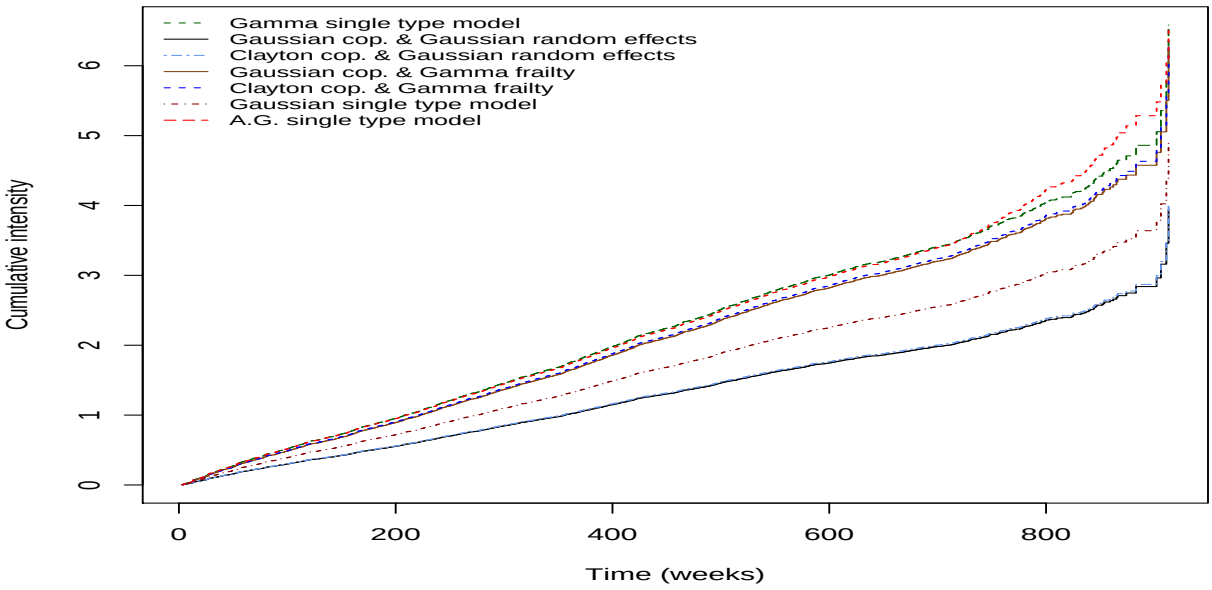


Figure 3.8 – Cumulative intensity functions for the BCC tumor type.

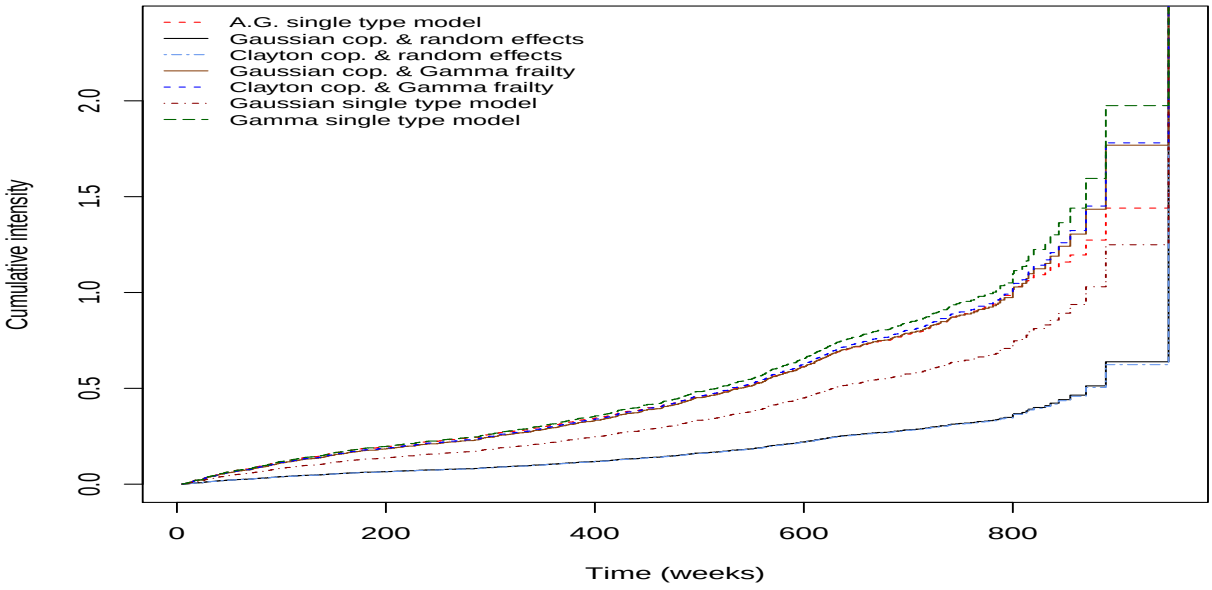


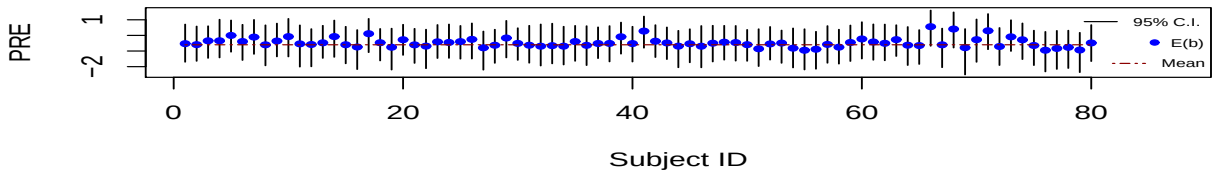
Figure 3.9 – Cumulative intensity functions for the SCC tumor type.

are useful, especially when investigating the heterogeneity over subjects. Figures 3.10, 3.11, 3.12, and 3.13 represent the 95% credible interval for the random effects of the Selenium and placebo group. The assumed means (red horizontal line) are 0 and -0.577 for the Gaussian and log-gamma distributions, respectively. The length of the vertical bars represents the length of the credible interval for each subject. Overall, the length of the intervals are seen to decrease as the number of events per subject increases which explains the reason of the length of credible intervals for placebo group to be longer than the Selenium group. When the average number of events for any group is closer to two events the predicted value of the random effects becomes closer to the assumed mean (0 or -0.577). For example, the average number of events for the treatment and placebo group for the BCC are 3.087 and 2.699 leads to the predicted values are close to 0 and -0.577 for Gaussian and log-gamma random effects, respectively. For the SCC type, the average number of events are 0.673 and 0.599 resulting in the predicted values for the subject with two event times larger than the mean values.

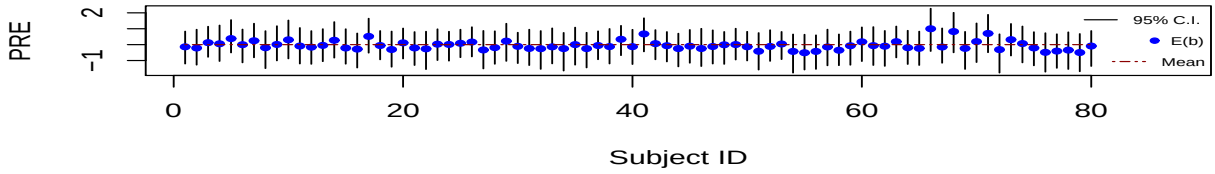
It can be seen in Figures 3.14, 3.15, 3.16, and 3.17 that there is a linear correlation between the number of events and the frailty terms for both the BCC and SCC cancer types. This further indicates that random effects increase with the increasing of the cancer recurrences. The frailties and random effects for both types have greater dependence associated small values higher rather than in the large values.

Martingale residuals are commonly used to examine the possible nonlinear effect of a variable (Therneau and Grambsch 2000). We apply martingale residuals in Figures 3.18, 3.19, 3.20, and 3.21 as a graphical examination for evaluating the multivariate frailty model (Mazroui et al. 2013). We can suggest that the proposed models fit the data well. Particularly, fitted models correctly predict, on average, the number of observed events for each cancer type during follow-up times for every subject. Martingale residuals for patient i for the event type j at the end of the follow-up time τ_i is calculated by $M_{ij} = N(\tau_i) - \hat{\Lambda}_{ij}(\tau_i)$.

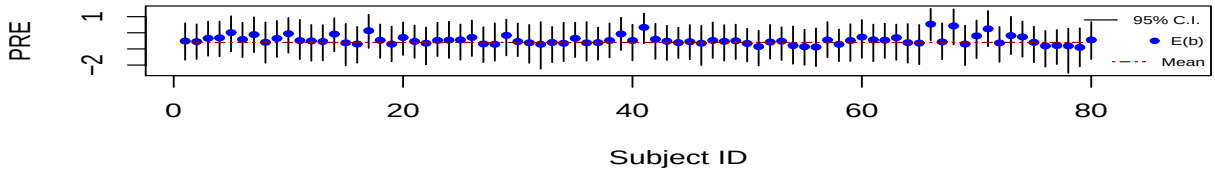
Table 3.11 calculates the sum of squared deviance residuals that can be applied for model comparisons. The sum of squared deviance residuals are more often appropriately ordered



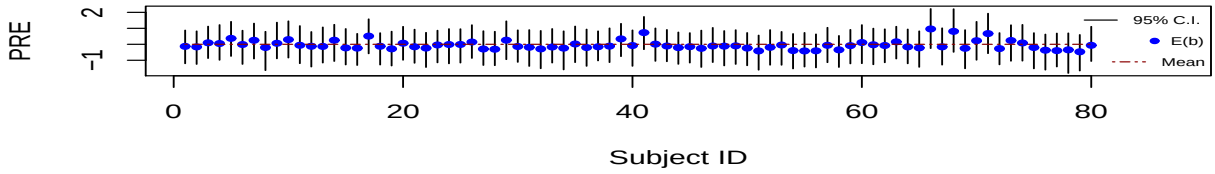
(a) Clayton Copula and gamma Frailty



(b) Clayton copula and Gaussian random effects

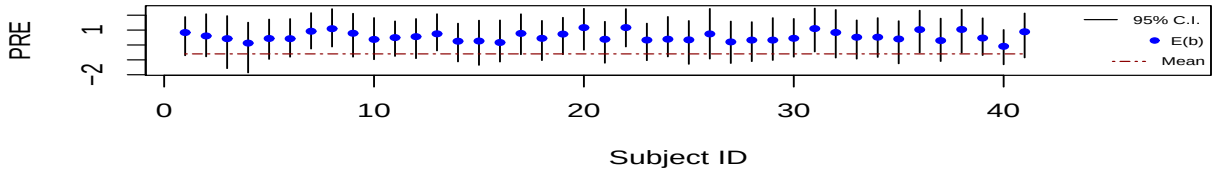


(c) Gaussian copula and gamma frailty

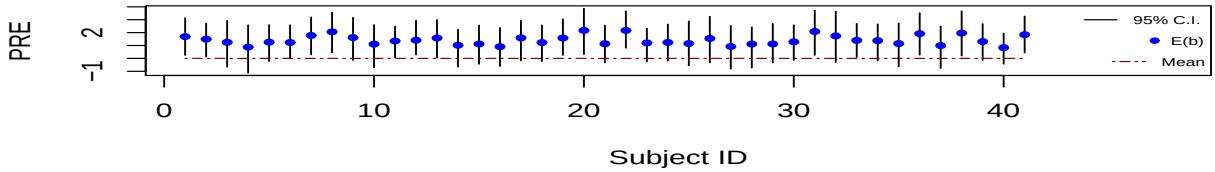


(d) Gaussian copula and Gaussian random Effects

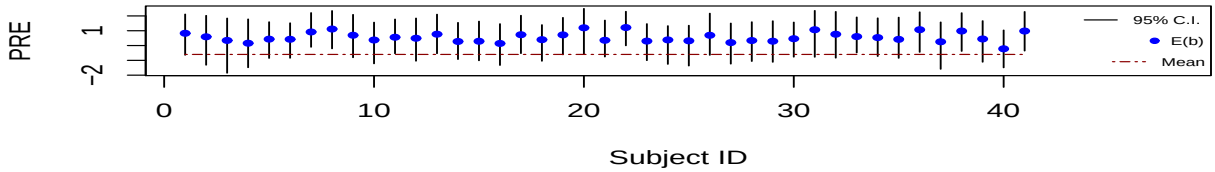
Figure 3.10 – Predicted random effects (PRE) for the patients with two events of the tumor type I and treatment group with a ref. line at means equal to zero for Gaussian random effects and one for gamma frailties (-0.577 for log-gamma random effects).



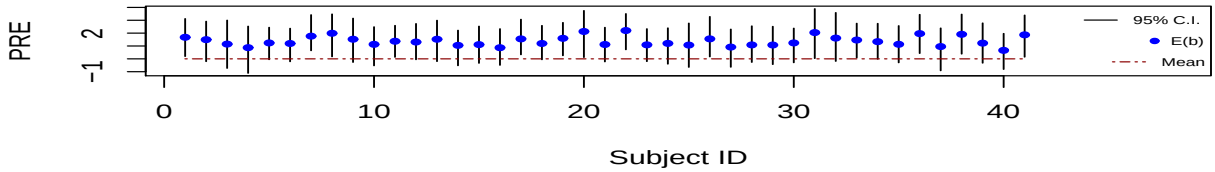
(a) Clayton copula and gamma frailty



(b) Clayton copula and Gaussian random effects

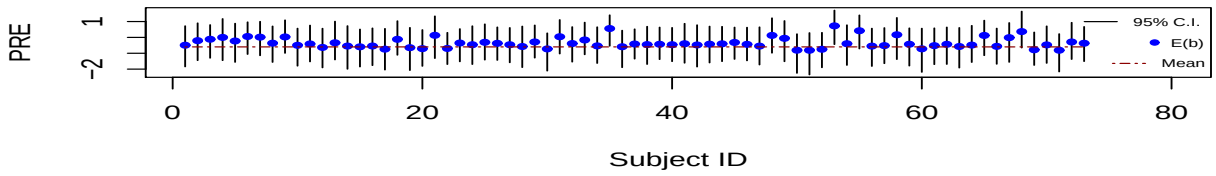


(c) Gaussian copula and gamma frailty

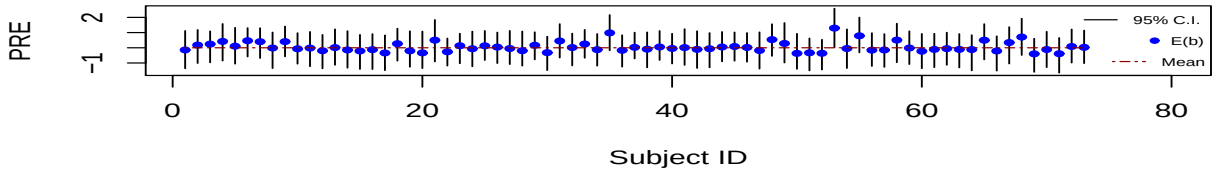


(d) Gaussian copula and Gaussian random effects

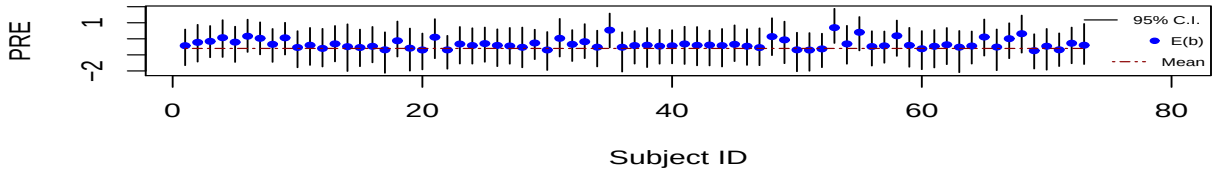
Figure 3.11 – Predicted random effects (PRE) for the patients with two events of the tumor type II and treatment group with a ref. line at means equal to zero for Gaussian random effects and one for gamma frailties (-0.577 for log-gamma random effects).



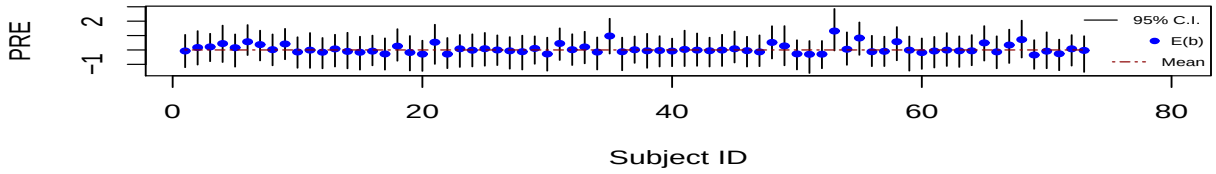
(a) Clayton copula and gamma frailty



(b) Clayton copula and Gaussian random effects

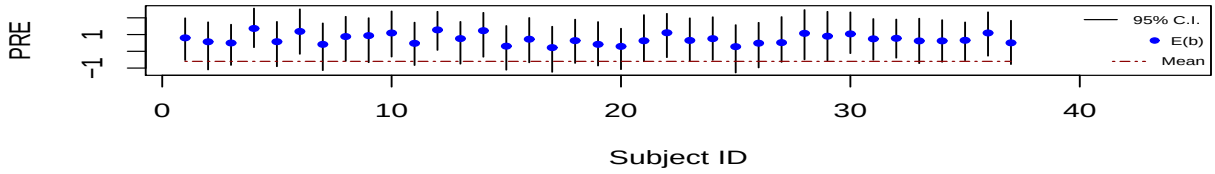


(c) Gaussian copula and gamma frailty

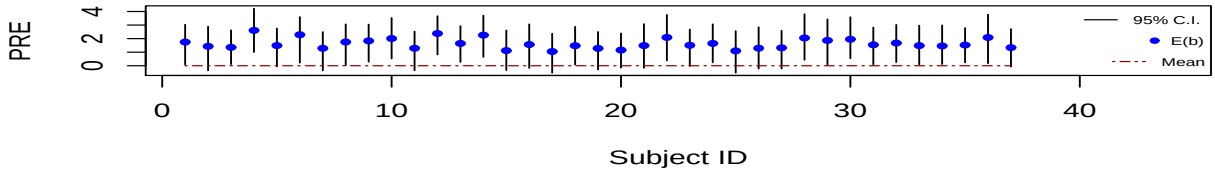


(d) Gaussian copula and Gaussian random effects

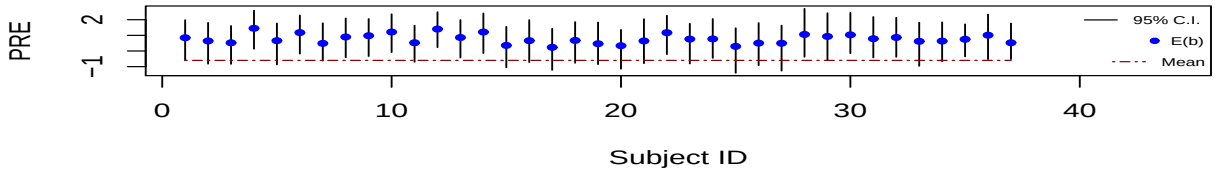
Figure 3.12 – Predicted random effects (PRE) for the patients with two events of the tumor type I and placebo group with a ref. line at means equal to zero for Gaussian random effects and one for gamma frailties (-0.577 for log-gamma random effects).



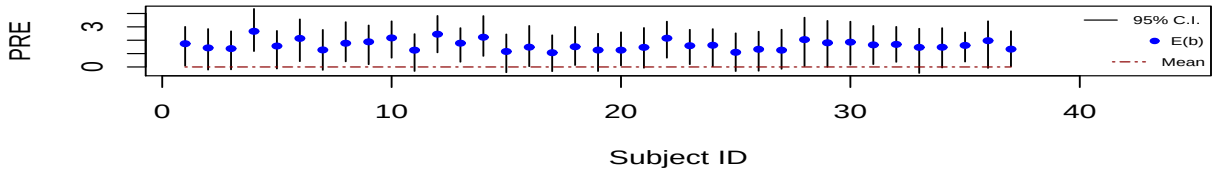
(a) Clayton copula and gamma frailty



(b) Clayton copula and Gaussian random effects



(c) Gaussian copula and gamma frailty



(d) Gaussian copula and Gaussian random effects

Figure 3.13 – Predicted random effects (PRE) for the patients with two events of the tumor type II and placebo group with a ref. line at means equal to zero for Gaussian random effects and one for gamma frailties (-0.577 for log-gamma random effects).

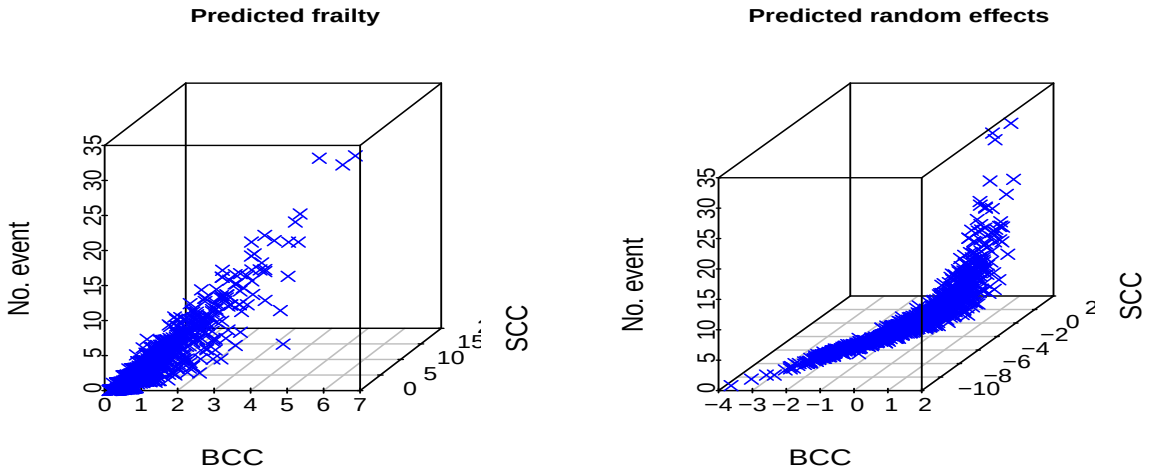


Figure 3.14 – Predicted frailties and random effects with the number of events for Clayton copula and gamma frailty.

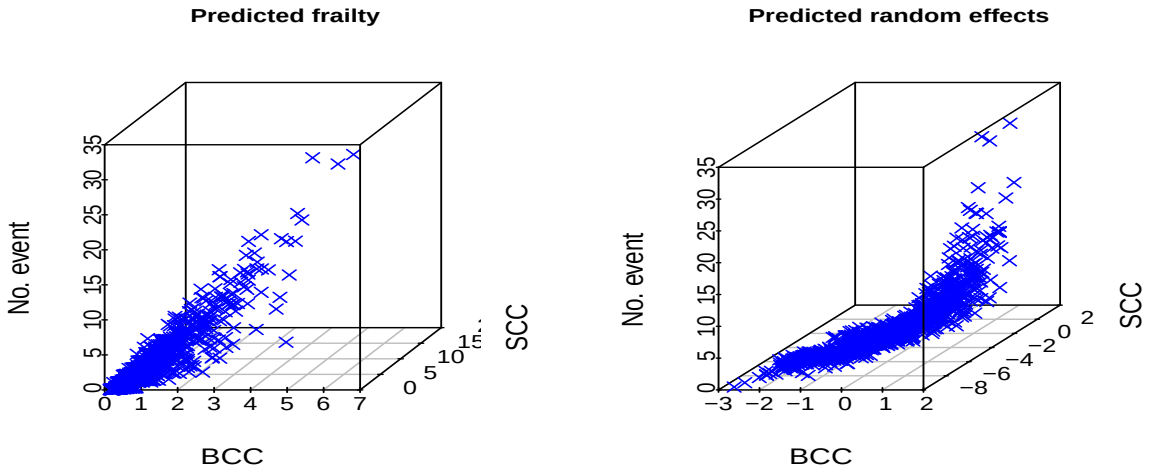


Figure 3.15 – Predicted frailties and random effects with the number of events for Clayton copula and Gaussian random effects.

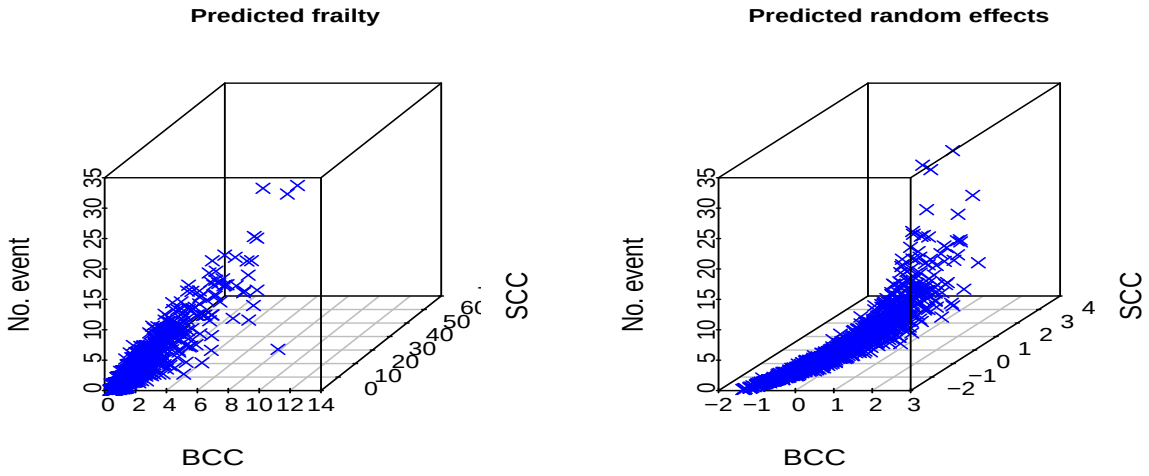


Figure 3.16 – Predicted frailties and random effects with the number of events for Clayton copula and Gaussian random effects.

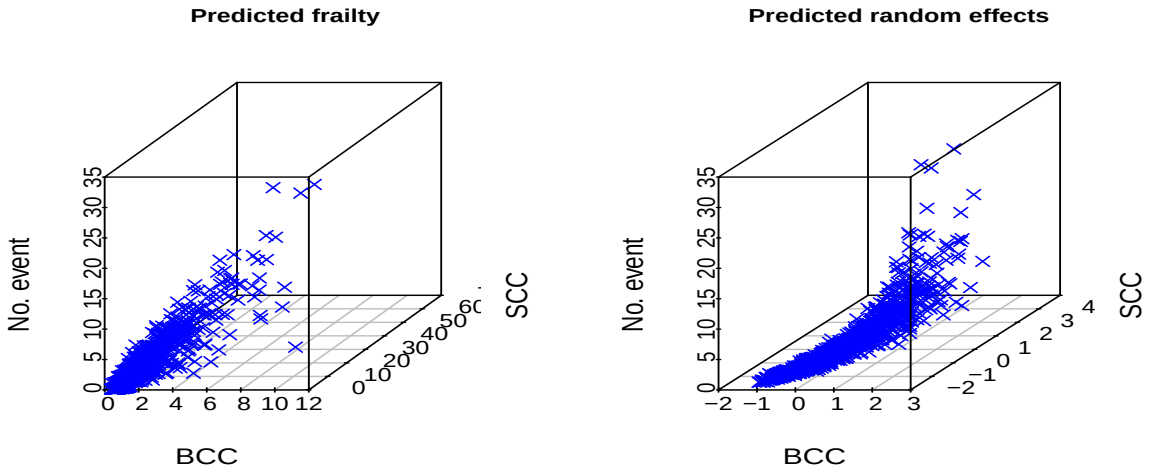


Figure 3.17 – Predicted frailties and random effects with the number of events for Gaussian copula and Gaussian random effects.

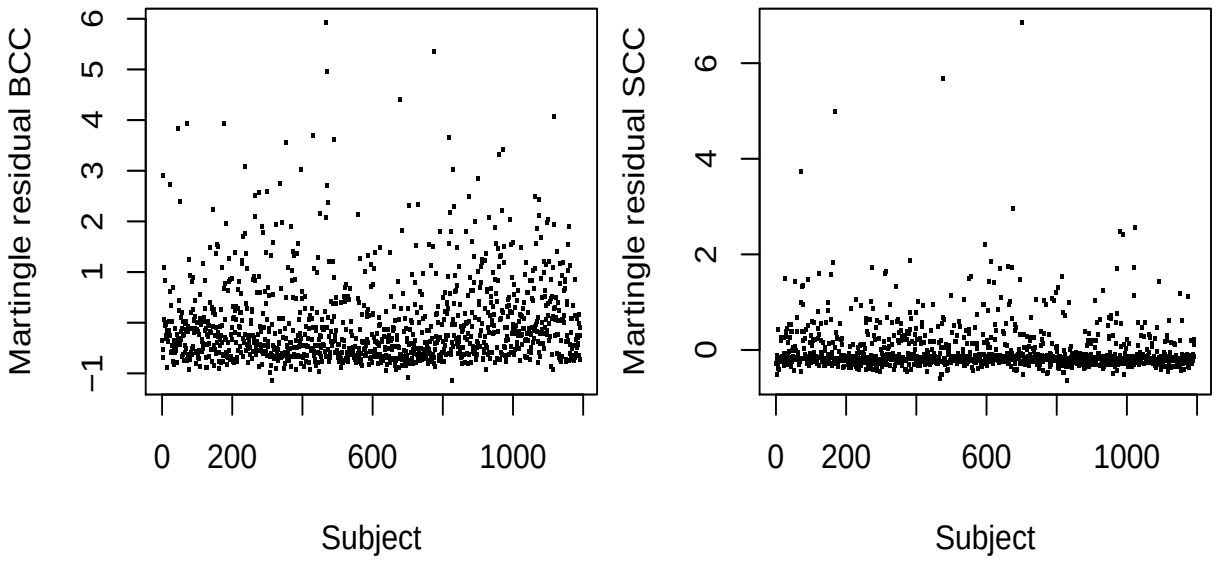


Figure 3.18 – Martingale residuals for Clayton copula and gamma frailties.

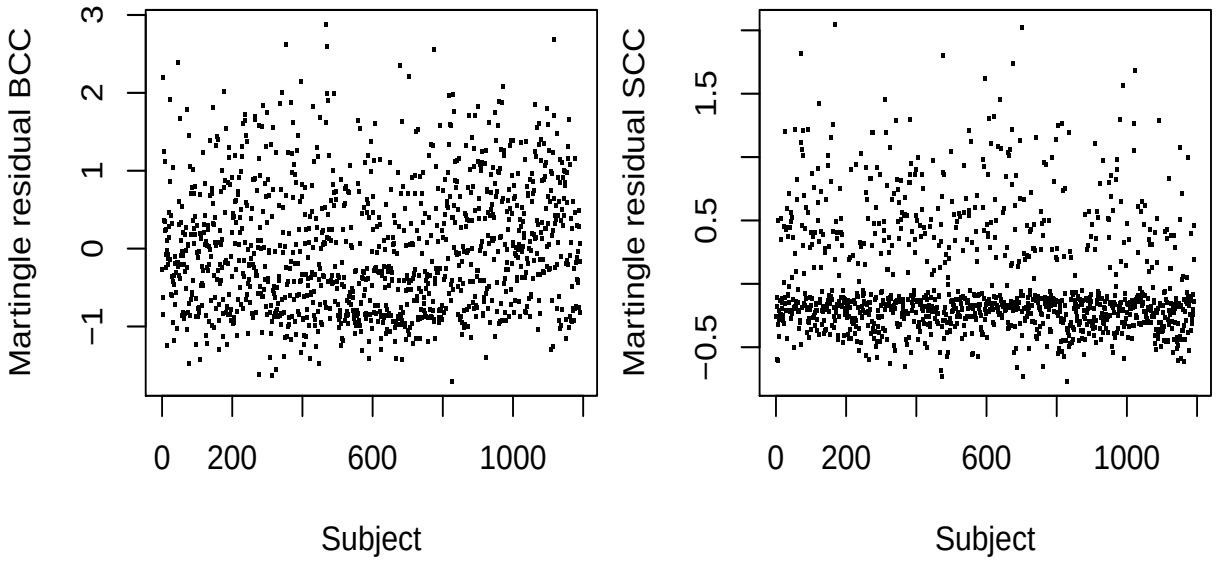


Figure 3.19 – Martingale residuals for Clayton copula and Gaussian random effects.

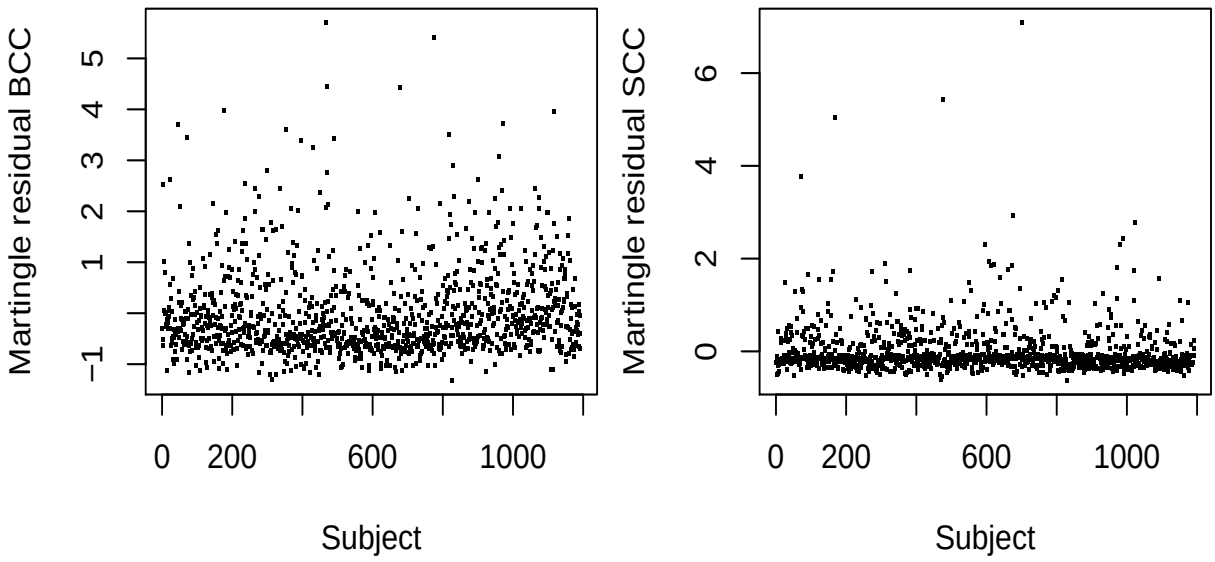


Figure 3.20 – Martingale residuals for Gaussian copula and gamma frailties.

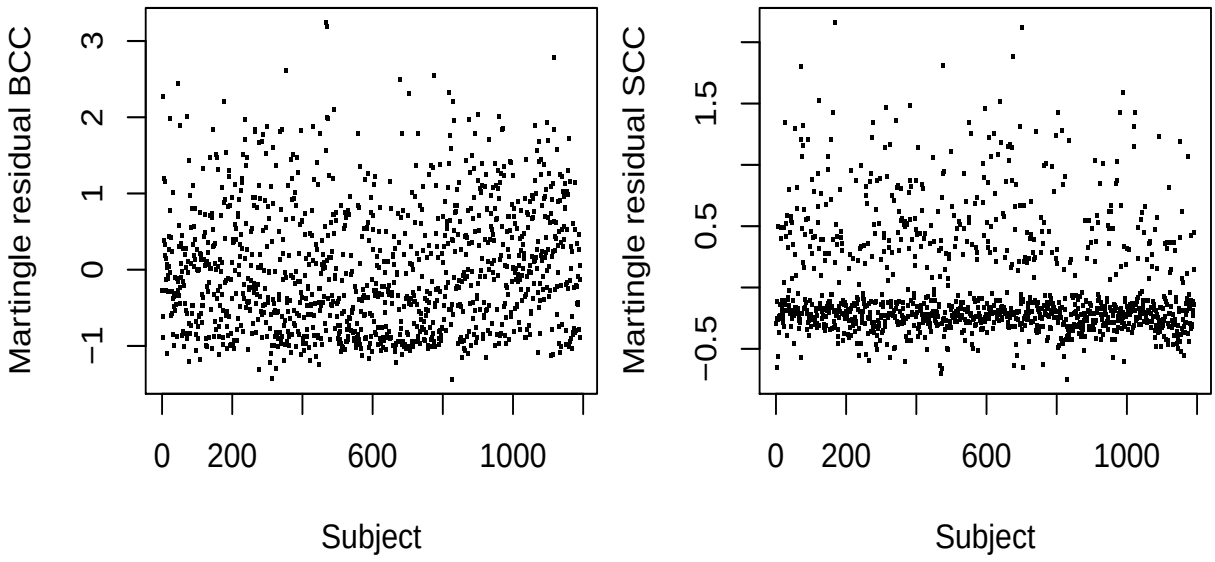


Figure 3.21 – Martingale residuals for Gaussian copula and Gaussian random effects.

Table 3.11 – Deviance residuals for the multi-type fitted models.

Type	Clayton & Gamma	Gaus. & Gamma	Clayton & Gaus.	Gaus. & Gaus.
BCC	355	349	400	394
SCC	307	300	326	333
Total	662	649	726	727

than the sum of squared martingale residuals. For the time independent covariates, the deviance residuals are approximated by Therneau and Grambsch (2000, chapter 4)

$$D_{ij} \approx \frac{N_j(\tau_i) - \hat{\Lambda}_{ij}(\tau_i)}{\sqrt{\hat{\Lambda}_{ij}(\tau_i)}},$$

where D_{ij} is the deviance residual for the individual i of event type j. Table 3.11 would suggest that the Gaussian copula and gamma frailty model is the best fitting of the skin cancer dataset.

3.8 Conclusions and Future Work

Usually, inference in frailty models requires a marginal likelihood approach, whereby the random effects are integrated out of the joint density consisting of response variables and random effects. This may involve the evaluation of analytically intractable integrals over the random effects distributions. To deal with this difficulties, several methods (e.g., numerical integration (Cook et al. 2010), Gaussian quadrature (Liu and Huang 2008), Laplace approximation (Yu et al. 2014) have been suggested but these still have difficulties in situations when the number of random components is large and their correlation structure needs to be modeled. Recently, the penalized maximum likelihood approach (Mazroui et al. 2013), which penalizes the baseline intensity in the marginal likelihood, has been proposed for parameter inference but it can not be used for directly inference of the frailties. Furthermore, it may be be difficult for their to converge with data with only a few numbers of events (Mazroui et al. 2012).

In this chapter, we have proposed multivariate correlated frailty models for multi-type of recurrent events. We are interested in the inference of subject-specific random effects as

well as the model parameters. In particular, two copula functions (Gaussian and Gamma) with Gaussian random effects and gamma frailties are introduced. For inference, we adopt an MCEM approach which provides a computation of appropriate credible intervals for the frailties and random effects. Also we apply the martingale and deviance residuals to illustrate goodness of fit and model selection criteria.

In the application, we can conclude that the choices of the combinations of the copula function and marginal distributions may have an influence on the inference of the regression coefficients and cumulative baseline intensity functions. However, the choice of copula function is not affecting the inference of the variance components (marginal parameters) and the prediction of the frailty or random effect. The length of the credible intervals of the frailty and random effects in the placebo group is, however, much larger than that for treatment group. The cumulative baseline intensity (rate) for BCC is much higher than that for SCC.

We assume a random baseline intercept (representing the random baseline intensity). A more general model is to use a multivariate random effect for the baseline and treatment for all or some of the event types. Beside Clayton and Gaussian copulas, different copula functions can be accommodating. It is also of my future interest to extend the models for more than two event types with a dependent terminal event.

The sensitivity analysis would be still recommended to demonstrate the influence of choices for the copula function and marginal distributions to the inference of the regression coefficients, variance terms, correlation (copula) parameter, and the baseline cumulative intensity functions. There is clearly scope for future work on the development of goodness of fit and criteria for selecting the best model.

3.A Appendices

3.A.1 Metropolis-Hastings Algorithm

In this part we describe in brief the Metropolis-Hastings algorithm. Our approach involves generating random samples $\boldsymbol{w}_i^{(m)}$ from the exact conditional distribution of the frailty terms (given the data) by MCMC sampling. From (3.13) the conditional distribution of $\boldsymbol{w}_i|\mathbf{D}_i$ can be expressed as

$$g_{\boldsymbol{w}_i|\mathbf{D}_i}\left(\boldsymbol{w}_i|\boldsymbol{\xi}\right) \propto L_i\left(\boldsymbol{\xi}, \boldsymbol{w}_i\right).$$

There are several approaches to select proposal functions, resulting in specific types of M-H algorithms. We adopt a random-walk Metropolis-Hastings algorithm. For the density distribution of the candidates we use copula density functions that match copula densities for the frailty terms $\boldsymbol{w}_i$ to propose candidate samples for $\boldsymbol{w}_i^{\star}$. For example, Gaussian and Clayton copulas and Gaussian marginal distributions with some proposal copula parameters $\boldsymbol{\alpha}_{proposal}$ is used. The Metropolis-Hastings algorithm can be described in Algorithm 3.2.

Algorithm: 3.2 A Metropolis-Hastings algorithm for the E-step in the MCEM algorithm.

---

1. Initialize $\boldsymbol{W}^{(0)} = (\boldsymbol{w}_1'^{(0)}, \ldots, \boldsymbol{w}_N'^{(0)})'$. At iteration m, $(m = 1, \ldots, M)$,
2. Sample $\boldsymbol{w}_i^{\star}$, for the i^{th} component of $\boldsymbol{W}$ from a proposal distribution $\boldsymbol{w}_i^{\star} \sim g(\boldsymbol{w}_i^{\star}|\boldsymbol{w}_i^{(m-1)})$.
3. Compute an acceptance ratio (probability)

$$r = \frac{g_{\boldsymbol{w}_i|\mathbf{D}_i}\left(\boldsymbol{w}_i^{\star}|\boldsymbol{\xi}\right) g\left[\boldsymbol{w}_i^{(m-1)}|\boldsymbol{w}_i^{\star}\right]}{g_{\boldsymbol{w}_i|\mathbf{D}_i}\left(\boldsymbol{w}_i^{(m-1)}|\boldsymbol{\xi}\right) g\left[\boldsymbol{w}_i^{\star}|\boldsymbol{w}_i^{(m-1)}\right]}.$$

4. Sample $u \sim$uniform $(0, 1)$. Set $\boldsymbol{w}_i^{(m)}$ to $\boldsymbol{w}_i^{\star}$ if $u < r$ and to $\boldsymbol{w}_i^{(m-1)}$ if $u > r$.
5. Repeat steps 2 to 4 for generating M random samples $\boldsymbol{W}^{(1)}, \ldots, \boldsymbol{W}^{(M)}$.

3.A.2 Expected Log-likelihood of Gaussian and Clayton Copula with Gamma Marginals

In this section we derive the expected log-likelihood $Q_2(\boldsymbol{\alpha})$ for Gaussian and Clayton copula with gamma marginal distributions. For instance, when frailty terms are Gaussian copula, $Q_2(\alpha)$ is (see e.g., Hoff 2009, page 110)

$$\begin{aligned} Q_2(\alpha) &= -\frac{1}{2}\log(\mid \boldsymbol{R}_J \mid) - \frac{1}{2}\mathrm{E}\left[\boldsymbol{q}_i'\left(\boldsymbol{R}_J^{-1} - \mathbf{I}_J\right)\boldsymbol{q}_i\right] \\ &= -\frac{1}{2}\log(\mid \boldsymbol{R}_J \mid) - \frac{1}{2}\mathrm{tr}\left[\left(\boldsymbol{R}_J^{-1} - \mathbf{I}_J\right)\mathrm{E}\left(\boldsymbol{q}_i'\boldsymbol{q}_i\right)\right]. \end{aligned}$$

It is important to mention that $\mathrm{E}(\cdot)$ is the expectation of the frailty conditional distribution. In such case when we consider a Clayton copula, $Q_2(\alpha)$ can be expressed as

$$\begin{aligned} Q_2(\alpha) = \sum_{i=1}^{N}\left\{[-1/(\alpha - J)]\,\mathrm{E}\left[\log\left(\sum_{j=1}^{J} u_{ij}^{-\alpha} - J + 1\right)\right] + \sum_{j=1}^{J}\mathrm{E}\log\left(u_{ij}^{-\alpha-1}\right)\right\} \\ + N\,\log\left\{(-\alpha)^J\left[\prod_{j=0}^{J-1}(-\frac{1}{\alpha} - j)\right]\right\}. \end{aligned}$$

For example, a bivariate Clayton copula, $Q_2(\alpha)$ is

$$\begin{aligned} g(w_{i1},\, w_{i2}) = \sum_{i=1}^{N}\left\{[-1/(\alpha-2)]\,\mathrm{E}\left[\log\left(u_{i1}^{-\alpha} + u_{i2}^{-\alpha} - 1\right)\right] + (-\alpha - 1)\,\mathrm{E}\left[\log(u_{i1}) + \log(u_{i2})\right]\right\} \\ + N\left[\log(1+\alpha)\right]. \end{aligned}$$

The marginal distributions, when w_{ij} are gamma $\left(\frac{1}{\alpha_j},\, \alpha_j\right)$ distributed with mean =1 and variance α_j, consequently $Q_2(\alpha_j)$ can be in the form

$$Q_2(\alpha_j) = \sum_{i=1}^{N} \left\{ \left(\frac{1}{\alpha_j} - 1 \right) \mathrm{E}\left[\log\left(w_{ij} \right) \right] - \frac{1}{\alpha_j} \mathrm{E}(w_{ij}) \right\} - N \left[\log \Gamma\left(\frac{1}{\alpha_j} \right) + \frac{1}{\alpha_j} \log\left(\alpha_j \right) \right].$$

3.A.3 Covariance Matrix for Estimated Parameters

We denote the complete log-likelihood $\mathcal{L}\left(\boldsymbol{\xi}, \boldsymbol{W}\right)$ in (3.11), its first derivative vector, and second derivative matrix simply by $\mathcal{L}\left(\boldsymbol{\xi}\right)$, $\mathcal{L}_1\left(\boldsymbol{\xi}\right)$, and $\mathcal{L}_2\left(\boldsymbol{\xi}\right)$, respectively. The second derivative matrix $\mathcal{L}_2\left(\boldsymbol{\xi}\right)$ is not the observed information matrix because the uncertainty about $\boldsymbol{W}$ has been taken into account. The components of the observed information matrix $\boldsymbol{I}\left(\hat{\boldsymbol{\xi}}\right)$ are obtained as follows (see e.g., Huang and Liu 2007, Vaida et al. 2000).

3.A.3.1 First-order Derivative Vector

The first order derivation vectors are defined by

$$\mathcal{L}_1\left(\boldsymbol{\xi}\right) = \begin{bmatrix} \mathcal{L}_1\left(\boldsymbol{\beta}\right) \\ \mathcal{L}_1\left(\boldsymbol{\lambda}\right) \\ \mathcal{L}_1\left(\boldsymbol{\alpha}\right) \end{bmatrix}, \mathcal{L}_1\left(\boldsymbol{\beta}\right) = \begin{bmatrix} \mathcal{L}_1\left(\boldsymbol{\beta}_1\right) \\ \vdots \\ \mathcal{L}_1\left(\boldsymbol{\beta}_J\right) \end{bmatrix}, \mathcal{L}_1\left(\boldsymbol{\lambda}\right) = \begin{bmatrix} \mathcal{L}_1\left(\lambda_{11}\right) \\ \vdots \\ \mathcal{L}_1\left(\lambda_{JK_J}\right) \end{bmatrix} \mathcal{L}_1\left(\boldsymbol{\alpha}\right) = \begin{bmatrix} \mathcal{L}_1\left(\alpha_1\right) \\ \vdots \\ \mathcal{L}_1\left(\alpha_J\right) \\ \mathcal{L}_1\left(\alpha\right) \end{bmatrix},$$

where

$$\boldsymbol{\mathcal{L}}_1\left(\boldsymbol{\beta}_j\right) = \frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}_j} = \sum_{i=1}^{m} \left[K_{ij}(\tau_i)\boldsymbol{x}_i - \boldsymbol{x}_i \Lambda_{0j}(\tau_i) \mathrm{E}\left(w_{ij}\right) \exp\left(\boldsymbol{x}_i' \boldsymbol{\beta}_j\right) \right],$$

$$\mathcal{L}_1\left(\lambda_{k_j}\right) = \frac{\partial \mathcal{L}}{\partial \lambda_{k_j}} = \frac{1}{\lambda_{k_j}} - \sum_{i \in R_{t\left(k_j\right)}} \mathrm{E}\left(w_{ij}\right) \exp\left(\boldsymbol{x}_i' \boldsymbol{\beta}_j\right), \; k_j = 1, \ldots, K_j,$$

with $\lambda_{k_j} = \lambda_{0j}\left(t_{j(k_j)}\right)$. Here $t_{j(k_j)}$ $(k_j = 1, \ldots, K_j)$ are the distinct event times for the j^{th} type. The first derivative for the gamma marginal's is

$$\mathcal{L}_1(\alpha_j) = \frac{\partial \mathcal{L}}{\partial \alpha_j} = \frac{N}{\alpha_j^2}\left[\log(\alpha_j) - 1 + \Gamma_1\left(\frac{1}{\alpha_j}\right) + -\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}\left[\log(w_{ij})\right] + \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(w_{ij})\right],$$

where Γ_1 is the digamma function which is the first derivative of $\log[\Gamma(\cdot)]$. The first derivative for Gaussian marginal's is

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha_j^2} = -\frac{1}{2}\left[\frac{N}{\alpha_j} + \frac{1}{\alpha_j^2}\sum_{i=1}^{N}\mathrm{E}\left(w_{ij}^2\right)\right].$$

3.A.3.2 Second-order Derivative Matrix

The components of the second derivative matrix are given as

$$\mathcal{L}_2(\boldsymbol{\xi}) = \begin{bmatrix} \mathcal{L}_2^{(\beta\beta)} & \mathcal{L}_2^{(\beta\lambda)} & \mathbf{0} \\ \mathcal{L}_2^{(\beta\lambda)} & \mathcal{L}_2^{(\lambda\lambda)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{L}_2^{(\alpha\alpha)} \end{bmatrix},$$

where the second derivative matrix components are

$$\mathcal{L}_2^{(\beta\beta)} = \begin{bmatrix} \frac{\partial^2\mathcal{L}}{\partial\beta_{11}\partial\beta_{11}} & \cdots & \frac{\partial^2\mathcal{L}}{\partial\beta_{11}\partial\beta_{Jp}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\mathcal{L}}{\partial\beta_{Jp}\partial\beta_{11}} & \cdots & \frac{\partial^2\mathcal{L}}{\partial\beta_{Jp}\partial\beta_{Jp}} \end{bmatrix}, \mathcal{L}_2^{(\lambda\lambda)} = \begin{bmatrix} \frac{\partial^2\mathcal{L}}{\partial\lambda_{11}\partial\lambda_{11}} & \cdots & \frac{\partial^2\mathcal{L}}{\partial\lambda_{11}\partial\lambda_{JK_J}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\mathcal{L}}{\partial\lambda_{JK_J}\partial\lambda_{11}} & \cdots & \frac{\partial^2\mathcal{L}}{\partial\lambda_{JK_J}\partial\lambda_{JK_J}} \end{bmatrix},$$

$$\mathcal{L}_2^{(\beta\lambda)} = \begin{bmatrix} \frac{\partial^2\mathcal{L}}{\partial\beta_{11}\partial\lambda_{11}} & \cdots & \frac{\partial^2\mathcal{L}}{\partial\beta_{11}\partial\lambda_{JK_J}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\mathcal{L}}{\partial\beta_{Jp}\partial\lambda_{11}} & \cdots & \frac{\partial^2\mathcal{L}}{\partial\beta_{JP}\partial\lambda_{JK_J}} \end{bmatrix}, \text{ and } \mathcal{L}_2^{(\alpha\alpha)} = \begin{bmatrix} \frac{\partial^2\mathcal{L}}{\partial\alpha_1\partial\alpha_1} & \cdots & \frac{\partial^2\mathcal{L}}{\partial\alpha_1\partial\alpha_J} & \frac{\partial^2\mathcal{L}}{\partial\alpha_1\partial\alpha} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial^2\mathcal{L}}{\partial\alpha_J\partial\alpha_1} & \cdots & \frac{\partial^2\mathcal{L}}{\partial\alpha_J\partial\alpha_J} & \frac{\partial^2\mathcal{L}}{\partial\alpha_J\partial\alpha} \\ \frac{\partial^2\mathcal{L}}{\partial\alpha\partial\alpha_1} & \cdots & \frac{\partial^2\mathcal{L}}{\partial\alpha\partial\alpha_J} & \frac{\partial^2\mathcal{L}}{\partial\alpha\partial\alpha} \end{bmatrix}.$$

The forms for the second derivative are

$$\frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\beta}_j\partial\boldsymbol{\beta}_{j'}} = \sum_{i=1}^{N}\left[-\boldsymbol{x}_i\boldsymbol{x}_i'\Lambda_{0j}(\tau_i)\mathrm{E}(w_{ij})\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_j\right)\right],$$

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda_{k_j}^2} = \left\{ -\frac{N_j \left[t_{(k_j)} \right]}{\lambda_{k_j}^2} \right\},$$

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda_{l_j} \partial \boldsymbol{\beta}_j} = - \sum_{i \in R_{t_{(l_j)}}} \boldsymbol{x}_i' \mathrm{E}\left(w_{ij}\right) \exp\left(\boldsymbol{x}_i' \boldsymbol{\beta}_j\right),$$

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha_j^2} = -\frac{2N}{\alpha_j^3} \left[\log\left(\alpha_j\right) - 1 + \Gamma_1\left(\frac{1}{\alpha_j}\right) - \frac{1}{N} \sum_{i=1}^{N} \mathrm{E}\left[\log\left(w_{ij}\right)\right] + \frac{1}{N} \sum_{i=1}^{N} \mathrm{E}(w_{ij}) \right] + \frac{N}{\alpha_j^3} \left[1 - \frac{1}{\alpha_j} \Gamma_2\left(\frac{1}{\alpha_j}\right) \right],$$

where Γ_2 is the trigamma function, that is the second derivative of $\log\left[\Gamma\left(\cdot\right)\right]$. The second derivative for Gaussian marginal's is

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha_j^2} = -\frac{1}{2} \left[-\frac{N}{\alpha_j^2} + \frac{2}{\alpha_j^3} \sum_{i=1}^{N} \mathrm{E}\left(w_{ij}^2\right) \right].$$

All other off-diagonal elements of the second derivatives are equal to zero.

Bibliography

Aalen, O. O., Bjertness, E. and Sønju, T.: 2007, Analysis of dependent survival data applied to lifetimes of amalgam fillings, *Statistics in Medicine* **14**(16), 1819–1829.

Abu-Libdeh, H., Turnbull, B. W. and Clark, L. C.: 1990, Analysis of multi-type recurrent events in longitudinal studies; application to a skin cancer prevention trial, *Biometrics* **16**(4), 1017–1034.

An, X. and Bentler, P. M.: 2012, Efficient direct sampling MCEM algorithm for latent variable models with binary responses, *Computational Statistics & Data Analysis* **56**(2), 231–244.

Andersen, P. K. and Gill, R. D.: 1982, Cox's regression model for counting processes: a large sample study, *The Annals of Statistics* **10**(4), 1100–1120.

Booth, J. G. and Hobert, J. P.: 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.

Chen, B. E. and Cook, R. J.: 2009, The analysis of multivariate recurrent events with partially missing event types, *Lifetime Data Analysis* **15**(1), 41–58.

Chen, B. E., Cook, R. J., Lawless, J. F. and Zhan, M.: 2004, Statistical methods for multivariate interval-censored recurrent events, *Statistics in Medicine* **24**(5), 671–691.

Chen, J., Zhang, D. and Davidian, M.: 2002, A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution, *Biostatistics* **3**(3), 347–360.

Clark, L. C., Combs, G. F., Turnbull, B. W., Slate, E. H., Chalker, D. K., Chow, J., Davis, L. S., Glover, R. A., Graham, G. F., Gross, E. G. et al.: 1996, Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skina randomized controlled trial, *JAMA: the Journal of the American Medical Association* **276**(24), 1957–1963.

Cook, R. J. and Lawless, J. F.: 2007, *The Statistical Analysis of Recurrent Events*, Springer, New York.

Cook, R. J., Lawless, J. F. and Lee, K.-A.: 2010, A copula-based mixed poisson model for bivariate recurrent events under event-dependent censoring, *Statistics in Medicine* **29**(6), 694–707.

Cox, D. R.: 1972, Regression models and life-tables, *Journal of the Royal Statistical Society. Series B (Methodological)* **34**(2), 187–220.

Dempster, A. P., Laird, N. M. and Rubin, D. B.: 1977, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)* **39**(1), 1–38.

Duchateau, L. and Janssen, P.: 2008, *The Frailty Model*, Springer, New York.

Duchateau, L., Janssen, P., Kezic, I. and Fortpied, C.: 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., Slate, E. H., Reid, M. E., Turnbull, B. W., Wilkins, P. A., Combs, G. F., Park, H. K., Gross, E. G., Graham, G. F., Stratton, M. S. 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.

Finkelstein, M.: 2008, *Failure Rate Modelling for Reliability and Risk*, Springer, New York.

Ha, I. D., Lee, Y. and Song, J.-K.: 2001, Hierarchical likelihood approach for frailty models, *Biometrika* **88**(1), 233–233.

Ha, I. D., Lee, Y. and Song, J.-K.: 2002, Hierarchical-likelihood approach for mixed linear models with censored data, *Lifetime Data Analysis* **8**(2), 163–176.

Ha, I. D., Sylvester, R., Legrand, C. and MacKenzie, G.: 2011, Frailty modelling for survival data from multi-centre clinical trials, *Statistics in Medicine* **30**(17), 2144–2159.

Hoff, P. D.: 2009, *A First Course in Bayesian Statistical Methods*, Springer, New York.

Hougaard, P.: 2000, *Analysis of multivariate survival data*, Springer, New York.

Huang, X. and Liu, L.: 2007, A joint frailty model for survival and gap times between recurrent events, *Biometrics* **63**(2), 389–397.

Huang, X. and Wolfe, R. A.: 2002, A frailty model for informative censoring, *Biometrics* **58**(3), 510–520.

Johansen, S.: 1983, An extension of Cox's regression model, *International Statistical Review* **51**(2), 165–174.

Klein, J. P. and Moeschberger, M. L.: 2003, *Survival Analysis*, 2nd edn, Springer, New York.

Liu, L. and Huang, X.: 2008, The use of Gaussian quadrature for estimation in frailty proportional hazards models, *Statistics in medicine* **27**(14), 2665–2683.

Liu, L., Wolfe, R. A. and Huang, X.: 2004, Shared frailty models for recurrent events and a terminal event, *Biometrics* **60**(3), 747–756.

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 Meyer, R.: 2005, Bayesian inference for recurrent events data using time-dependent frailty, *Statistics in Medicine* **24**(8), 1263–1274.

Mazroui, Y., Mathoulin-Pélissier, S., MacGrogan, G., Brouste, V. and Rondeau, V.: 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., Mathoulin-Pelissier, S., Soubeyran, P. and Rondeau, V.: 2012, General joint frailty model for recurrent event data with a dependent terminal event: Application to follicular lymphoma data, *Statistics in Medicine* **31**(11-12), 1162–1176.

McGilchrist, C. and Aisbett, C.: 1991, Regression with frailty in survival analysis, *Biometrics* **47**(2), 461–466.

McGilchrist, C. and Yau, K.: 2008, Survival analysis with time dependent frailty using a longitudinal model, *Australian Journal of Statistics* **38**(1), 53–60.

McLachlan, G. J. and Krishnan, T.: 2007, *The EM Algorithm and Extensions*, Vol. 382, Wiley-Interscience, New Jersey.

Moreno, E. S. S.: 2008, *Nonparametric Frailty Models for Clustered Survival Data*, PhD thesis, Cornell University.

Prentice, R. L., Williams, B. J. and Peterson, A. V.: 1981, On the regression analysis of multivariate failure time data, *Biometrika* **68**(2), 373–379.

Ripatti, S., Larsen, K. and Palmgren, J.: 2002, Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm, *Lifetime Data Analysis* **8**(4), 349–360.

Ripatti, S. and Palmgren, J.: 2000, Estimation of multivariate frailty models using penalized partial likelihood, *Biometrics* **56**(4), 1016–1022.

Robert, C. and Casella, G.: 2009, *Introducing Monte Carlo Methods with R*, Springer, New York.

Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V. and Soubeyran, P.: 2007, Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events, *Biostatistics* **8**(4), 708–721.

Song, P. X.-K.: 2000, Multivariate dispersion models generated from Gaussian copula, *Scandinavian Journal of Statistics* **27**(2), 305–320.

Therneau, T. M. and Grambsch, P. M.: 2000, *Modeling Survival Data: Extending the Cox Model*, Springer, New York.

Vaida, F., Xu, R. et al.: 2000, Proportional hazards model with random effects, *Statistics in Medicine* **19**(24), 3309–3324.

Wei, G. C. and Tanner, M. A.: 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.

Wei, L.-J., Lin, D. Y. and Weissfeld, L.: 1989, Regression analysis of multivariate incomplete failure time data by modeling marginal distributions, *Journal of the American Statistical Association* **84**(408), 1065–1073.

Yau, K. and McGilchrist, C.: 1998, ML and REML estimation in survival analysis with time dependent correlated frailty, *Statistics in Medicine* **17**(11), 1201–1213.

Yu, Z., Liu, L., Bravata, D. M. and Williams, L. S.: 2014, Joint model of recurrent events and a terminal event with time-varying coefficients, *Biometrical Journal* **56**(2), 183–197.

Chapter 4 General Conclusions and Areas for Future Work

4.1 Summary

Many studies presented a variety of intensity models for the single-type with or without informative censoring which are depend on different approaches to model the baseline intensity and frailty terms as well as different estimation methods. The baseline intensity function is either assumed to be completely unspecified (semiparametric model) or to follow a distribution depending on a low-dimensional unknown parameter vector (e.g., Weibull, Gompertz). Other studies fit the shared frailty model by using splines to model the baseline intensity function. Splines with specified number of knots for the baseline hazard are parametric models. But similar to the models with piecewise constant baseline intensity, it is much more flexible compared to the classical parametric models assuming a Gompertz or Weibull distribution. If the number of pieces becomes large, the models show similar flexibility as semiparametric models. In many situations, it is reasonable to expect fitting parametric baseline intensity is unrealistic.

For the estimation methods, the expectation-maximization (EM) algorithm is the most general estimation procedure for the semiparametric models. Evaluating the conditional distribution for the frailty in the algorithm for which the numerical integration can be used with consideration to the frailty dimensions. MCMC methods are the general alternative to calculate this conditional distribution for high dimensional frailty terms. The penalized partial likelihood (PPL) approach and Markov chain Monte Carlo (MCMC) methods are mostly used to fit parametric frailty models. The PPL is not flexible enough to be used in high dimensions frailty models as well as a bootstrap technique is used to obtain an estimate of the frailty variance(see e.g., Liu and Huang 2008). These estimation methods are discussed in detail by Duchateau and Janssen (2008) and Therneau and Grambsch (2000). Different choices

of frailty distributions are possible. Gamma and lognormal distributions for the univariate frailty and multivariate Gaussian in case of multivariate random effects are commonly used.

Recently, Mazroui et al. (2013) present a shared multivariate frailty models for two types of recurrent events using penalized likelihood estimation where the baseline functions are approximated by a parametric piecewise constant and M-spline. Furthermore, their method could be difficult to converge with the data with few numbers of events.

Cook et al. (2010) describe a bivariate Poisson model where a Clayton copula function is used to model the only positive association between two gamma random effects. An EM algorithm with a numerical integration is used in the E-step to approximate the conditional distribution of the two marginal gamma frailties. However, it is quite difficult to be used, as a general solution, to deal with different cases and high dimensional random effects. Furthermore, It may be not useful when a complete inference for the subject-specific random effects or frailties is desired. Methods of the generalized linear models and survival analysis are used to estimate the parametric and semiparametric models, respectively and the bootstrap technique is used for the variance estimation.

In Chapter 3 we propose multi-type models. Gaussian copula function is used as an example to describe equal degrees of positive and negative dependence and Calyton copula for strong left tail dependence and relatively weak right tail dependence. Gamma and Gaussian distributions are introduced to model the marginal frailties and random effects. MCEM algorithms with MCMC sampling routines is used to evaluate the correlated frailty conditional distribution in the E-step of the algorithm. We consider adapting Louis' method for variance estimation which derives the form of the score vector and the matrix second derivatives and evaluate them at the last round samples of the MCEM algorithm.

A multivariate frailty model for the multi-type recurrent event data using a Monte Carlo EM algorithm is developed in Chapter 2. In this model, a multivariate Gaussian distribution is used to model the correlated random effects. The baseline intensity functions are modeled in a complete non-parametric way (different from the piecewise parametric model). We propose

an implementation of the MCEM algorithm to estimate the model parameters. In the E-step, a Metropolis-Hastings sampler is adapted to obtain the expectation of the conditional distribution for the random effects. Different from numerical integration, the Monte Carlo sampling is more flexible in dealing with multi-type events even when the number of event types is large, the number of events per subjects and number of subjects are small, and the censoring rate is high. At the same time it gives us the advantage of making inferences on the subject-specific random effects and frailties. In the M-step, maximum likelihood estimates of the fixed effects using expectation of numerical optimizations, the variance components, and nonparametric estimators of the baseline cumulative intensity functions are obtained. The observed information matrix for the estimated parameters is derived and calculated by using Louis' formula. The proposed model has a good performance for simulated datasets with both different censoring rates and different number of events per subject. We apply the developed methodology to the dataset collected as a part of the Nutritional Prevention of Cancer trial.

Chapter 3 presents copula based semiparametric multivariate frailty models for the multi-type recurrent event data with applications to a multi-type skin cancer. We generalize the multivariate Gaussian assumption of the frailty terms and allow the frailty distributions to have more features than the symmetric, unimodal properties of the normal density. More flexible approaches to modeling the correlated frailty using copula functions are introduced. Copula functions provide tremendous flexibility, especially in taking the advantages of a variety of choices for the marginal distributions and correlation structures. Gaussian and Clayton copulas with gamma and Gaussian margins are used to model the association between the random effects or frailties.

In Chapter 3, semiparametric intensity models for multi-type recurrent events based on a collaboration of the MCEM as a general option to estimate the unknown parameters of high-dimensional frailty models and copula functions for modeling the multivariate frailty are presented. Complete estimation procedures for fixed effects, nonparametric baseline intensity functions, and copula parameters are introduced. We propose appropriate martingale residu-

als that can be used in a graphical way to assess the model goodness of fit. Deviance residuals are calculated for model selection criteria. Louis' formula is used for the variance estimates and predictions of frailties are obtained. Performance of proposed models are evaluated by simulation studies. Applications are illustrated through a dataset collected from a clinical trial of patients with multi-type skin cancer.

4.2 Conclusions

In the applications of the methods, it is shown that using single-type models instead of a multi-type frailty model when there is significant dependence between the processes lead to unreliable estimates. In this case, the association between different event types needs to be taken into account in order to obtain accurate inferences. On the other hand, if no association exists between the processes, a more restricted model might be acceptable. The multivariate Gaussian assumption on the correlated frailty is restricted with the symmetric, unimodal properties of the Gaussian density and so may be unrealistic. This could raise concerns over the validity of parameter inferences. Copula functions provide a variety of choices for the marginal distributions and the correlation structure which can be used in the context of modeling the multivariate correlated frailty terms and random effects.

In a simple frailty structures, The PPL estimation method and MCEM with numerical integration and Laplace approximation techniques for the E-step work comparatively well but they are less truthful for studies with the number of events per subject and the number of subjects are small as well as for models with high-dimensional correlated frailties or random effects. On the other hand, the MCEM with MCMC algorithm in the E-step is considered a general estimation method regardless the of number of subjects, the number of events per subject, and the correlation structures of the frailties at the expensive computing price (Ripatti et al. 2002, Xu et al. 2009).

4.3 Future Work

In this dissertation, our considerations are mainly on developing general methodologies for modeling multi-type recurrent event data. In the proposed methods, MCEM is the used procedure to estimate the unknown parameters. Louis' formula is applied to obtain the information matrix. The baseline intensities are left unspecified. Multivariate Gaussian distribution and copula functions (Gaussian and Clayton) are applied to fit the multivariate correlated random effects (or frailty terms). The generality of the proposed models enables them to be adapted for the specific especial cases. For example, it is straightforward to extend the developed models to multivariate recurrent event types (more than two), time-varying coefficients with applications of several event types (Yu et al. 2014), and time-dependent and/or nonlinear covariates. These methods can be extended to use different combinations of copula functions and marginal distributions for the frailties and random effects.

In many chronic disease settings, the occurrence of event is associated with an elevated risk of death. Deaths are awkward in the analysis of recurrent events because they prevent further events. In addition, many subjects who experience multiple recurrences tend to withdraw or terminate from the study because of health related reasons. These withdrawals violate the assumed independent censoring. One could employ this work to handle the applications of multi-type recurrent events with the presence of death and/or dependent censoring. In some other situations, where recurrent events can be one of several distinct types with more than one type of terminal event process, we could also modify the proposed models. When competing risks exist for the terminal events, the models can be used to jointly model different types of terminal events and recurrent event processes.

The models introduced in this dissertation can be particularly appropriate when data are clustered at several hierarchical levels. For instance, when the application is based on recurrent infection times of patients from different hospitals or locations. In such cases it is important to estimate the parameters of interest as accurately as possible by taking into

account the hierarchical structure of the data. The MCEM algorithm can be adapted for nested frailty models with multi-type recurrent events (Rondeau et al. 2006). The proposed methods can also be used when uncorrelated or correlated random effects are used for the baseline intensity and treatment effect across centers or individuals (Ha et al. 2011, Vaida et al. 2000).

In MCEM algorithms, the M-steps are usually very efficient due to closed-form solutions, while the Monte Carlo E-steps often rely on MCMC routines. The problems caused by using MCMC algorithms include two main issues. First, the use of an MCMC algorithm in each E-step, leading to problems of slow convergence and requiring substantial experiences for convergence diagnosis, as a result, will greatly increase modeling effort. Second, the efficiency of these MCEM algorithms is heavily affected by these computationally intensive MCMC algorithms. Consequently, the automated MCEM presented in the context of generalized mixed models (Robert and Casella 2009, McLachlan and Krishnan 2007, Booth and Hobert 1999, Levine and Fan 2004) can be adapted for the current models, and for studying the efficacy of the different routines to reduce the cost of computing in the future work.

For assessing the goodness of fit and model selection criteria for the fitted multi-type modes, martingale and deviance residuals are used. Analysis of these residuals was applied to the BCC and SCC skin cancer types. While the adequateness can be tested by visual inspection of these residual, such approach is very subjective. It is attractive to develop more objective methods for model checking and selecting criteria (see e.g., Xu et al. 2009, Donohue et al. 2011).

Clustered and repeated measures data are very common in biomedical applications, for example when one or more variables are measured on each patient at a number of hospital visits, or when a number of questions are asked at a series of interviews. The generalized linear mixed model can be applied for fully parametric subject-specific inference for clustered or repeated measures responses in the exponential family (Gueorguieva 2001). The methods introduced in this dissertation can be adapted to deal with situations when multiple outcome

variables in the exponential family are present. The responses can be combined in a multi-type model by commanding a multivariate distribution for the variable-specific random effects. This allows maximum-likelihood estimation approaches to be extended to the multivariate case.

Bibliography

Booth, J. G. and Hobert, J. P.: 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.

Cook, R. J., Lawless, J. F. and Lee, K.-A.: 2010, A copula-based mixed poisson model for bivariate recurrent events under event-dependent censoring, *Statistics in Medicine* **29**(6), 694–707.

Donohue, M., Overholser, R., Xu, R. and Vaida, F.: 2011, Conditional akaike information under generalized linear and proportional hazards mixed models, *Biometrika* **98**(3), 685–700.

Duchateau, L. and Janssen, P.: 2008, *The Frailty Model*, Springer, New York.

Gueorguieva, R.: 2001, A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family, *Statistical Modelling* **1**(3), 177–193.

Ha, I. D., Sylvester, R., Legrand, C. and MacKenzie, G.: 2011, Frailty modelling for survival data from multi-centre clinical trials, *Statistics in Medicine* **30**(17), 2144–2159.

Levine, R. A. and Fan, J.: 2004, An automated (Markov chain) Monte Carlo EM algorithm, *Journal of Statistical Computation and Simulation* **74**(5), 349–360.

Liu, L. and Huang, X.: 2008, The use of Gaussian quadrature for estimation in frailty proportional hazards models, *Statistics in medicine* **27**(14), 2665–2683.

Mazroui, Y., Mathoulin-Pélissier, S., MacGrogan, G., Brouste, V. and Rondeau, V.: 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.

McLachlan, G. J. and Krishnan, T.: 2007, *The EM Algorithm and Extensions*, Vol. 382, Wiley-Interscience, New Jersey.

Ripatti, S., Larsen, K. and Palmgren, J.: 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 Casella, G.: 2009, *Introducing Monte Carlo Methods with R*, Springer, New York.

Rondeau, V., Filleul, L. and Joly, P.: 2006, Nested frailty models using maximum penalized likelihood estimation, *Statistics in Medicine* **25**(23), 4036–4052.

Therneau, T. M. and Grambsch, P. M.: 2000, *Modeling Survival Data: Extending the Cox Model*, Springer, New York.

Vaida, F., Xu, R. et al.: 2000, Proportional hazards model with random effects, *Statistics in Medicine* **19**(24), 3309–3324.

Xu, R., Gamst, A., Donohue, M., Vaida, F. and Harrington, D. P.: 2009, Using profile likelihood for semiparametric model selection with application to proportional hazards mixed models, *Statistica Sinica* **19**(2), 819–842.

Yu, Z., Liu, L., Bravata, D. M. and Williams, L. S.: 2014, Joint model of recurrent events and a terminal event with time-varying coefficients, *Biometrical Journal* **56**(2), 183–197.