2016

Semiparametric Joint Dynamic Modeling of a Longitudinal Marker, Recurrent Competing Risks, and a Terminal Event

Piaomu Liu
University of South Carolina

Semiparametric Joint Dynamic Modeling of a Longitudinal Marker, Recurrent Competing Risks, and a Terminal Event

by

Piaomu Liu

Bachelor of Arts
Carleton College 2009

Master of Science
The University of Iowa 2011

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Statistics

College of Arts and Sciences

University of South Carolina

2016

Accepted by:

Edsel Peña, Major Professor

Yanyuan Ma, Committee Member

Lianming Wang, Committee Member

Alexander McLain, Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

Acknowledgments

I acknowledge support from the National Science Foundation (NSF) Grant DMS 1106435 and National Institute of Health (NIH) Grant 1P30GM103336-01A1. I also thank the Department of Statistics, USC for the assistantship support it has provided me during my graduate studies.

I would like to express my deepest gratitude for my advisor, Dr. Edsel A. Peña, for his guidance throughout my graduate study at USC - Columbia. Your contribution to research in recurrent event and lifetime data analysis, as well as your passion for the statistical science and discovery have been some of the greatest inspirations I have to pursue my academic interests. I believe that none of the progress I have made in research would be possible without your encouragement and your belief in my potential of growth and success. I am so grateful for being your student, for I have also learned from you that being a teacher is be excited about students' growth, and to love learning.

To my thesis committee members, Dr. Yanyuan Ma, Dr. Lianming Wang, and Dr. Alexander McLain, I would like to thank you for your insightful comments and constructive criticism on the thesis. I have had a tremendous learning experience from all of you.

I would like to thank the Department of Statistics, the College of Arts and Sciences, and the Graduate School at USC for providing financial sport for my conference trips. I would also like to extend my immeasurable appreciation to all the faculty and staff in the department of Statistics.

To my mom and aunt Liya, I have been very blessed to have your unconditional

love and support over the years. You have been my best cheerleaders, and you are the heroines in my life.

Abstract

The joint modeling framework has found extensive applications in cancer and other biomedical research. For example, recent initiatives and developments in precision medicine call for appropriate prognostic tools to assist individualized or personalized approaches in cancer diagnosis and treatment. Data generated by clinical trials and medical research often include correlated longitudinal marker measurements and time-to-event information, which are possibly a recurrent event, competing risks, and a survival outcome. Primary interests of joint modeling include the association between the longitudinal marker measurements and time-to-event data, as well as predictions of survival probabilities of new observational units from the same population.

The dissertation deals with joint dynamic modeling of a longitudinal marker, recurrent competing risks, and a terminal event. To tackle the problem of simultaneously modeling three types of data processes, we begin by proposing joint dynamic models of recurrent competing risks (RCR) and a terminal event (TE). We adopt the counting process approach of survival analysis to specify the joint models, where history of data, or filtration is considered. Intensity processes of the recurrent competing risks (RCR) and the terminal event (TE) also includes fixed covariates, past event occurrences as well as impact of possible interventions. Impact of past event occurrences is important to be accounted for in the intensity processes because it is reasonable to assume that if a unit has experienced a certain type of recurrent event many times up to time t, the probability of a new event occurrence of this unit could either increase or decrease compared to those who have experienced fewer event occurrences, depending on the context of the data. Consequently, as the differ-

ent aspects of the intensity processes change over time, the intensity processes evolve dynamically.

A frailty variable, or latent variable, which is unobserved, is often used to induce association in survival analysis. We introduce a frailty variable Z into the joint dynamic model proposed previously. For parameter estimation, we propose semi-parametric inference procedures for the joint models with no-frailty and with frailty cases. Nelson-Aalen type of estimators are derived for baseline hazards, and partial likelihoods are obtained to estimate the unknown finite-dimensional parameters in the joint models. We illustrate the inference procedures on simulated datasets. Simulation studies with moderate sample sizes are performed to understand large/finite-sample properties of the proposed estimators. We also address the issue of predicting terminal event (TE) probabilities of a new unit from the same population, and provide an example for the simulated population.

When correlated longitudinal marker measurements, recurrent competing risks event occurrences, and status of a terminal event are the data of interest, we propose a joint dynamic model that link the three data processes together. The joint dynamic model consists of the longitudinal marker (LM) submodel, the recurrent competing risks (RCR) submodal, and the terminal event (TE) submodel. For each observational unit, the marker is measured at irregularly spaced times within the monitoring period or until the terminal event happens. The longitudinal marker submodel is a mixed model with a fixed linear time trend. A frailty variable, or the random effect in the submodel induces both within-subject correlation as well as the association between the longitudinal marker process and the recurrent competing events. Additionally, this random variable induces correlation between the longitudinal marker values and the terminal event. The joint models capture dependence structure of the data from the following two aspects: firstly, past longitudinal marker history affect the intensity process of the recurrent competing events, including that of the terminal event; at the same time, past event occurrences of RCR also affect the mean process of the longitudinal marker. Secondly, the frailty variable represents all other unobserved variables that induce associations between the different processes (LM, RCR, and TE). Built upon the aforementioned joint dynamic model of RCR and TE, the proposed joint dynamic models (of LM, RCR and TE) possess similar dynamic nature. A semiparametric inference procedure involving an EM algorithm is proposed. Future research activities for this joint model are indicated.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In many engineering, biomedical, public health, economics, sociology, and psychology settings, of main interest is the time to occurrence of some terminal event for an experimental unit (e.g., patient, experimental mouse, cell line, computer chip, medical device, diagnostic machine, county, city, stock, etc.). Terminal events typically possess a detrimental character such as death, onset of a disease, metastasis of cancer, outbreak of an epidemic, occurrence of a stroke, occurrence of a heart attack, failure of a medical device, among others. However, in some cases, it may also have a beneficial nature such as when getting cured of a disease, being released from a hospital, containment of an epidemic, among others.

Studies dealing with the times to a terminal event often encounter the phenomenon of right-censoring and/or left-truncation (cf. [28], [11], [2]) where for some units the exact time to event occurrence is not observed. Early work in survival analysis dates back to [29], where a nonparametric estimator of survival function is proposed.

In addition, it is of importance to take into consideration concomitant variables that could be impacting the time to occurrence of the terminal event. Popular models for the time to occurrence of a terminal event such as the Cox proportional hazards model (CPHM) ([10]), the accelerated failure time model (AFTM) ([28]), and Aalen's additive hazards model ([2]) incorporate both right-censoring and/or left-truncation and the covariate information. Statistical analysis approaches for these models could be found, for instance, in the books by (cf. [28], [11], [17], [5], and [2]).

Many works modeling recurrent event data have emerged in recent years. References can be found in (cf. [47], [6], [64], [3], [62], [44], [42], [39], [43]), among others. Interests of statistical inference in these papers often lie in the stochastic mechanism that governs the occurrences of the recurrent event. Relevant random quantities like the gap-times (or inter-event times), or the distribution of the kth (k = 1,2, $\cdots$) occurrence of the recurrent event are studied, and many models also take into account information of covariates. The impact of interventions after each event occurrence and the effect of accumulating event occurrences on the unit were taken into account leading to general classes of dynamic recurrent event models (cf. [42]; [43]). There are also papers dealing with the joint modeling of a terminal event and recurrent events such as those by [18]. These joint models could have important practical implications since one may utilize information regarding the occurrences of the recurrent event to improve knowledge of the occurrence of the terminal event, in particular in the context of personalized and/or precision medicine approaches. In monitoring occurrences of a recurrent event, there is usually a random window of observation, with the upper limit of this observation window determined by either an administrative constraint, a censoring time, or the time to occurrence of a terminal event. This induces a complex probabilistic structure on the observed random entities induced by the so-called sum-quota accrual scheme and which leads to size-biased sampling phenomena.

Real life applications often involve several competing recurrent events. Such applications include analyses of a variety of datasets (cf. [46], [14], [57], [56]). Instead of considering only one recurrent event, there could be several recurrent events which all have an impact on the time to occurrence of the terminal event. The importance of these competing recurrent events in terms of the terminal event could vary, and these competing recurrent events could also be affecting each other. Similar to a single recurrent event, interest of analyzing recurrent competing risks data focuses on the stochastic processes that generate the observed recurrent competing events.

Typically, the random quantities associated with the stochastic processes include estimating cumulative incidence functions (cf. [14], [15], [31]), gap time distributions (cf. [50]), distribution of event counts (cf. [18]), and inferences on covariate information for both fixed (cf. [14], [10], [43]) and time-varying cases (cf. [46], [16], [32]).

Joint modeling approaches have been proposed to simultaneously study recurrent event and terminal event processes together (cf. [24], [34], [53], [13]). In these cases, the recurrent event is correlated with the terminal event of interest. We hope to utilize information from recurrent competing risks to understand a terminal event process, and propose a *joint* dynamic model to *simultaneously* model these processes together in Chapter 2 and Chapter 3 . The proposed models possess a dynamic feature owing to the stochastic modeling framework, making the models more realistic and easy to interpret. Parameter estimation procedures for both no-frailty and frailty cases are proposed . Predictions of a new unit's terminal event time using information from recurrent competing risks are also discussed. Potential applications of the model include but not limited to personalized medicine, precision medicine, prediction of adverse health outcomes, and prediction of destructive event in nature like an earthquake and in an economy such as recession or dropping of Dow-Jones Industrial Average by more than 5% based on information from some recurrent competing risks.

In biomedical and clinical research, a longitudinal marker is also often correlated with a recurrent event, or recurrent competing risks. In such situations, joint models that simultaneously link a longitudinal marker, or multiple longitudinal markers and a recurrent (competing) risks are constructed (cf. [22], [65], [13], [25], [33], [7]). The joint modeling approach has also been applied to data resulting from many biomedical and clinical studies where longitudinal measurements are associated with a survival outcome. One of the earlier motivations of the joint modeling framework came from HIV research and a key question of interest was to understand the association between a CD4 load profile and time to disease progression (cf. [40], [58], [66], [59], and [60]).

The models deal with the association between the longitudinal marker process and the recurrent (competing) event processes through latent classes (cf. [22]), random effects (cf. [33]) or latent zero-mean Gaussian processes (cf. [65], [13]). In addition to longitudinal measurements, other potentially useful information available from observations often include baseline covariates, demographic information and treatment assignment, which are commonly collected and recorded in clinical research (cf. [60], [21]). When the longitudinal marker is correlated with a survival outcome, joint modeling framework shows superiority over modeling the two processes separately.

Goals of the longitudinal marker and survival outcome joint models usually focus on associations between the marker(s) and the survival outcome, as well as prediction of survival times (cf. [63], [59], [55], [49], [60], [9], [23], [61], [26]). In the two submodels that comprise of the joint model, dependence between the longitudinal measurements and survival outcome is often realized through either the past history of the longitudinal marker, or a shared random effect (cf. [60]). Shared random effect models (cf. [34], [36], [38]) are also common approaches which use shared random variable, frailty variable or latent classes induce association between the longitudinal marker and survival outcome.

In recent years, the joint modeling framework is seen in developing joint models for longitudinal marker, recurrent (competing) risks and a survival outcome, and a terminal event as well (cf. [35], [30], [8]). For instance, in [35], a joint model is constructed using a dataset where CD4 counts as well as occurrences of a recurrent disease and death were recorded on patients infected with HIV. It was observed that longer survival of patients are associated with higher CD4 counts and lower rate of disease re-occurrence. At the same time, the longitudinal measurements are also associated with rate of disease occurrences. To model the correlated processes, the authors used random effects to induce associations. However, compared to joint models of LM/RCR or RCR/TE, the type of joint model LM/RCR/TE has been attempted less.

Within the joint modeling framework, dynamic joint modeling build the history of the longitudinal measurements, and often takes into account past history of the longitudinal measurements when updating the hazard of a recurrent event. Since the longitudinal marker measurements are associated with the terminal event, the past history of the longitudinal maker also affects the hazard of the terminal event (cf. [48], [51], [37], [52], [8]). Consequently, the hazard of terminal event also updates as the longitudinal marker process evolves. In biomedical applications, the dynamic joint modeling provides a way to predict survival probabilities of new observational units from the same population. (cf. [51], [8])

In Chapter 4, we propose a dynamic joint model that link longitudinal measurements, correlated recurrent competing risks and a terminal event together. The joint model consists of the longitudinal marker (LM) submodel, the recurrent competing risks (RCR) submodel, and the terminal event (TE) submodel. The joint model takes into account past history of the longitudinal marker when modeling the intensity processes of recurrent competing risks (RCR). History of the longitudinal marker also affects the intensity process of the terminal event (TE). Because both the longitudinal marker (LM) and recurrent competing risks (RCR) are associated with the TE, a frailty variable is introduced to model the between-process dependence. The frailty variable also induces association between the RCR and TE. By modeling the dependence structure among the processes, we hope that the joint dynamic model will strengthen the analyses of the LM, RCR and TE. Additionally, built upon the joint dynamic model of RCR and TE in Chapter 2 and Chapter 3, the proposed model contributes to research in joint modeling by allowing the intensity process of recurrent competing risks evolve while taking into consideration of the impact of past event history, and the possible impact of interventions. Parameter estimation procedure is proposed and demonstrated on a simulated dataset. Areas of applications of the model include but not limited to biomedical, clinical and reliability type of research. The dynamic aspect of the model will make the model applicable in precision medicine. We conclude the dissertation in Chapter 5 with some remarks and future studies.

Chapter 2

Joint Dynamic Modeling of Recurrent Competing Risks and a Terminal Event

2.1 Introduction

A General Class of Dynamic Models

In [45] and [44], nonparametric estimation procedures and asymptotic properties of estimators of recurrent event data are introduced. The papers extended the approach that make use of the martingale property of a counting process to estimate the intensity process. A Nelson-Aalen type of estimator is described to estimate the baseline hazard and a Kaplan-Meier type of estimator is described to estimate the baseline survival functions.

To incorporate possible interventions after each recurrent event occurrence and impact of covariate information, a general class of dynamic models for recurrent event data was proposed in [42]. Baseline hazard of the intensity process are estimated using estimators in [45] and [44] and captures the dynamic nature of the relevant stochastic processes as time passes. A comprehensive review about this class of dynamic models can be found in [41]. A general class of semi-parametric models of recurrent event data is also developed in [43].

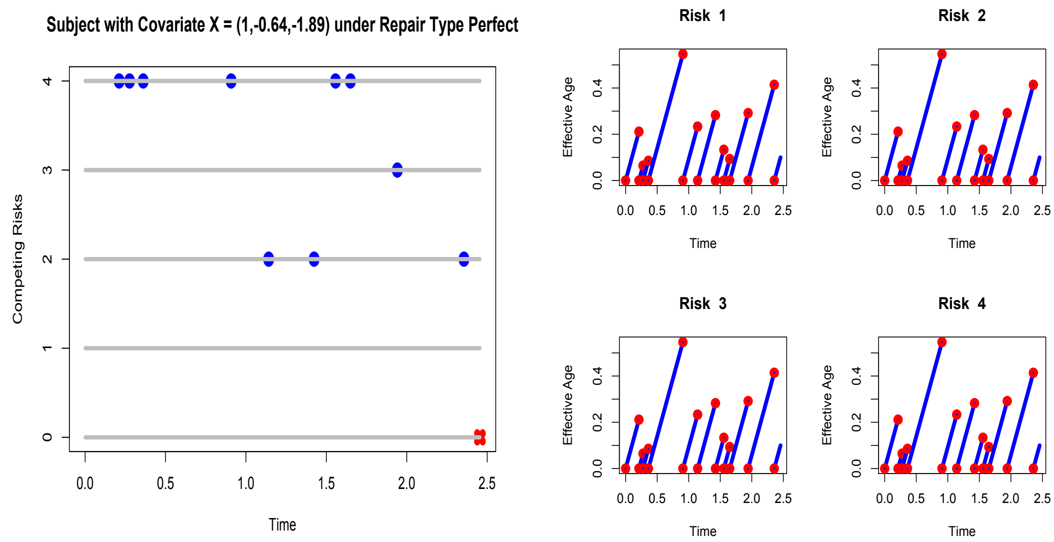

Figure 2.1: Data of a Single Unit Observed Under Perfect Repair (*Left*: Data of a Single Unit *Right*: Effective Ages of Competing Risks under Perfect Repair)

Effective Age

Effective age processes can be used to model impact of interventions (cf. [20]). An intervention could be repair of failed machine components, or surgery doctors perform on a patient to treat certain diseases. Since there are multiple recurrent competing risks involved in our joint model, the effective age processes are risk-specific. Suppose the qth risk of unit i is observed to experience $j = 1, 2, \cdots, N^\dagger_{qi}(s-)$ events on monitoring interval [0, s]. An effective age process is a process whose sample paths are nonnegative, increasing, left-continuous, and differentiable on the time interval between two successive event occurrences $(S_{j-1}, S_j]$. For the proposed joint model, we consider the following two kinds of effective age processes

$$\mathcal{E}_q(v)^{PER} = v - S_{N^\dagger_\bullet(v-)}; \qquad \mathcal{E}_q(v)^{PAR} = v - S_{qN^\dagger_q(v-)}$$

In Figure 2.1, event occurrences are plotted for a single observational unit under perfect repair where Q = 4. After an event happens, interventions take place. Under perfect repair, effective ages are set to 0 for all recurrent competing risks. Therefore, all units start anew after each event occurrence. In Figure 2.2, data are plotted for

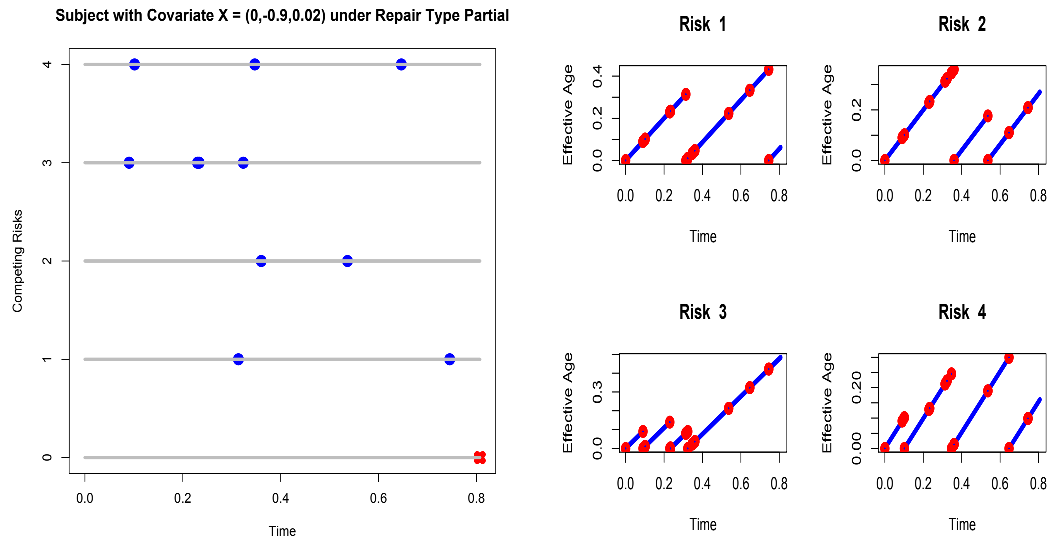

Figure 2.2: Data of a Single Unit Observed Under Partial Repair (*Left*: Data of a Single Unit *Right*: Effective Ages of Competing Risks under Partial Repair)

a single observational unit under partial repair. In this case, an intervention only happens to the risk to which an event has just happened. Resetting effective age to 0 does not happen to other risks. Notice that the first event under partial repair is risk 4, and the right panel of Figure 2.2 indicates that only risk 4 resets its effective age to 0.

Some Stochastic Processes

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be some probability space. Define $\mathbf{F} = \{\mathcal{F}_s | 0 \leq s \leq s^*\}$ a history or filtration on the same probability space. $N^\dagger_{qi}(s)$ and $N^\dagger_{0i}(s)$ are counting processes and $Y^\dagger_i(s)$ are predictable processes with respect to $\mathbf{F}$. For the ith unit, we first describe the stochastic processes:

1. $\{N^\dagger_{qi}(s) : s \geq 0\}$: counting process for the qth competing risk.

2. $\{N^\dagger_{0i}(s) : s \geq 0\}$: counting process for the terminal event.

3. $\{Y^\dagger_i(s) : s \geq 0\}$: at-risk process.

4. $\{\mathcal{E}_{qi}(s) : s \geq 0\}$: the effective age process.

2.2 Data

For $i = 1, 2, \cdots, n$, the observables are

$$\mathbf{D}_i \;=\; \{(N^{\dagger}_{qi}(s), N^{\dagger}_{0i}(s), Y^{\dagger}_i(s), \mathcal{E}_{qi}(s)) : s \geq 0, X_i, \tau_i\}.$$

τ_i is the subject specific random censoring time. We assume that τ_i is independent of $N^{\dagger}_{qi}(s)$ and $N^{\dagger}_{0i}(s)$. In Figure 2.3, observations for 50 independent units are plotted. All observed Q recurrent competing risks for each unit are plotted on the same line using different colored symbols. Censoring is indicated using a green cross and a red cross is employed to show when a unit experiences the terminal event by the end of the monitoring period.

For the ith unit, and $q = 1, 2, \cdots, Q$, our proposed model is

$$\begin{aligned} \mathcal{L}_c(v, \Theta) &\equiv \mathbf{P}\left\{\bigcap_{q=1}^{Q}[dN^{\dagger}_{qi}(v) = dn_{qi}(v)]; [dN^{\dagger}_{0i}(v) = dn_{0i}(v)]|\mathcal{F}_{v-}\right\} \\ &= \left\{\prod_{q=1}^{Q}[dA_{qi}(v)]^{dn_{qi}(v)}[1 - dA_{qi}(s)]^{1-dn_{qi}(v)}\right\} \\ &\times \left\{[dA_{0i}(v)^{dn_{0i}(v)}[1 - dA_{0i}(v)]^{1-dn_{0i}(dv)}\right\} \end{aligned} \tag{2.1}$$

with

$$\begin{aligned} A^{\dagger}_{qi}(s) &= \int_0^t Y^{\dagger}_i(v)\rho_q(\mathbf{N}^{\dagger}_i(v-); \alpha_q)\exp(X_i\beta_q)\lambda_{q0}(\mathcal{E}_{qi}(v))dv \\ A^{\dagger}_{0i}(s) &= \int_0^s Y^{\dagger}_i(v)\rho_0(\mathbf{N}^{\dagger}_i(v-); \gamma)\exp(X_i\beta_0)\lambda_0(v)dv \end{aligned} \tag{2.2}$$

where $dn_{qi}(v), dn_{0i}(v) \in \{0, 1\}$ and $\sum_{q=1}^{Q} dn_{qi}(v) + dn_{0i}(v) \leq 1$. And $\mathbf{N}^{\dagger}_i(v-) = \left(N^{\dagger}_{1i}(v-), N^{\dagger}_{2i}(v-), \cdots, N^{\dagger}_{Qi}(v-)\right)^T$ is a $Q \times 1$ vector of recurrent competing risks, $\{\mathcal{E}_{qi}(s) : s \geq 0\}$ is the effective age process, and X_i is a p $\times$ 1 fixed covariate vector.

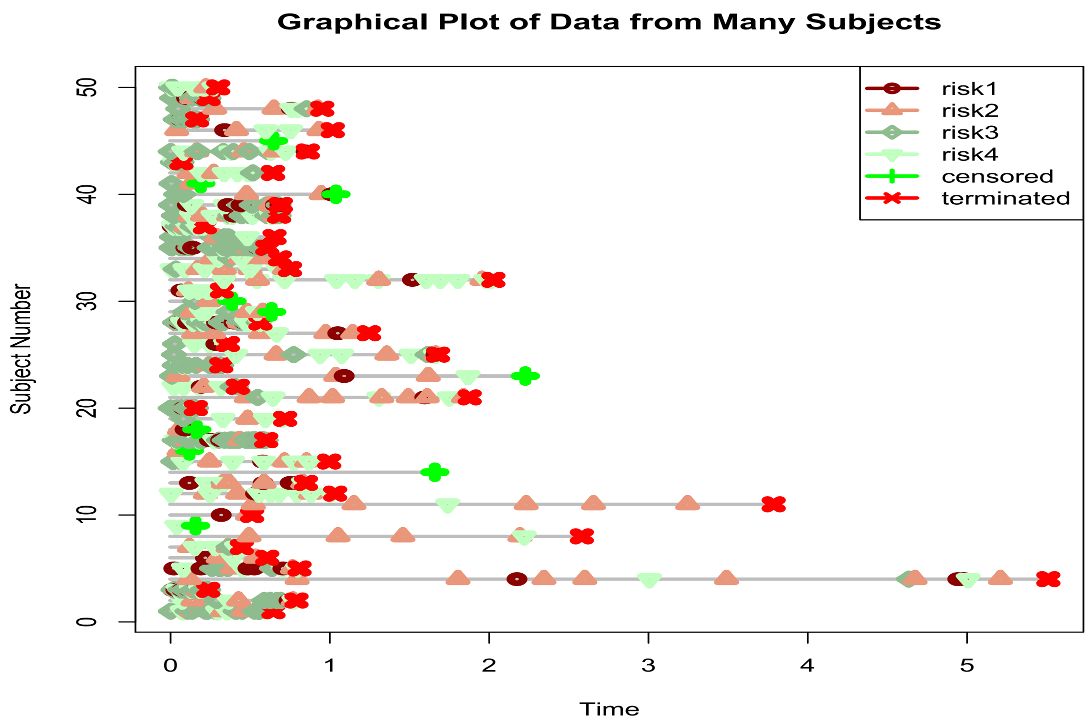

Figure 2.3: An example of a data set with recurrent competing risks and a terminal event for $n = 50$ subjects and $Q = 4$ types of competing recurrent events. The four different colored symbols other than '$\times$' and '+' indicate occurrences of the different recurrent competing risks. A symbol '$\times$' indicates occurrence of the terminal event, whereas a '+' indicates that the terminal event is not observed but instead τ, the time of the end of the monitoring period, is observed, so the terminal event time is right-censored.

For ease of notation, we also define the following

$$a^{\dagger}_{qi}(v)dv = Y^{\dagger}_i(v)\rho_q(\mathbf{N}^{\dagger}_i(v-);\alpha_q)\exp(X_i\beta_q)\lambda_{q0}(\mathcal{E}_{qi}(v))dv \tag{2.3}$$

$$a^{\dagger}_{0i}(v)dv = Y^{\dagger}_i(v)\rho_0(\mathbf{N}^{\dagger}_i(v-);\gamma)\exp(X_i\beta_0)\lambda_0(v)dv \tag{2.4}$$

2.3 Model Description

Let $\Theta^{(nofrail)}$ be the vector of unknown parameters. On $[0, s^*]$, unknown parameters of infinite dimensions $\Lambda_{q0}(s)$ and $\Lambda_0(s)$ are the cumulative baseline hazards for the recurrent competing risks, and the terminal event, respectively. The $\Theta^{(nofrail)}$ vector is

$$\begin{aligned}\Theta^{(nofrail)} =& \{\gamma, \beta_0, (\alpha_q, \beta_q; q = 1, 2, \cdots, Q)\} \\ & \cup\{(\Lambda_{q0}(s), q = 1, 2, \cdots, Q), \Lambda_0(s) | 0 \leq s \leq s^*\}\end{aligned} \tag{2.5}$$

where for $q = 1, 2, \cdots, Q, \gamma$ and α_q are $Q \times 1$ vectors; and β_0 and β_q are $p \times 1$ vectors. Consequently, the full likelihood of the model of n observational units is

$$\mathcal{L}_F(\Theta, s) = \prod_{i=1}^{n} \prod_{v=0}^{s} \mathbf{P}\{\bigcap_{q=1}^{Q} [dN_{qi}^{\dagger}(v) = dn_{qi}(v)]; [dN_{0i}^{\dagger}(v) = dn_{0i}(v)] | \mathcal{F}_{v-}\} \tag{2.6}$$

Another important dynamic aspect of the joint model is to include impact of past event occurrences on the intensity process. Existing work has shown the significance of past event occurrences on the recurrent competing risks (cf. Latouche et. al (2015)). It is reasonable to assume that high event occurrences of certain risk right before time t increases the chance of a new event happening during a small time interval $[t, t+dt)$ conditional on $\mathcal{F}_{t-}$. The ρ functions ρ_q and ρ_0 incorporate the impact of past event occurrences into the intensity processes for a single unit are

$$\begin{aligned}\rho_q(\mathbf{N}_i^{\dagger}(t-); \alpha_q) &= 1 + \mathbf{N}_i^{\dagger}(t-)\alpha_q \\ \rho_0(\mathbf{N}_i^{\dagger}(t-); \gamma) &= \exp(\mathbf{N}_i^{\dagger}(t-)^T \gamma)\end{aligned} \tag{2.7}$$

The γ are Q $\times$ 1 vectors of coefficients. One may interpret the α_q coefficients as follows: if $\alpha_q > 0$, past event occurrences increases the instantaneous probability of a new event for the qth risk; if $\alpha_q < 0$, past event occurrences decreases the instantaneous probability of a new event occurrence for the qth risk; when $\alpha_q = 0$, past event occurrences does not impose further impact on the qth risk's instantaneous probability. Elements of vector γ have similar interpretations of their impact on the intensity process of the terminal event.

Generalized At-risk Process

The observable effective age processes for the Q risks are $\{\mathcal{E}_q(s) : s \geq 0\}, q = 1, 2, \cdots, Q$. In [44], a doubly-indexed process is defined to utilize the martingale approach in [1] and [19]. The paper extends work in [54] and proposes a doubly-indexed process as in equation 2.8 operating on both the calendar times and gap times. To make inferences about $\Lambda_{q0}(s)$ at any given calendar time $s \geq 0$ instead of $\Lambda_{q0}(\mathcal{E}_{qi}(s))$, we adopt the doubly-indexed process approach. Under this framework, we define doubly-indexed processes for stochastic processes $N_q^\dagger(s), M_q^\dagger(s)$ and $A_q^\dagger(s), s \geq 0, q = 1, 2, \cdots, Q$:

Therefore, for unit i, the stochastic processes become

$$\begin{aligned} Z_{qi}(s,t) &= I\{\mathcal{E}_{qi}(s) \leq t\} && (2.8) \\ N_{qi}(s,t) &= \int_0^s I(\mathcal{E}_{qi}(v) \leq t) N_{qi}^\dagger(dv) = \int_0^s Z_{qi}(v,t) N_{qi}^\dagger(dv) && (2.9) \\ A_{qi}(s,t) &= \int_0^s I(\mathcal{E}_{qi}(v) \leq t) A_{qi}^\dagger(dv) = \int_0^s Z_{qi}(v,t) A_{qi}^\dagger(dv) && (2.10) \\ M_{qi}(s,t) &= N_{qi}(s,t) - A_{qi}(s,t) = \int_0^s Z_{qi}(v,t) M_{qi}^\dagger(dv) && (2.11) \end{aligned}$$

To estimate baseline hazards, we also derive a doubly-indexed process for each at-risk process:

Proposition 1.

$$Y_{qi}(s,w) = \sum_{j=1}^{N_{qi}^\dagger\left((s\wedge\tau_i)-\right)+1} I\left\{w \in \left(\mathcal{E}_{qi}(S_{ij-1}), \mathcal{E}_{qi}(S_{ij})\right]\right\} \frac{\kappa_{qi}(\mathcal{E}_{qij}^{-1}(w))}{\mathcal{E}'_{qi}(\mathcal{E}_{qij}^{-1}(w))},$$

where $\kappa_{qi}(\mathcal{E}_{qij}^{-1}(w)) = \rho_q(N_{qi}^\dagger(\mathcal{E}_{qij}^{-1}(w)-); \alpha_q) \exp(X_i^T \beta_q)$*, and* $\mathcal{E}_{qij}^{-1}(.)$ *is the inverse of* $\mathcal{E}_{qij}(v) = \mathcal{E}_{qi}(v) I\{v \in (S_{ij-1}, S_{ij}]\}$.

A Semiparametric Approach

In this section, we develop a semi-parametric inference procedure to estimate $\Theta^{(nofrail)}$. We first obtain expressions for baseline hazards given finite-dimensional parameters. Note that $M_{qi}(s,t)$ in section 2.4 is a martingale with respect to calendar time s given an effective age t. For a doubly-indexed counting process $N_{qi}(s,t)$, we obtain an identity

$$M_{qi}(s,t) = N_{qi}(s,t) - \int_0^t Y_{qi}(s,w)\Lambda_{q0}(dw)$$

Summing over all units, we obtain

$$\sum_{i=1}^{n} M_{qi}(s,dw) = \sum_{i=1}^{n} N_{qi}(s,dw) - \sum_{i=1}^{n} Y_{qi}(s,w)\Lambda_{q0}(dw)$$

Since $\mathbf{E}[\sum_{i=1}^{n} M_{qi}(s,dw)] = 0$, we have $\mathbf{E}\left[\sum_{i=1}^{n} N_{qi}(s,dw)\right] = \mathbf{E}\left[\sum_{i=1}^{n} Y_{qi}(s,w)\Lambda_{q0}(dw)\right]$. Define the aggregated at-risk process for $q = 1, 2, \cdots, Q$ as

$$\begin{aligned} S_{q0}(s,w|\alpha_q,\beta_q) &= \sum_{i=1}^{n} Y_{qi}(s,w|\alpha_q,\beta_q) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{N_{qi}^{\dagger}[(s\wedge\tau_i)-]+1} I[w \in \left(\mathcal{E}_{qi}(S_{ij-1}), \mathcal{E}_{qi}(S_{ij})\right)] \frac{\kappa_{qi}(\mathcal{E}_{qij}^{-1}(w))}{\mathcal{E}_{qi}'(\mathcal{E}_{qij}^{-1}(w))} \end{aligned}$$

where $\kappa_{qi}(\mathcal{E}_{qij}^{-1}(w)) = \rho_q(N_{qi}^{\dagger}(\mathcal{E}_{qij}^{-1}(w)-);\alpha_q)\exp(X_i^T\beta_q)$.

Consequently, given the finite - dimentional parameters, an Aalen-Nelson type of "estimator" for recurrent competing risks cumulative baseline hazards is

$$\hat{\Lambda}_{q0}(s,t|\alpha_q,\beta_q) = \int_0^t \frac{\sum_{i=1}^{n} N_{qi}(s,dw)}{S_{q0}(s,w|\alpha_q,\beta_q)} \tag{2.12}$$

The expression of $\hat{\Lambda}_{q0}(s,t|\alpha_q,\beta_q)$ above involves unknown parameters α_q and β_q. Therefore, to obtain the cumulative baseline hazard estimator, the unknown finite-dimensional parameters are to be estimated, and then plugged in the expression in 2.12. For the terminal event process, since no effective age is involved

$$\begin{aligned} M_i^{\dagger}(s) &= N_i^{\dagger}(s) - \int_0^s Y_i^{\dagger}(t)\lambda_0(t)\rho_0(\mathbf{N}_i^{\dagger}(t-);\gamma)\exp(X_i^T\beta_0)dt \\ &= N_i^{\dagger}(s) - \int_0^s Y_i^{\dagger}(t)\rho_0(\mathbf{N}_i^{\dagger}(t-);\gamma)\exp(X_i^T\beta_0)\Lambda_0(dt) \end{aligned}$$

We also define the aggregated at-risk process as

$$S_0(v|\gamma, \beta_0) = \sum_{i=1}^{n} Y_i^{\dagger}(v|\gamma, \beta_0) = \sum_{i=1}^{n} I[(\tau_i \wedge S_i) \geq v]\kappa_{0i}(v)$$

where $\kappa_{0i}(v) = \rho_0(N_i^{\dagger}(v-); \gamma) \exp(X_i^T \beta_0)$ with S_i being time-to-terminal event. Given the finite-dimentional parameter, we obtain an Aalen-Nelson type of "estimator" for the terminal event cumulative baseline hazard

$$\hat{\Lambda}_0(t|\gamma, \beta_0) = \int_0^t \frac{\sum_{i=1}^n N_{0i}^{\dagger}(dv)}{S_0(v|\gamma, \beta_0)} \tag{2.13}$$

Similar to $\hat{\Lambda}_{q0}(s, t|\alpha_q, \beta_q)$ in equation 2.12, $\hat{\Lambda}_0(t|\gamma, \beta_0)$ will be considered as an estimator after estimates of the finite-dimensional parameters are computed and plugged in the above expression.

Define $T_i = \min(\tau_i, S_i)$, to estimate the finite-dimentional parameters, we obtain the partial likelihood process by substituting baseline hazards expressions using equation 2.12 and equation 2.13 in the full likelihood (see equation 2.6).

$$\begin{aligned}\mathcal{L}_p(\Theta) &= \prod_{i=1}^{n} \left\{ \prod_{q=1}^{Q} \prod_{j=1}^{N_{qi}^{\dagger}(\tau_i)} \left[\frac{\kappa_{qi}(S_{ij})}{S_{q0}(s, \mathcal{E}_{qi}(S_{ij})|\alpha_q, \beta_q)} \right]^{dN_{qi}^{\dagger}(S_{ij})} \right\} \\ &\times \left\{ \left[\frac{\kappa_{0i}(T_i)}{S_0(T_i|\gamma, \beta_0)} \right]^{dN_{0i}^{\dagger}(T_i)} \right\}\end{aligned}$$

Estimating equations of finite-dimensional parameters are

$$\sum_{i=1}^n \int_0^{\tau_i} \left[\frac{\frac{\partial}{\partial \alpha_q} \rho_q(N_{qi}^{\dagger}(v-); \alpha_q))}{\rho_q(N_{qi}^{\dagger}(v-); \alpha_q))} - \frac{\frac{\partial}{\partial \alpha_q} S_{q0}(s, \mathcal{E}_{qi}(v)|\alpha_q, \beta_q)}{S_{q0}(s, \mathcal{E}_{qi}(v)|\alpha_q, \beta_q)} \right] N_{qi}^{\dagger}(dv) = 0$$

$$\sum_{i=1}^n \int_0^{\tau_i} \left[X_i - \frac{\frac{\partial}{\partial \beta_q} S_{q0}(s, \mathcal{E}_{qi}(v)|\alpha_q, \beta_q)}{S_{q0}(s, \mathcal{E}_{qi}(v)|\alpha_q, \beta_q)} \right] N_{qi}^{\dagger}(dv) = 0$$

$$\sum_{i=1}^n \int_0^{\tau_i} \left[\frac{\frac{\partial}{\partial \gamma} \rho_0(N_i^{\dagger}(v-); \gamma)}{\rho_0(N_i^{\dagger}(v-); \gamma)} - \frac{\frac{\partial}{\partial \gamma} S_0(v|\gamma, \beta_0)}{S_0(v|\gamma, \beta_0)} \right] N_{0i}^{\dagger}(dv) = 0$$

$$\sum_{i=1}^n \int_0^{\tau_i} \left[X_i - \frac{\frac{\partial}{\partial \beta_0} S_0(v|\gamma, \beta_0)}{S_0(v|\gamma, \beta_0)} \right] N_{0i}^{\dagger}(dv) = 0$$

To solve for finite-dimensional parameters, we implement a Newton-Ralphson algorithm. PLEs of the recurrent competing risks ($q = 1, 2, \cdots, Q$) and terminal event processes of baseline survivor functions are

$$\hat{\bar{F}}_{q0}(s, t) = \prod_{w=0}^{t} [1 - \hat{\Lambda}_{q0}(s, dw)]; \qquad \hat{\bar{F}}_0(t) = \prod_{w=0}^{t} [1 - \hat{\Lambda}_0(dw)]$$

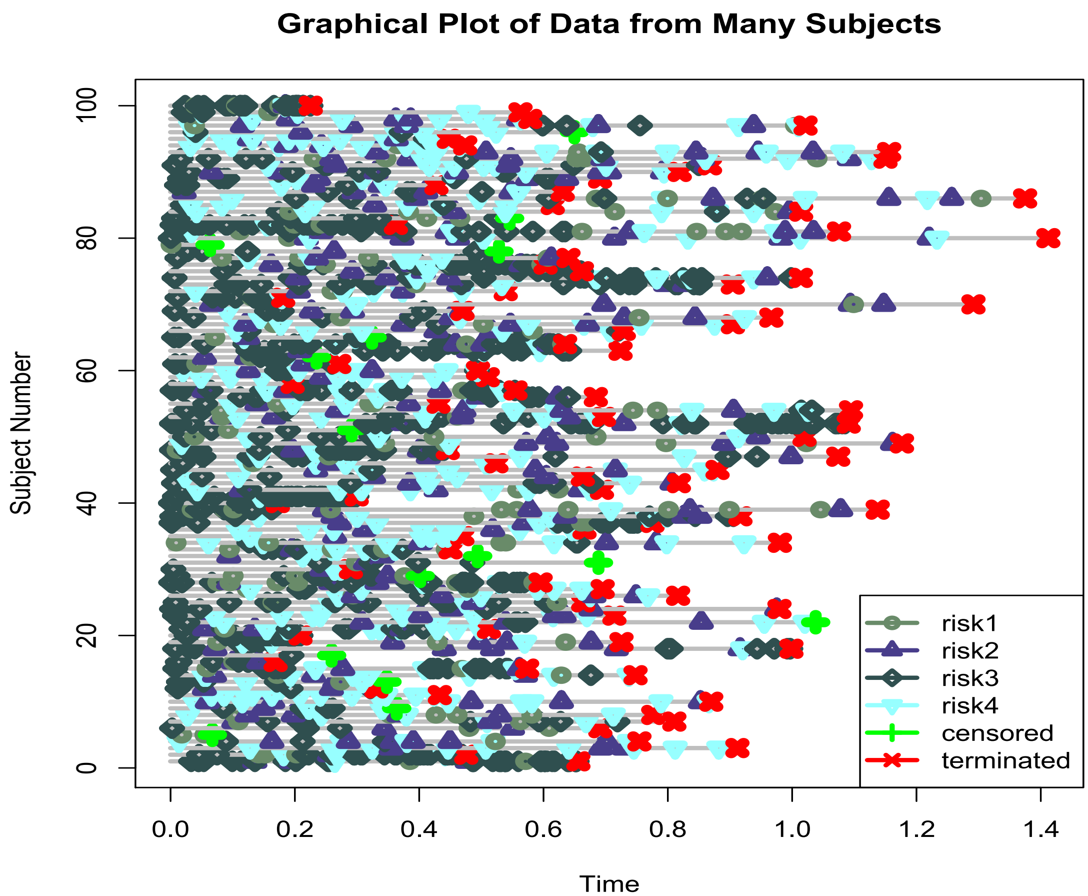

Figure 2.4: A Simulated Data Set For Estimation (No-Frailty): n = 100

2.5 Parameter Estimates on a Simulated Dataset

We implement the parameter estimation procedure on a simulated dataset shown in Figure 2.4. Table 2.1 displays true parameters used to generate the dataset. Results of the proposed estimation procedure are shown in Table 2.2. Survival functions of the baseline hazards are also plotted in Figure 2.5. The red curves are the true survival functions while the blue colored wiggly curves are estimates.

In this demonstration, we used a more simplified model for the ρ functions in equations 2.2, wherein $\alpha_q = (\alpha_q I\{q = j\}, j = 1, 2, \cdots, 4)$, for $q = 1, 2, \cdots, 4$, that is, all elements are zeros except for the qth position. As such the parameter was simply identified as $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$.

Table 2.1: Parameters of Simulated Dataset For Parameter Estimation

Risk(q)	α_q	β_q	κ_q^s	λ_q
TE (0)	(0.1, 0.1, 0.05, 0.5)	(0.3, -0.4, 0.5)	2	0.4
1	0.25	(-0.2, 0.1, 0.3)	1	2
2	0.2	(0.3, 0.1, 0.05)	2	3
3	0.1	(0.3, -0.1, 0.4)	0.5	5
4	0.05	(0, 1, -0.5)	1.5	4

Table 2.2: Finite-Dimensional Paramester Estimates (with No-Frailty)

Risk(q)	$\hat{\alpha}_q$	$\hat{\beta}_q$
TE (0)	(0.09, 0.37, 0.04, 0.41)	(0.20, -0.14, 0.57)
1	0.18	(0.03, 0.19, 0.21)
2	0.09	(0.24, 0.02, 0.00)
3	0.09	(0.19, -0.07, 0.38)
4	0.06	(-0.15, 1.08, -0.55)

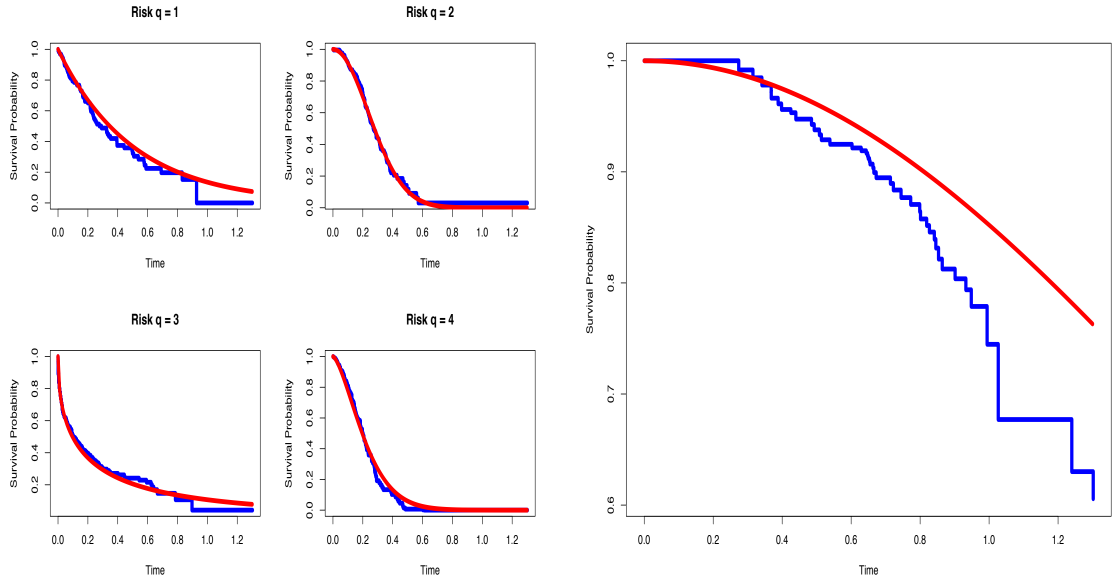

Figure 2.5: PLE Survival Functions of Baseline Hazards (No-frailty)

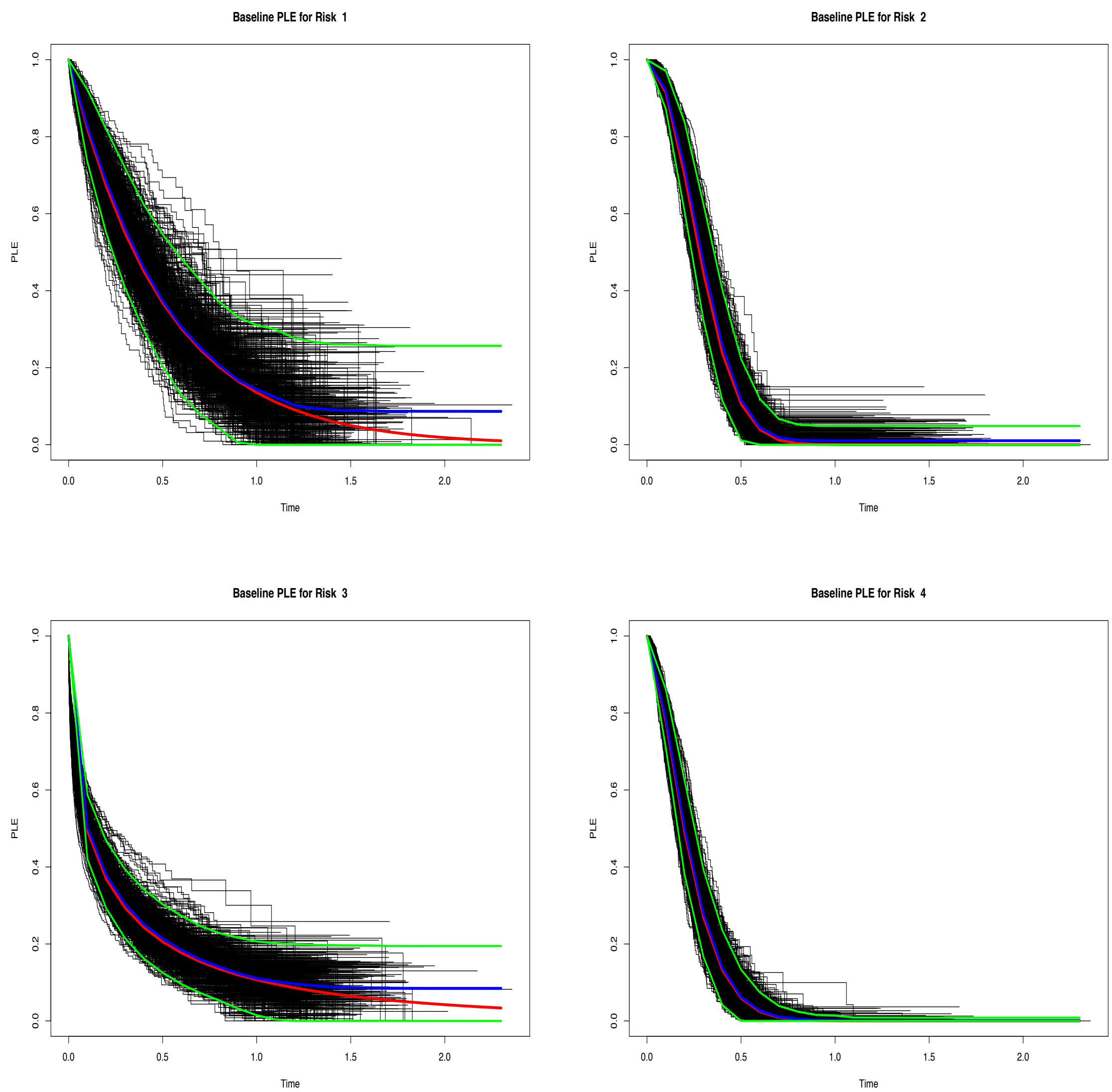

Figure 2.6: Overlaid plots of simulated PL estimates of the baseline survivor functions for the recurrent competing risks under the no-frailty model.

2.6 A Small Simulation Study

To understand the large/finite-sample properties of the estimators, we perform a simulation study from $m = 500$ repeated data samples. In the simulation study the model parameters were Q = 4 recurrent competing risks, p = 3 covariates with $X_1 \sim Ber(.5)$ and $(X_2, X_3) \sim BVN(0, 0, 1, 1, \rho = .5)$, and with partial type effective ages. To generate the simulation data, the baseline hazard functions for the recurrent competing

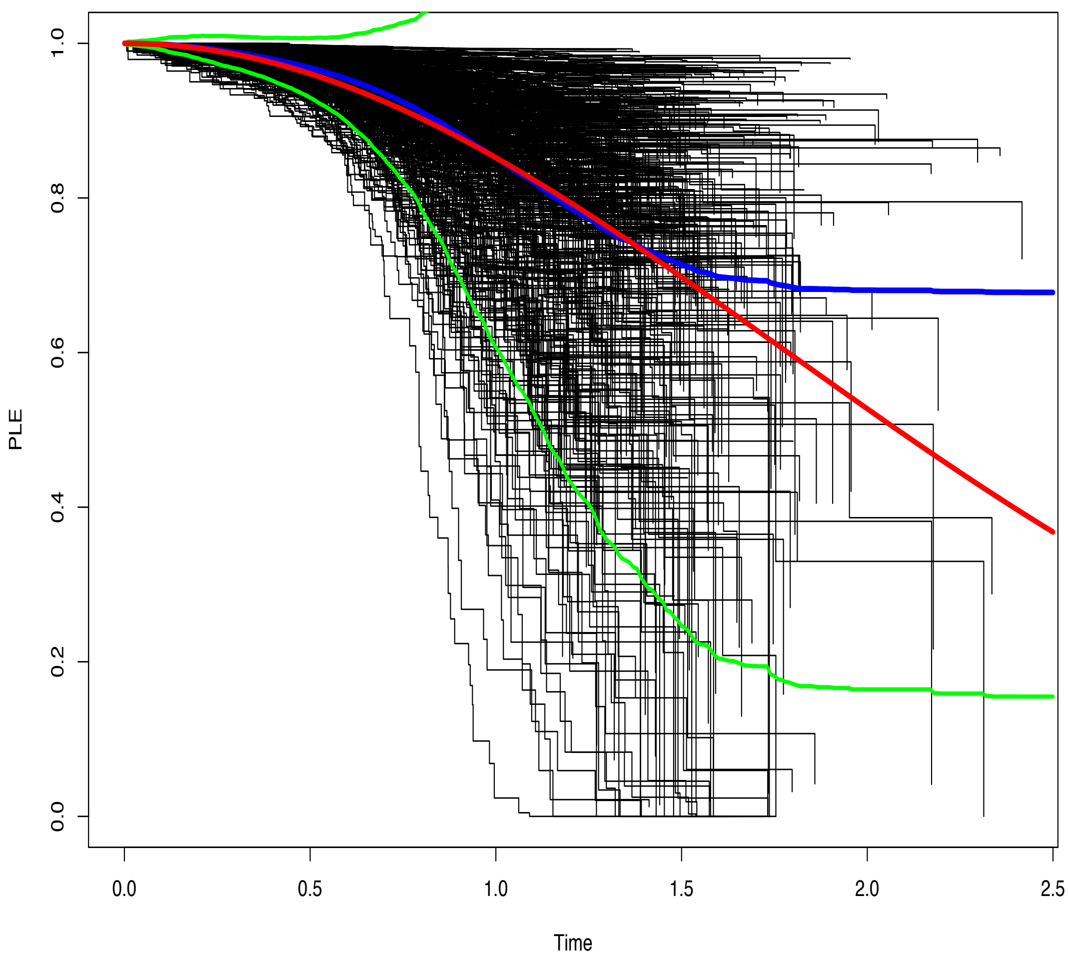

Figure 2.7: Overlaid plots of simulated PL estimates of the baseline survivor function for the terminal event portion under the no-frailty model (PL Estimates with True SF and Mean Curve of Estimates)

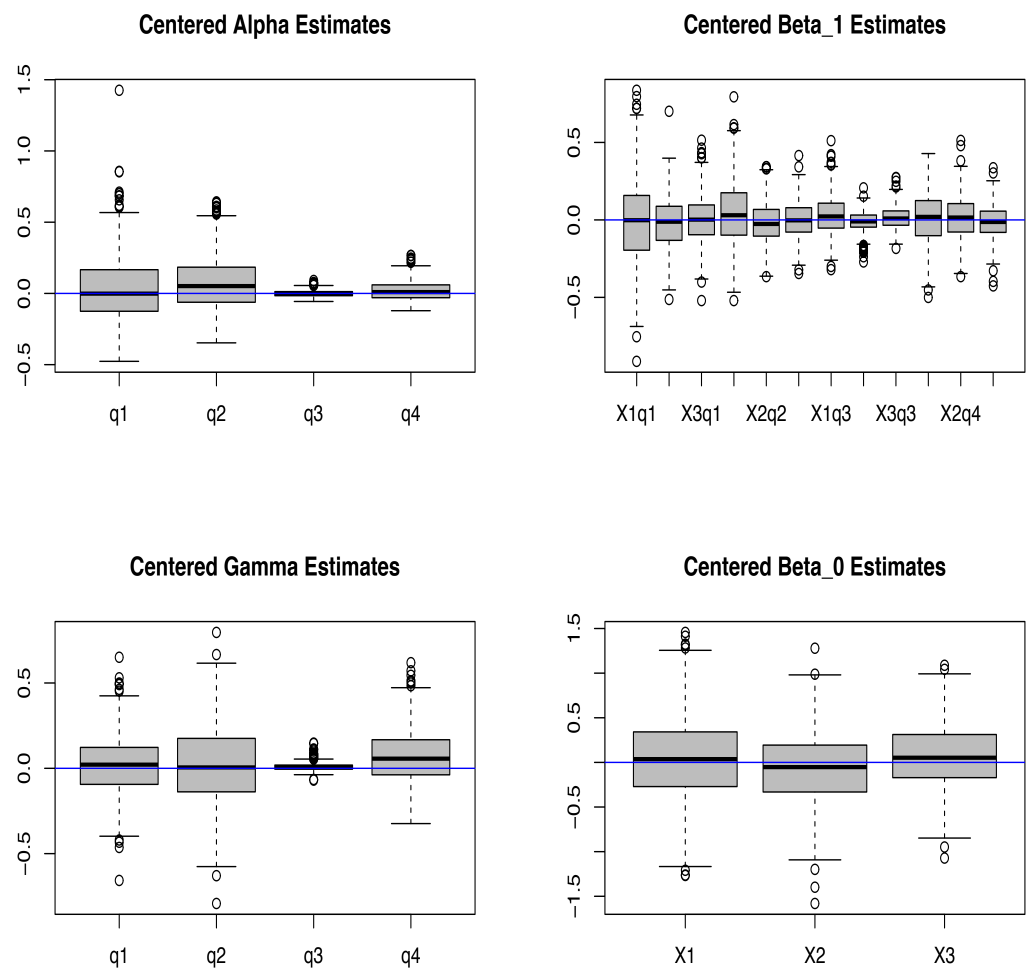

Figure 2.8: Boxplot of Finite Dimensional Parameter Estimates from 500 Data Samples

risks and terminal event are Weibull hazard functions with shape parameters and scale parameters shown in Table 2.1. The Weibull distribution of the terminal event baseline hazard function has shape parameter $\kappa_0 = 2$ and scale $\lambda_0 = 0.4$. However, The proposed joint dynamic model is semiparametric in nature. The baseline hazards of RCR and TE are left unspecified in the model, and the cumulative baseline hazards are estimated nonparametrically. Therefore, the shape and scale parameters associated with the Weibull distributions are not an interest of the parameter estimation, and the Weibull distribution are used for the purpose of data generation.

For the PLE of baseline hazard, $q = 1, 2, \cdots, Q$, Figure 2.6 depicts overlaid plots of the PL estimates of the baseline survivor functions for each of the four recurrent competing risks. Recall that the true baseline survivor functions are of the Weibull type. These are indicated by the red-colored curves, while the blue-colored curves represent the mean of the 500 estimates, respectively. The green curves represent the two standard deviation curves from the mean curve with the standard deviation computed from the 500 replicates. Note the close agreement between the true curve and the mean curve. Figure 2.7 presents the overlaid plot of PL estimates for the baseline survivor function for the terminal event portion. Again, the true curve is of a Weibull-type and note the close agreement between the true curve (red) and the mean curve (blue). Figure 2.8 is the boxplot of finite-dimensional parameter estimates of the 500 samples from the population. Observe that the mean of the estimates are close to their true values. We also observe that the standard errors of the estimates of parameters for the third recurrent competing risk are smaller than those for the other risks. This is because there were more event occurrences for this risk.

We also plot histograms of the 500 estimates for each estimator. Figure 2.9 and 2.10 display the histograms. The green colored bars indicate the average of the 500 replicates. Theoretical investigation of large sample properties of the estimators will be pursued in the future.

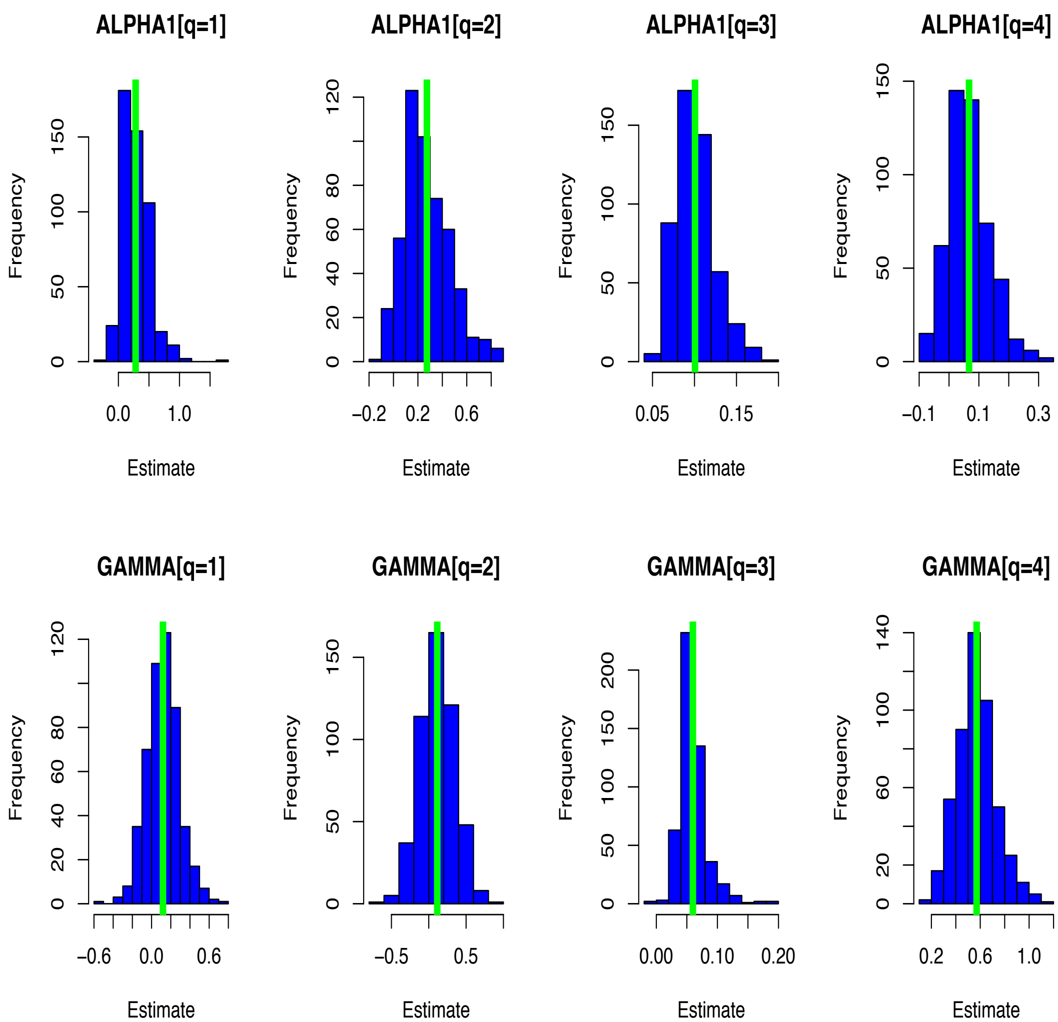

Figure 2.9: Histogram of Parameter Estimates From 500 Data Samples (Estimates of Unknown Parameters In the ρ Functions)

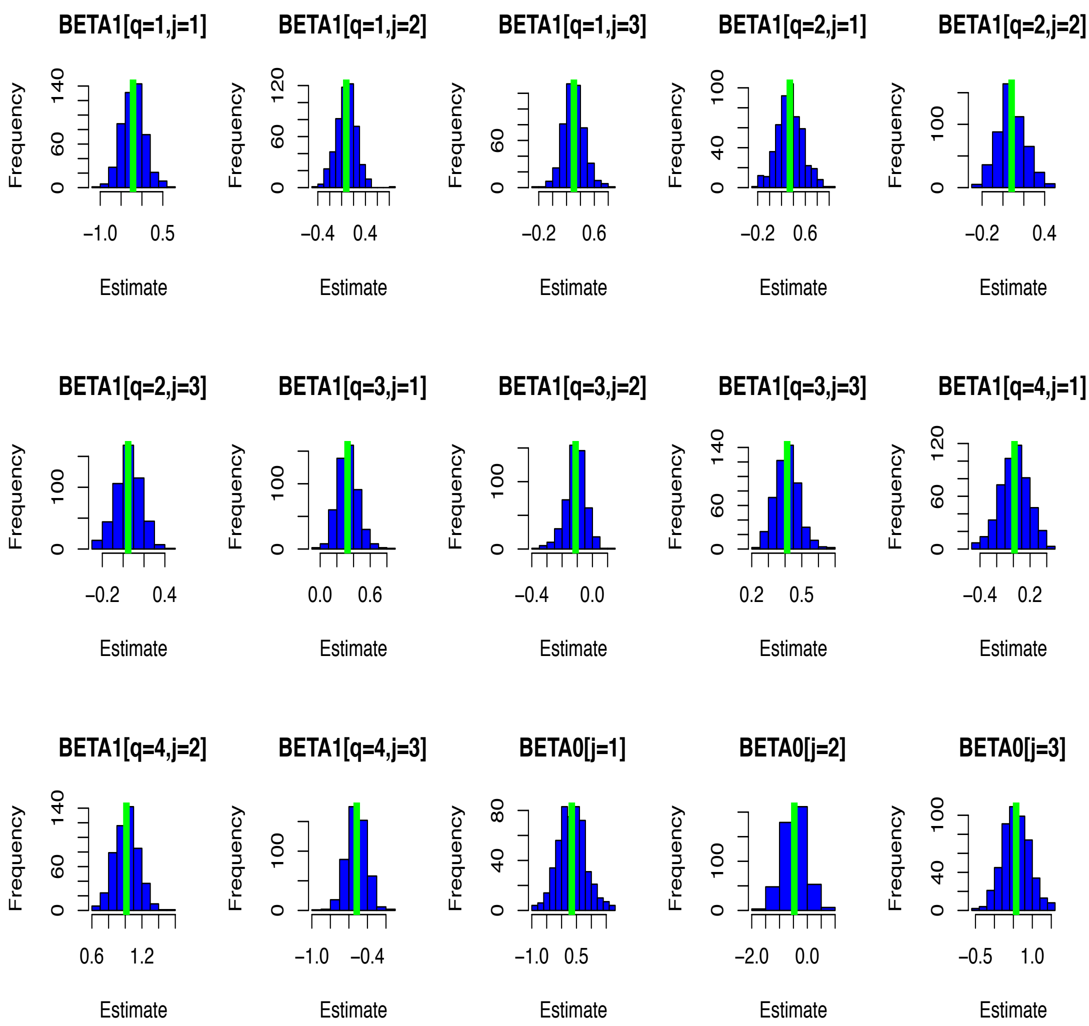

Figure 2.10: Histogram of Parameter Estimates From 500 Data Samples (Estimates of Unknown Parameters Associated With the Covariates)

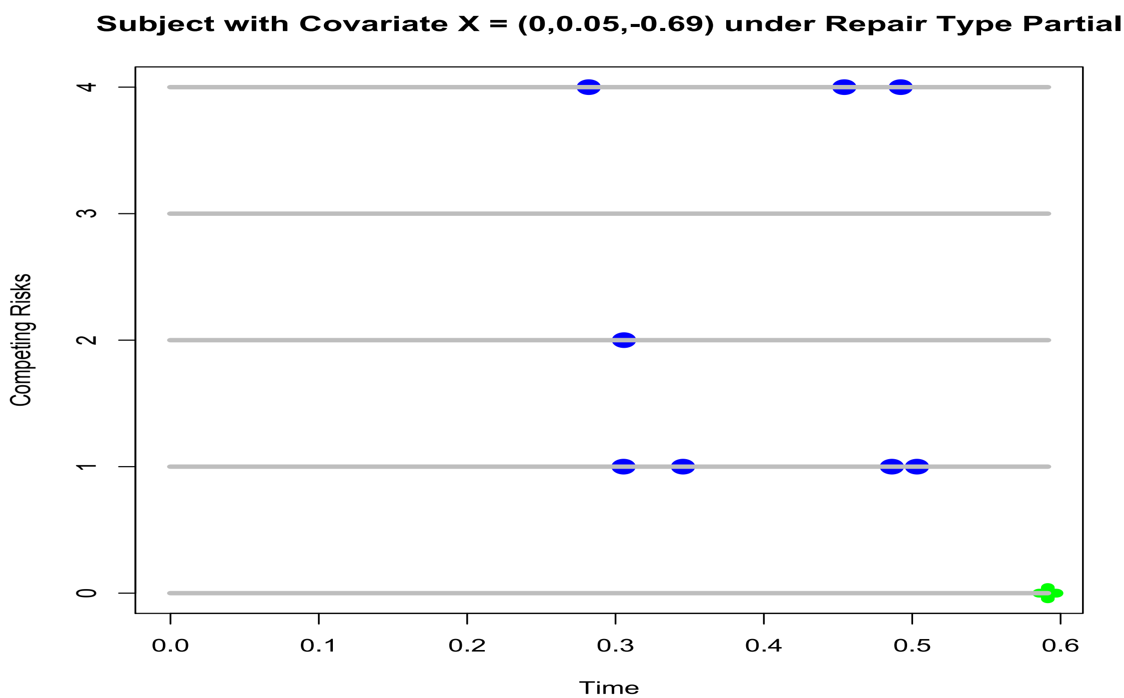

Figure 2.11: Past Event Occurrences of New Unit 0 Under Partial Repair

2.7 Predictions of Terminal Event Survival Probabilities

A major goal of joint dynamic modeling is to predict terminal event survival probabilities for a new unit 0 incorporating information from past recurrent competing risk occurrences. When the statistical inference procedure of the proposed joint dynamic modeling approach is implemented, parameter estimates are obtained. For a new unit 0 that has not experienced the terminal event, shown in Figure 2.11, we propose a simulation approach to predict time-to-terminal event survival probabilities. For illustration purposes, we use the no-frailty case under partial repair as an example.

A single new unit 0 that has survived the monitoring period is considered for the prediction problem. In Figure 2.12, 500 paths of terminal event times, which

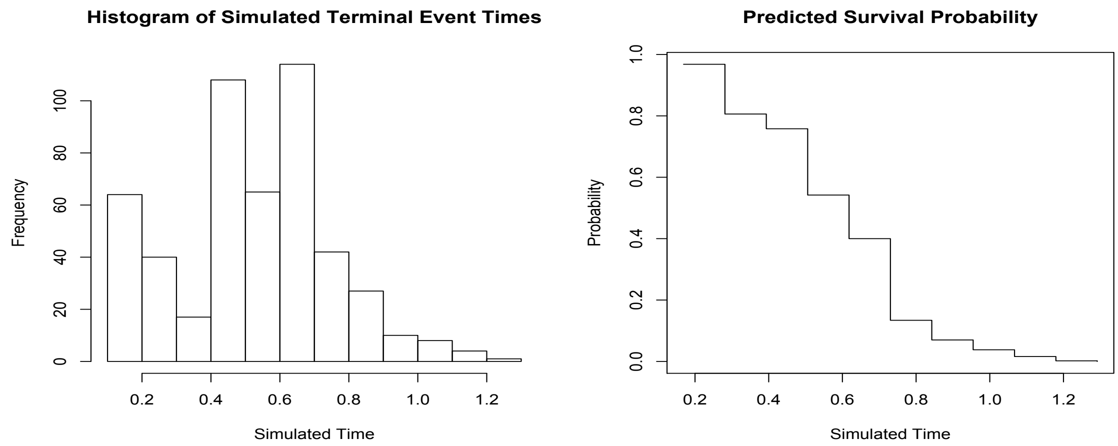

Figure 2.12: Predicted Terminal Event Times (No-frailty, Partial Repair, 500 Paths)

Table 2.3: Mean, Median and SD of Predicted Residual TE Times Under Partial Repair

Mean	Median	S.E
0.531	0.524	0.224

Table 2.4: Simulated TE Survival Probabilities

Time	0.169	0.393	0.618	0.843	1.067
Probabilities	0.968	0.758	0.400	0.070	0.016

starts from the end of the monitoring time τ, are simulated. Past event occurrences, and parameter estimates are considered when generating the paths. Distribution of such simulated survival times is plotted in the left panel of the plot while survival probabilities are plotted in the right panel in Figure 2.12.

We compute terminal event survival probabilities using the 500 simulated paths. In medical research, or cancer prognostics, these terminal event probabilities will provide predictions taking into account an individual patient's past event history. Table 2.3 summarizes the simulated residual lifetimes, and Table 2.4 provides predicted survival probabilities for the new unit 0 after the end of monitoring period τ.

Chapter 3

Joint Dynamic Modeling of Recurrent Competing Risks and A Terminal Event With Frailty

3.1 Introduction

In real-life applications, recurrent competing risks observed from the same unit are often correlated. However, factors contributing to such associations are usually unobserved. To account for these unobserved factors, a latent variable Z is introduced to induce associations between different recurrent competing risks described in section 2.1. In reliability and survival analysis literature, this unobserved Z variable is called a frailty. Many previous models have assumed Z to follow a gamma distribution. We adopt such an assumption in our models and assume $Z \sim Ga(\xi, \xi), \xi > 0$, avoiding identifiability issues. Note that when $\xi \to \infty$, the model is identical to the model in Chapter 2 .

3.2 Data

Frailty variable Z is unobserved. The observables of the frailty model are the same as those in Chapter 2. $\mathbf{D}_i = \{(N_{qi}^{\dagger}(s), N_{0i}^{\dagger}(s), Y_i^{\dagger}(s), \mathcal{E}_{qi}(s)) : s \geq 0, X_i, \tau_i\}, i = 1, 2, \cdots, n$. τ_i is the subject specific random censoring time. We assume that τ_i is independent of $N_{qi}^{\dagger}(s)$ and $N_{0i}^{\dagger}(s)$.

3.3 Model Description

To include the frailty variable in the joint dynamic models developed in Chapter 2, we specify the intensity processes as follow: For the ith observational unit, and $q = 1, 2, \cdots, Q$, the intensity processes are

$$\begin{aligned} A^{\dagger}_{qi}(s|Z_i) &= Z_i \int_0^t Y^{\dagger}_i(v)\rho_q(\mathbf{N}^{\dagger}_i(v-);\alpha_q)\exp(X_i\beta_q)\lambda_{q0}(\mathcal{E}_{qi}(v))dv \\ A^{\dagger}_{0i}(s|Z_i) &= Z_i \int_0^s Y^{\dagger}_i(v)\rho_0(\mathbf{N}^{\dagger}_i(v-);\gamma)\exp(X_i\beta_0)\lambda_0(v)dv \end{aligned}$$

For ease of notation, we also define the following

$$\begin{aligned} a^{\dagger}_{qi}(v|Z_i)dv =& Z_iY^{\dagger}_i(v)\rho_q(\mathbf{N}^{\dagger}_i(v-);\alpha_q)\exp(X_i\beta_q)\lambda_{q0}(\mathcal{E}_{qi}(v))dv \\ a^{\dagger}_{0i}(v|Z_i)dv =& Z_iY^{\dagger}_i(v)\rho_0(\mathbf{N}^{\dagger}_i(v-);\gamma)\exp(X_i\beta_0)\lambda_0(v)dv \end{aligned} \tag{3.1}$$

Incorporating the frailty variable Z into the likelihood from Section 2.2, and conditional on Z_i, the likelihood of the frailty case for the ith unit is

$$\begin{aligned} \mathcal{L}_c(v,\Theta|Z_i) &\equiv \prod_{v=0}^{s} \mathbf{P}\{\bigcap_{q=1}^{Q}[dN^{\dagger}_{qi}(v) = dn_{qi}(v)]; [dN^{\dagger}_{0i}(v) = dn_{0i}(v)]|\mathcal{F}_{v-}, Z_i\} \\ &= \{\prod_{q=1}^{Q}[dA_{qi}(v|Z_i)]^{dn_{qi}(v)}[1 - dA_{qi}(v|Z_i)]^{1-dn_{qi}(v)}\} \\ &\times \{[dA_{0i}(v|Z_i)]^{dn_{0i}(v)}[1 - dA_{0i}(v|Z_i)]^{1-dn_{0i}(v)}\} \end{aligned}$$

where $dn_{qi}(v), dn_{0i}(v) \in \{0, 1\}$ and $\sum_{q=1}^{Q} dn_{qi}(v) + dn_{0i}(v) \leq 1$.

3.4 Parameter Estimation

Impact of Frailty Z

Similar to the no-frailty case, parameters are estimated semi-parametrically. The impact of frailty Z with regards to estimation is seen in the aggregated generalized at-risk processes. For the recurrent competing risks processes, we define some new aggregated generalized at-risk processes. For $q = 1, 2, \cdots, Q$,

$$\begin{aligned} S_{q0}(s, w|\alpha_q, \beta_q, \mathbf{z}) &= \sum_{i=1}^{n} Y_{qi}(s, w|\alpha_q, \beta_q, z_i) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{N_{qi}^{\dagger}[(s\wedge\tau_i)-]+1} z_i I[w \in \left(\mathcal{E}_{qi}(s_{ij-1}), \mathcal{E}_{qi}(s_{ij})\right)] \frac{\kappa_{qi}(\mathcal{E}_{qij}^{-1}(w))}{\mathcal{E}_{qi}'(\mathcal{E}_{qij}^{-1}(w))} \end{aligned}$$

where $\kappa_{qi}(\mathcal{E}_{qij}^{-1}(w)) = \rho_q(N_{qi}^{\dagger}(\mathcal{E}_{qij}^{-1}(w)-); \alpha_q) \exp(X_i^T \beta_q)$. For the terminal event process,

$$S_0(v|\gamma, \beta_0, \mathbf{z}) = \sum_{i=1}^{n} Y_i^{\dagger}(v|\gamma, \beta_0, z_i) = \sum_{i=1}^{n} z_i I[(\tau_i \wedge S_i) \geq v] \kappa_{0i}(v)$$

where $\kappa_{0i}(v) = \rho_0(\mathbf{N}_i^{\dagger}(v-); \gamma) \exp(X_i^T \beta_0)$ with S_i being time-to-terminal event. Following the approach in [44], given values of the finite-dimensional parameters and frailty $\mathbf{Z} = \mathbf{z}$, we estimate baseline hazards of the recurrent competing risks and the terminal event with frailty using the expressions below

$$\hat{\Lambda}_{q0}(s, t|\mathbf{z}, \alpha_q, \beta_q) = \int_0^t \frac{\sum_{i=1}^n N_{qi}(s, dw)}{S_{q0}(s, w|\mathbf{z}, \alpha_q, \beta_q)}; \qquad \hat{\Lambda}_0(t|\mathbf{z}, \gamma, \beta_0) = \int_0^t \frac{\sum_{i=1}^n N_{0i}^{\dagger}(dv)}{S_0(v|\mathbf{z}, \gamma, \beta_0)}. \tag{3.2}$$

Plugging in the estimates of the finite-dimensional parameters, the PLEs of the baseline survival functions of the recurrent competing risks ($q = 1, 2, \cdots, Q$) and the terminal event processes, conditional on $\mathbf{Z} = \mathbf{z}$, are

$$\hat{\bar{F}}_{q0}(s, t|\mathbf{z}) = \prod_{w=0}^{t} [1 - \hat{\Lambda}_{q0}(s, dw|\mathbf{z})]; \qquad \hat{\bar{F}}_0(t|\mathbf{z}) = \prod_{w=0}^{t} [1 - \hat{\Lambda}_0(dw|\mathbf{z})].$$

An EM Algorithm

We develop an EM algorithm (cf. [12]) to estimate the finite-dimensional parameters. With the introduction of frailty Z, the vector of unknown parameters becomes $\Theta = \Theta^{(nofrail)} \cup \{\xi\}$ in the frailty case. Assuming $\mathbf{Z} = \mathbf{z}$ is known, we obtain the full likelihood process as below

$$\mathcal{L}^{\dagger}[s^*|\Theta, \mathbf{Z}=\mathbf{z}, \mathbf{D}(s^*)] = \prod_{i=1}^{n} \Big\{ \frac{\xi^{\xi}}{\Gamma(\xi)} z_i^{\xi-1} \exp(-\xi z_i)$$

$$\times \prod_{v=0}^{s^*} \prod_{q=1}^{Q} \Big(z_i Y_i^{\dagger}(v) \lambda_{q0}(\mathcal{E}_{qi}(v)) \rho_q \Big[N_{qi}^{\dagger}(v-); \alpha_q \Big] \exp(X_i^T \beta_q) \Big)^{N_{qi}^{\dagger}(dv)}$$

$$\times \exp\Big(-\int_0^{s^*} z_i Y_i^{\dagger}(v) \lambda_{q0}(\mathcal{E}_{qi}(v)) \rho_q \Big[N_{qi}^{\dagger}(v-); \alpha_q \Big] \exp(X_i^T \beta_q) dv \Big)$$

$$\times \Big(z_i Y_i^{\dagger}(v) \lambda_0(v) \rho_0 [\mathbf{N}_i^{\dagger}(v-); \gamma] \exp(X_i^T \beta_0) \Big)^{N_i^{\dagger}(dv)}$$

$$\times \exp\Big(-\int_0^{s^*} z_i Y_i^{\dagger}(v) \lambda_0(v) \rho_0 \Big[\mathbf{N}_i^{\dagger}(v-); \gamma \Big] \exp(X_i^T \beta_0) dv \Big) \Big\}$$

To compute conditional distribution of $Z_i, i = 1, 2, \cdots, n$, we use the fact that

$$Z|\Theta, \mathbf{D}(s^*) \propto \mathcal{L}^{\dagger}[s^*|\Theta, \mathbf{Z}=\mathbf{z}, \mathbf{D}(s^*)] \prod_{i=1}^{n} f(Z_i|\xi).$$

We then obtain $Z_i|\mathbf{D}(s^*), \Theta \overset{iid}{\sim} Ga(\alpha, \beta)$, with

$$\begin{aligned} \alpha(s^*) &= \xi + \sum_{q=1}^{Q} N_{qi}^{\dagger}(s^*-) + N_{0i}^{\dagger}(s^*-) \\ \beta(s^*) &= \xi + \sum_{q=1}^{Q} \int_0^{s^*} A_{qi}^{\dagger}(dv) + \int_0^{s^*} A_{0i}^{\dagger}(dv) \end{aligned}$$

For $i = 1, 2, \cdots, n$, the conditional expectation of $Z_i|\mathbf{D}(t), \Theta$ is

$$\begin{aligned} \mathbf{E}[Z_i|\mathcal{F}_{t-}, X_i, \Theta] &= \frac{\alpha(t)}{\beta(t)} = \frac{\xi + \sum_{q=1}^{Q} N_{qi}^{\dagger}(t-) + N_{0i}^{\dagger}(t-)}{\xi + \sum_{q=1}^{Q} \int_0^t A_{qi}^{\dagger}(dv) + \int_0^t A_{0i}^{\dagger}(dv)} \\ \mathbf{E}[\log(Z_i)|\mathcal{F}_{t-}, X_i, \Theta] &= \mathbf{DG}(\xi + \sum_{q=1}^{Q} N_{qi}^{\dagger}(t-) + N_{0i}^{\dagger}(t-)) \\ &+ \log[\mathbf{E}\{Z_i|\mathcal{F}_{t-}, X_i, \Theta\}] - \log(\xi + \sum_{q=1}^{Q} \int_0^t A_{qi}^{\dagger}(dv) + \int_0^t A_{0i}^{\dagger}(dv)) \end{aligned}$$

where $\mathbf{DG}(\alpha) = \frac{d}{d\alpha} \log \Gamma(\alpha)$.

The EM algorithm is described as follows:

E-step: Obtain conditional expectation of the full log-likelihood with respect to $\mathbf{Z}|(\mathbf{D}(t-), \Theta)$. Let

$$\widehat{Z_i} = \mathbf{E}[Z_i|\mathcal{F}_{t-}, X_i, \Theta]; \qquad \widehat{logZ_i} = \log(\mathbf{E}[Z_i|\mathcal{F}_{t-}, X_i, \Theta])$$

$$\begin{aligned}
\mathbf{E}(\log \mathcal{L}_c^{\dagger}[s^*|\Theta, \mathbf{Z}, \mathbf{D}(s^*)) &= n\xi \log \xi - n \log \Gamma(\xi) \\
&+ \sum_{i=1}^{n} \widehat{logZ_i}(\sum_{q=1}^{Q} N_{qi}^{\dagger}(s^*) + N_{0i}^{\dagger}(s^*) + \xi - 1) \\
&- \sum_{i=1}^{n} \widehat{Z_i}\Big(\xi + \int_0^{s^*} [\sum_{q=1}^{Q} A_{qi}^{\dagger}(dv) + A_{0i}^{\dagger}(dv)]\Big) \\
&+ \sum_{i=1}^{n} \Big(\sum_{q=1}^{Q} \int_0^{s^*} \log a_{qi}^{\dagger}(v) N_{qi}^{\dagger}(dv) + \int_0^{s^*} \log a^{\dagger}(v) N_{0i}^{\dagger}(dv)\Big)
\end{aligned}$$

M-step: When values of the finite-dimensional parameters in Θ and the frailty variables $\mathbf{Z}$ are given, baseline hazards of the recurrent competing risks and the terminal event can be estimated non-parametrically as in 3.4. Small changes to the estimating equations are needed to incorporate the frailty variable. For α_q and $\beta_q, q = 1, 2, \cdots, Q$:

$$\sum_{i=1}^{n} \int_0^{\tau_i} \left[\frac{\frac{\partial}{\partial \alpha_q} \rho_q(N_{qi}^{\dagger}(v-); \alpha_q))}{\rho_q(N_{qi}^{\dagger}(v-); \alpha_q))} - \frac{\frac{\partial}{\partial \alpha_q} S_{q0}(s, \mathcal{E}_{qi}(v)|\alpha_q, \beta_q, \mathbf{z})}{S_{q0}(s, \mathcal{E}_{qi}(v)|\alpha_q, \beta_q, \mathbf{z})} \right] N_{qi}^{\dagger}(dv) = 0$$

$$\sum_{i=1}^{n} \int_0^{\tau_i} \left[X_i - \frac{\frac{\partial}{\partial \beta_q} S_{q0}(s, \mathcal{E}_{qi}(v)|\alpha_q, \beta_q, \mathbf{z})}{S_{q0}(s, \mathcal{E}_{qi}(v)|\alpha_q, \beta_q, \mathbf{z})} \right] N_{qi}^{\dagger}(dv) = 0$$

For γ and β_0:

$$\sum_{i=1}^{n} \int_0^{\tau_i} \left[\frac{\frac{\partial}{\partial \gamma} \rho_0(N_i^{\dagger}(v-); \gamma)}{\rho_0(N_i^{\dagger}(v-); \gamma)} - \frac{\frac{\partial}{\partial \gamma} S_0(v|\gamma, \beta_0, \mathbf{z})}{S_0(v|\gamma, \beta_0, \mathbf{z})} \right] N_{0i}^{\dagger}(dv) = 0$$

$$\sum_{i=1}^{n} \int_0^{\tau_i} \left[X_i - \frac{\frac{\partial}{\partial \beta_0} S_0(v|\gamma, \beta_0, \mathbf{z})}{S_0(v|\gamma, \beta_0, \mathbf{z})} \right] N_{0i}^{\dagger}(dv) = 0$$

We follow the algorithm described below to estimate all parameters:

1. Initialize $\hat{\mathbf{Z}}^{(0)} = \mathbf{1}_{n\times 1}, \hat{\alpha}_q^{(0)}, \hat{\beta}_q^{(0)}, \hat{\gamma}^{(0)}, \hat{\beta}_0^{(0)}$.
2. Obtain $\hat{\Lambda}_{q0}^{(0)}(.), q = 1, 2 \cdots, Q$ and $\hat{\Lambda}_0^{(0)}(.)$.
3. Update to $\hat{\mathbf{Z}}^{(1)}$ using $\{\hat{\Lambda}_{q0}(.), \hat{\alpha}_q^{(0)}, \hat{\beta}_q^{(0)}, q = 1, 2, \cdots, Q; \hat{\gamma}^{(0)}, \hat{\beta}_0^{(0)}, \hat{\Lambda}_0(.)\}$.

Table 3.1: Finite-Dimensional Parameter Estimates (with Frailty)

Risk(q)	$\hat{\alpha}_q$	$\hat{\beta}_q$
TE (0)	(0.29, 0.27, 0.04, 0.51)	(-0.01, -0.41, 0.75)
1	-0.02	(0.28, -0.09, 0.44)
2	0.24	(0.40, -0.01, 0.19)
3	0.10	(0.42, -0.21, 0.59)
4	0.07	(0.18, 0.77, -0.41)

4. Obtain $\hat{\xi}^{(0)}$. Define $\mathcal{L}_\xi[s^*|\Theta, \mathbf{D}(s^*)] = \mathbf{E}(\log \mathcal{L}_c^\dagger[s^*|\Theta, \mathbf{Z}, \mathbf{D}(s^*))$ as in the **E** step. $\hat{\xi} = \arg\max_{(\xi)} \mathcal{L}_\xi[s^*|\Theta^{(0)}, \mathbf{D}(s^*)]$.
5. With $\hat{\mathbf{Z}}^{(1)}, \hat{\Lambda}_{q0}^{(0)}(.)$ and $\hat{\Lambda}_0^{(0)}(.)$, we obtain $\hat{\alpha}_q^{(1)}, \hat{\beta}_q^{(1)}, \hat{\gamma}^{(1)}, \hat{\beta}_0^{(1)}$.
6. Reset $\hat{\mathbf{Z}}^{(1)}$ to $\hat{\mathbf{Z}}^{(0)}$, $\hat{\alpha}_q^{(1)}, \hat{\beta}_q^{(1)}, \hat{\gamma}^{(1)}, \hat{\beta}_0^{(1)}$ to $\hat{\alpha}_q^{(0)}, \hat{\beta}_q^{(0)}, \hat{\gamma}^{(0)}, \hat{\beta}_0^{(0)}$.

Repeat steps 2 - 5 until $|(\mathbf{Z}^{(0)}, \Theta^{(0)}) - (\mathbf{Z}^{(1)}, \Theta^{(1)})| <$ tol. For example, tol $= 10^{-7}$. The convergence criterion only applies to the finite-dimensional parameters in Θ.

3.5 Parameter Estimates on a Simulated Dataset

Using the same parameters in Table 2.1, we simulate a second dataset to illustrate the estimation procedure in Section 3.4. ξ, the parameter of frailty variable Z is set to be 2. Hence, frailty variable vector $\mathbf{Z}$ are generated using $Gamma(2, 2)$.

$\hat{\xi}$, the estimate of ξ is 2.767. Estimates of other finite-dimensional parameters are displayed in Table 3.1. Estimates of baselines of the survival functions are plotted in Figure 3.1.

3.6 A Small Simulation Study

In this small simulation study, we simulated $m = 50$ datasets with sample size $n = 30$. The simulation setup is the same as in Section 2.5. Parameters in Table 2.1 are used to generate the data. Assuming a $Ga(\xi, \xi)$ over the frailty variable $Z_i, i = 1, 2, \cdots, 30, \xi = 2$ is used in data generation. Simulation results are presented in this section.

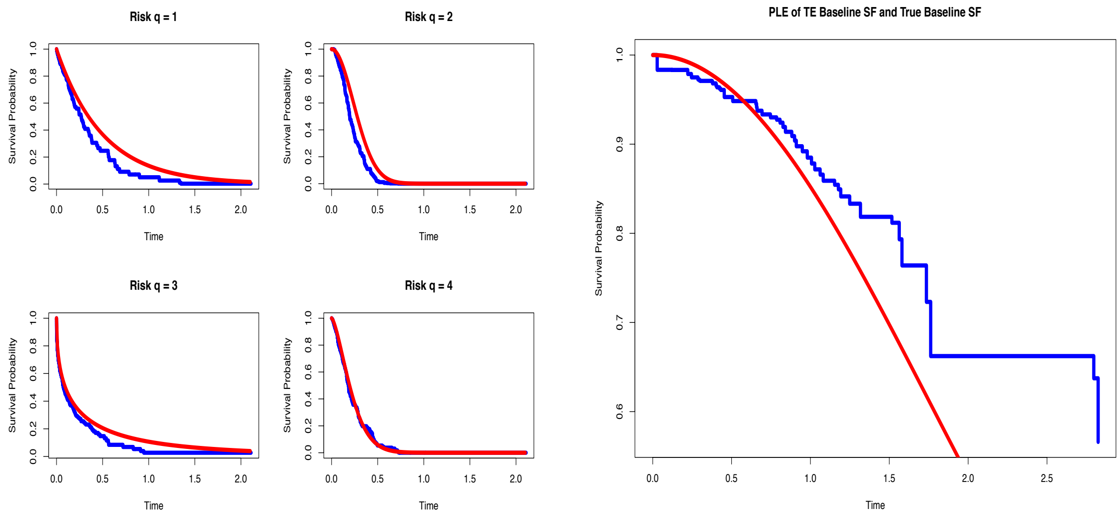

Figure 3.1: *Left*: PLE Survival function of Recurrent Competing Risks; *Right*: PLE Survival function of Terminal Event

In Figure 3.2, for the qth recurrent competing risk, $q = 1, 2, \cdots, 4$, 50 estimates of the baseline survival function are plotted using the simulated datasets. Recall the true survival function of the baseline hazard of the recurrent competing risks is generated using a Weibull distribution. Red curves represent the true survival functions. The mean of the 50 estimates are indicated using the blue curves, and standard deviations of the 50 replicates are also computed. The green curves are two standard deviations of the mean curves. According to the plots, the four blue mean curves are very close to the truth, and most of the estimates are within two standard deviations of the mean curves. In Figure 3.3, 50 estimates of the terminal event baseline hazard are plotted. The red curves represent the true Weibull baseline survival probabilities, and the blue curves are the mean of the 50 replicates. The green curves are two standard deviations of the mean curves, where the standard deviation is calculated using the 50 replicates.

In Figure 3.4, boxplots of the 50 centered estimates of the finite-dimensional parameters are created for the recurrent competing risks. For every estimator, the mean of the 50 replicates is approximately 0; and standard errors vary across different es-

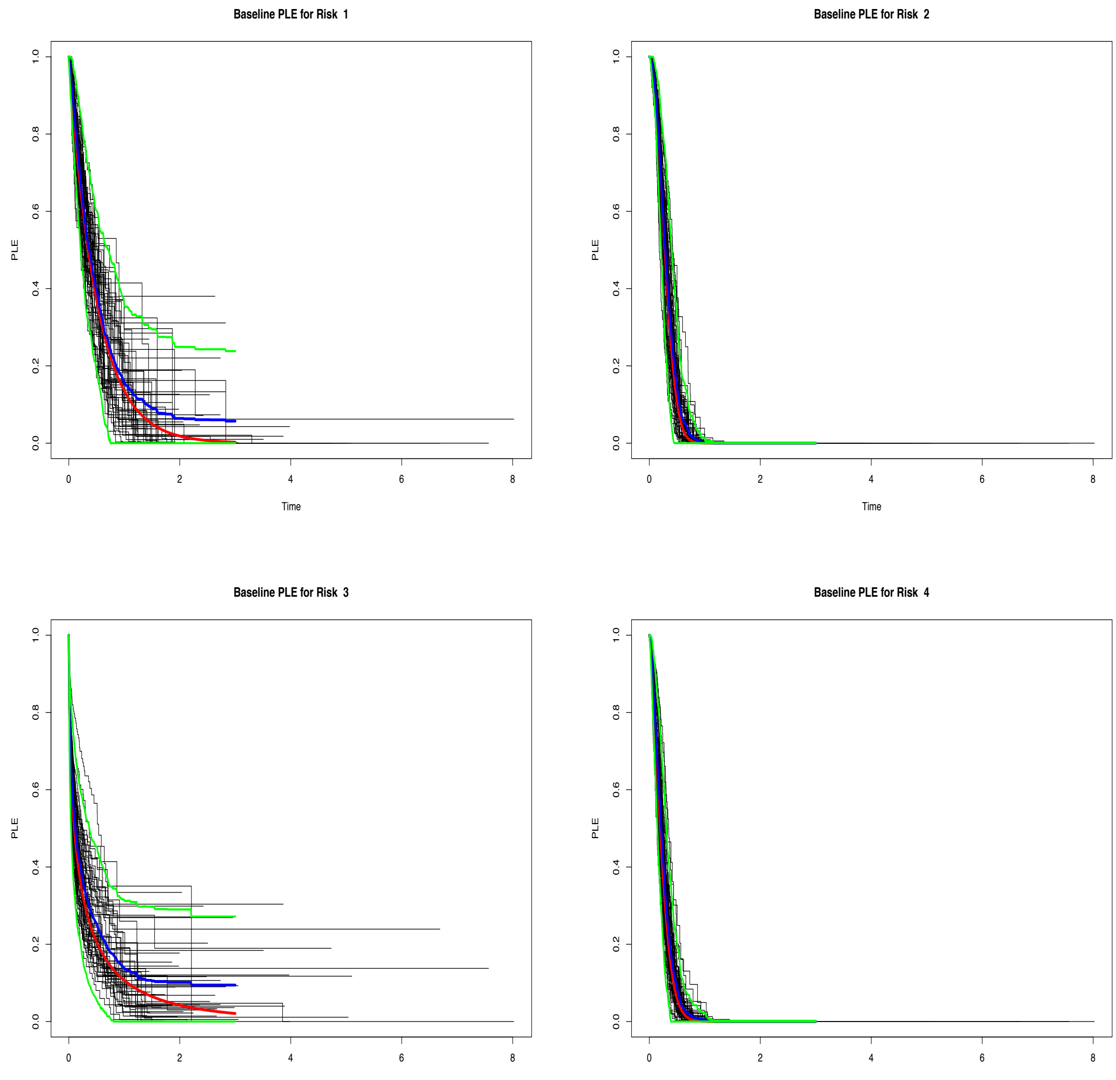

Figure 3.2: Overlaid plots of simulated PL estimates of the baseline survivor functions for the recurrent competing risks under the frailty model.

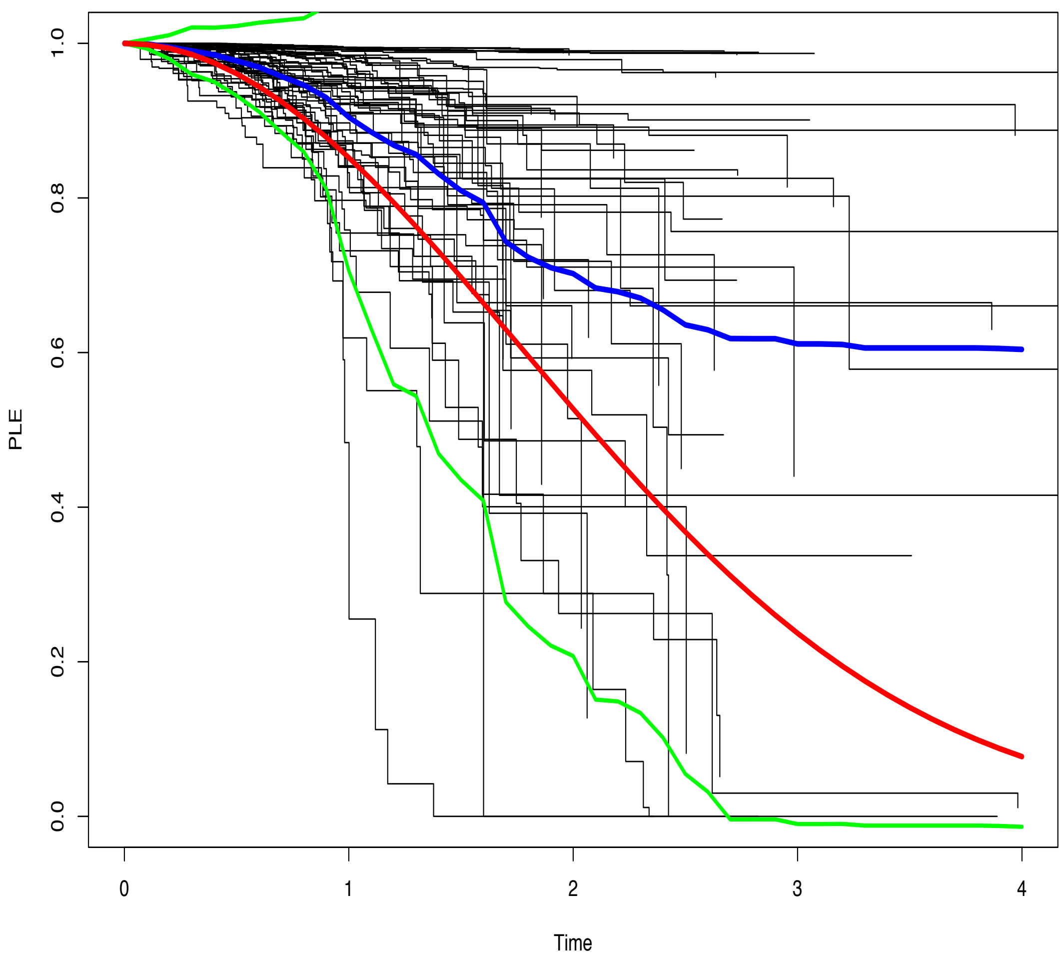

Figure 3.3: Overlaid plots of simulated PL estimates of the baseline survivor function for the terminal event portion under the frailty model (PL Estimates with True SF and Mean Curve of Estimates)

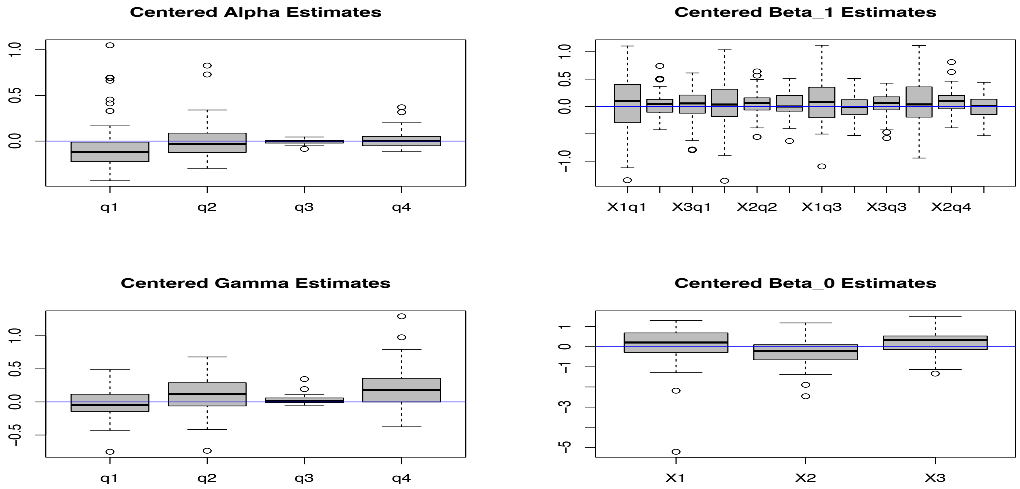

Figure 3.4: Boxplots of centered simulated estimates under the frailty model.

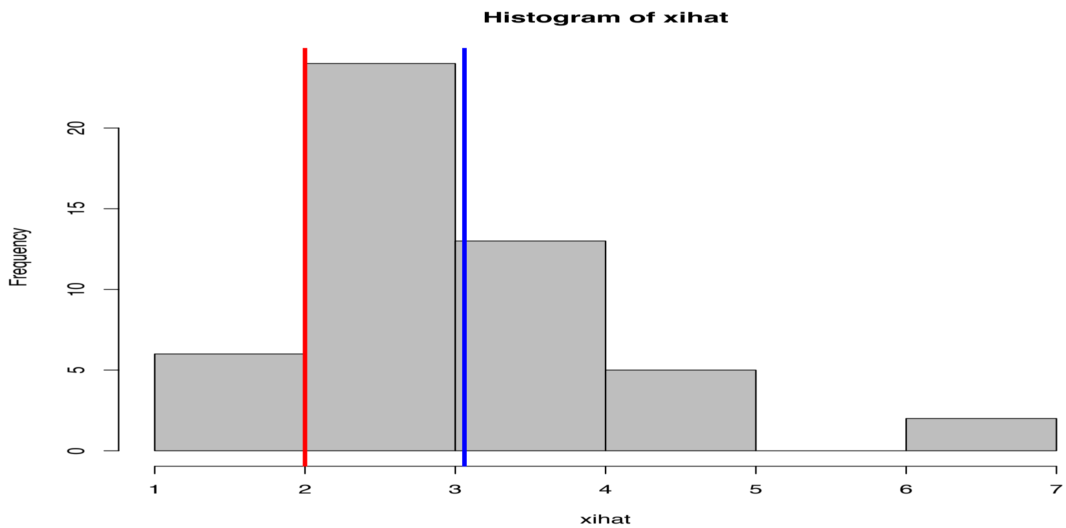

Figure 3.5: Histogram of simulated ξ estimates under the frailty model.

timators. For example, standard errors of the estimators associated with $q = 3$ RCR tend to be smaller compared to other risks. One possible explanation may be that the risk experiences more event occurrences. Therefore, there is more information about the risk. This is similar to what is observed in Section 2.5. In Figure 3.5, histogram of the ξ estimates is plotted. The blue vertical line represents the mean of the 50 ξ estimates, which is slightly larger than the true value of 2. The true parameter value of ξ is represented using a red vertical line.

A larger scale simulation study is necessary and will be performed in the future. However, from these modest results we see that the estimation procedure for the model with frailties appear to be working.

Chapter 4

Semiparametric Joint Dynamic Models for Longitudinal Markers, Correlated Recurrent Competing Risks, and a Terminal Event

4.1 Introduction

Dependence structure of correlated longitudinal marker and time-to-event data has been a key interest of research under the joint modeling framework (cf. [4], [7], [52]). Recent literature provided examples where the profile of the longitudinal marker is modeled, and the rate of time-to-event processes takes into account past history of the longitudinal marker (cf. [27]). Other works often use a shared random effect (cf. [36]) or latent classes (cf. [22]) when modeling possible association between the marker process and time-to-event data. In this Chapter, we tackle the problem

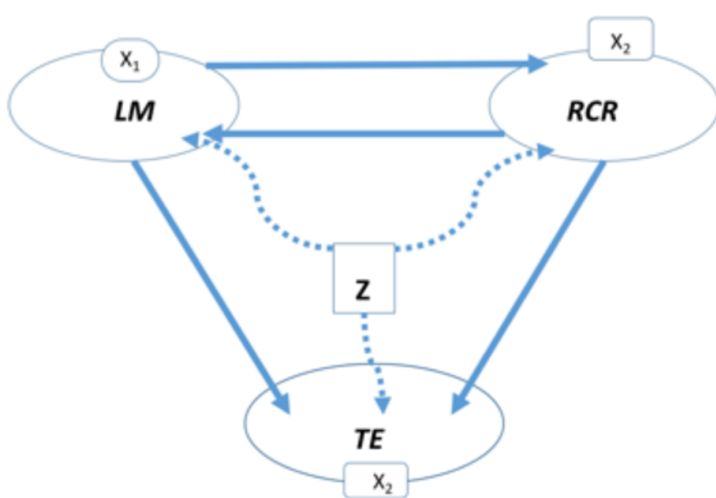

Figure 4.1: Dependence Structure of the LM, RCR and TE

of simultaneously modeling a longitudinal marker (LM), recurrent competing risks (RCR) and a terminal event (TE). The proposed joint dynamic models continue to consider effective age processes and the two types of repair, which are considered in Chapter 2 and Chapter 3, for the RCR and TE submodels. Figure 4.1 is a pictorial representation of the dependence structure of the data. The solid blue arrows indicate the possible dependence between LM and RCR, RCR and TE, or LM and TE; and the dotted lines connecting frailty variable Z to indicate possible associations between the different processes due to the unobserved random variable. Fixed covariate vector X_1 contribute to the mean of the LM, and X_2 is the fixed covariate that contribute to intensity processes of the RCR and TE.

4.2 Data

For the ith unit from the population, $W_i(t_{ij})$ is the value of the continuous longitudinal marker (LM) at time $t_{ij}, i = 1, 2, \cdots, n; j = 1, 2, \cdots, m_i$. The recurrent competing risks (RCR) and terminal event (TE) are observed. X_{i1} is the fixed $p_1 \times 1$ covariate vector contributing to the mean of LM, which may be gender, and demographic characteristics associated with the unit. X_{i2} is the fixed $p_2 \times 1$ covariate vector contributing to the intensity processes of RCR and TE, which may include treatment assignment, and other characteristics associated with the unit. τ_i is the random monitoring time, and assumed to follow an Exponential distribution. The observables for the joint dynamic LM/RCR/TE is

$$\begin{aligned}\mathbf{D}_i =& \{(N^l_{qi}(s), N^l_{0i}(s), Y^l_i(s), \mathcal{E}_{qi}(s), q = 1, 2, \cdots, Q) : s \geq 0;\\ &(W_i(t_{ij}), j = 1, 2, \cdots, m_i); X_{i1}, X_{i2}, \tau_i\}\end{aligned}$$

In Figure 4.2 and Figure 4.3, we show two examples of observational units. Observed LM values, RCR event occurrences and TE event occurrence are plotted. Notice that in Figure 4.3, the unit experiences terminal event before the end of the

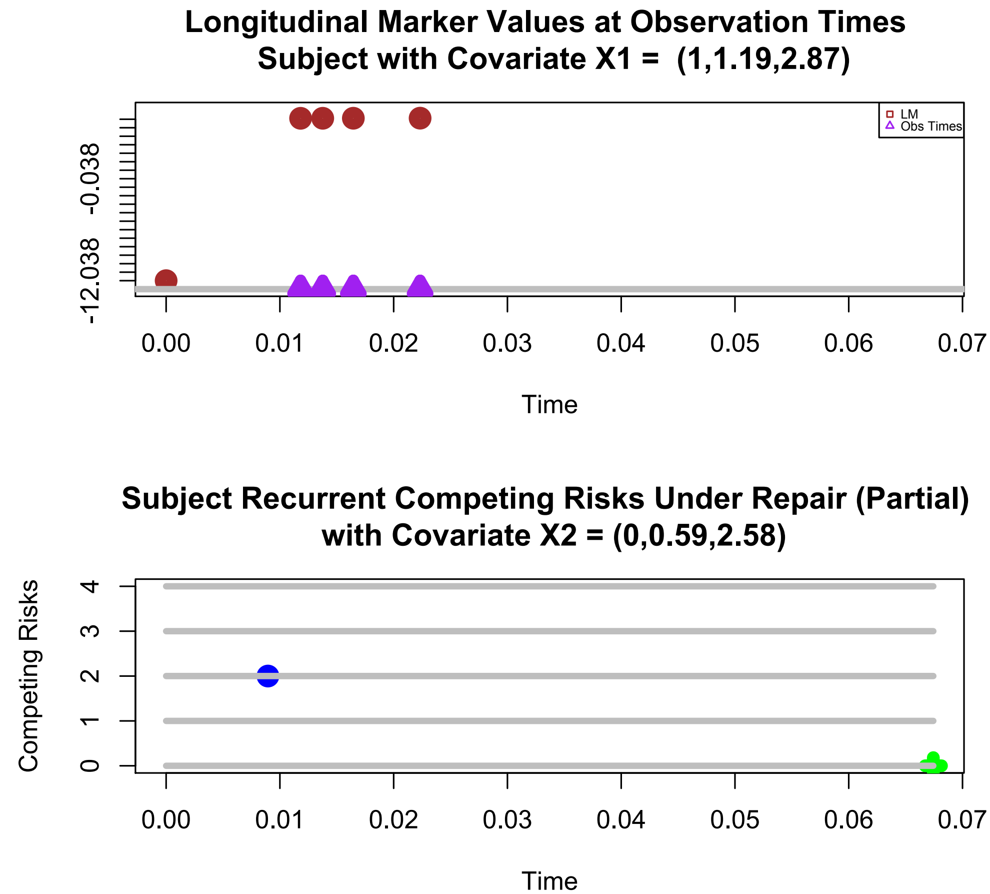

Figure 4.2: Data of a Single Unit Observed Under Partial Repair (Censored TE)

monitoring period τ_i while in Figure 4.2, the TE of the unit is censored. We assume that the random monitoring time τ_i is independent of the data processes, and follows an Exponential distribution with some parameter.

4.3 Model Description

The LM Submodel

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be some probability space. For a single unit i, define $\mathbf{F} = \{\mathcal{F}_s | 0 \leq s \leq s^*\}$ a history or filtration on the same probability space. $N^l_{qi}(s)$ and $N^l_{0i}(s)$ are counting processes, $W_i(s)$ is the longitudinal process, and $Y^l_i(s)$ is predictable processes with respect to $\mathbf{F}$.

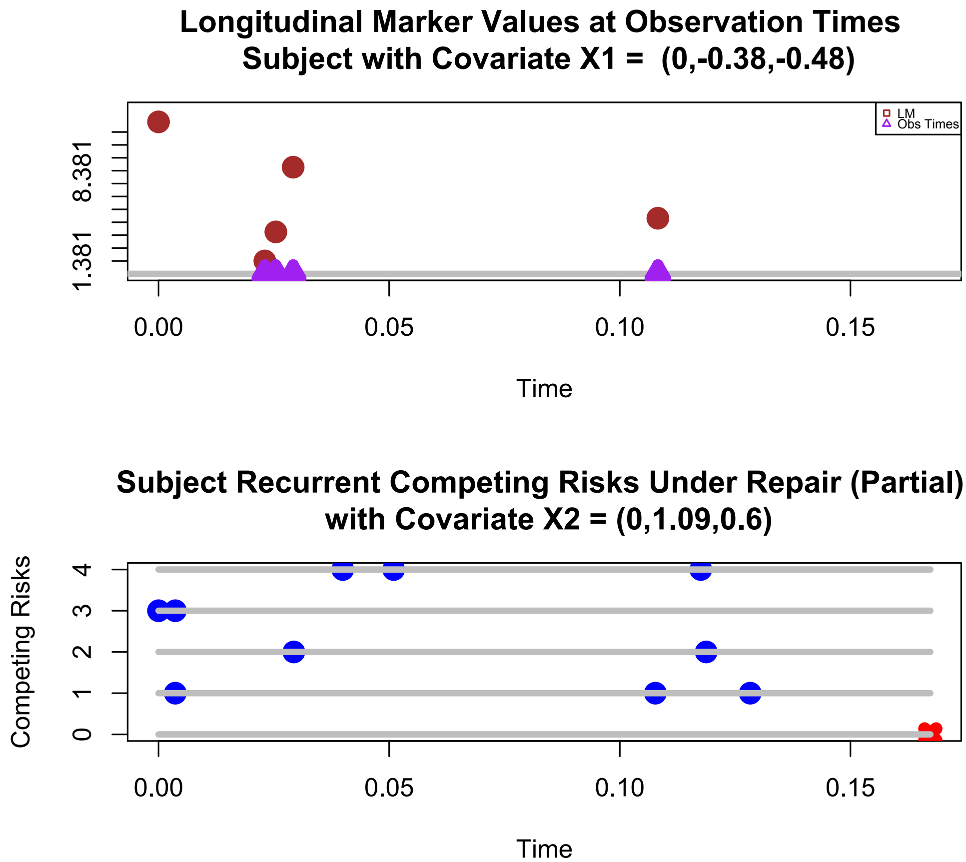

Figure 4.3: Data of a Single Unit Observed Under Partial Repair (Not Censored TE)

Conditional on the history $\mathbf{F}$, the mean of the longitudinal process takes into account fixed intercept, linear time trend, fixed covariates in X_{i1}, and impact of past recurrent competing event occurrences. Frailty variable Z_i is also the random effect in the mixed model. Accounting for possible association between LM measurements within the same subject, the mean process of this LM conditional model has to consider previous LM measurement values. Although the longitudinal marker is continuous, we only observe its values at $t_{ij} \geq 0, j = 1, 2, \cdots, m_i$, which are predetermined discrete times, for the ith observational unit. So, for $j = 1, 2, \cdots, m_i$, the submodel for LM is

$$
\begin{aligned}
W_i(t_{ij})|(\Theta^l, \mathcal{F}_{t_{ij}-}, Z_i) = \beta_0^l + \beta_t^l t_{ij} + X_{i1}\beta_1^l + \zeta^l Z_i + \mathbf{N}_i^l(t_{ij}-)\xi^l \\
+ \delta\Big[W_i(t_{ij-1}) - (\beta_0^l + \beta_t^l t_{ij-1} + X_{i1}\beta_1^l \\
+ \zeta^l Z_i + \mathbf{N}_i^l(t_{ij-1}-)\xi^l)\Big] + \epsilon_i(t_{ij})
\end{aligned}
\tag{4.1}
$$

and let

$$
\Theta^l = \{\beta_0^l, \beta_t^l, \beta_1^l, \zeta^l, \xi^l, \delta, \nu, \sigma\} \tag{4.2}
$$

where β_1^l and ξ^l are $p \times 1$ and $Q \times 1$ vectors, respectively. Assume for $v \geq 0, \epsilon_i(v) \overset{iid}{\sim} N(0, \sigma^2)$, random effect $Z_i \overset{iid}{\sim} N(0, \nu^2)$. Define $\mathbf{L}(v-) = \max\{t_{ij} : t_{ij} \leq v-, v \geq 0, j = 1, 2, \cdots, m_i\}$, and

$$
\mu_i(v|Z_i, \Theta^l, \mathcal{F}_{v-}) = \beta_0^l + \beta_t^l v + X_{i1}^t \beta_1^l + \zeta^l Z_i + \mathbf{N}_i^l(v-)\xi^l \tag{4.3}
$$

Subsequently, we write equation 4.1 in a more compact fashion

$$
\begin{aligned}
W_i(t_{ij})|(\Theta^l, \mathcal{F}_{t_{ij}-}, Z_i) = \mu_i(t_{ij}|Z_i, \Theta^l, \mathcal{F}_{t_{ij}-}) + \delta\Big[W_i(\mathbf{L}(t_{ij}-)) \\
- \mu_i(\mathbf{L}(t_{ij}-)|Z_i, \Theta^l, \mathcal{F}_{t_{ij}-})\Big] + \epsilon_i(t_{ij})
\end{aligned}
\tag{4.4}
$$

The RCR Submodel

History of the LM is accounted for in the intensity process of RCR. We consider a similar setting for the RCR as in Chapter 2. For $q = 1, 2, \cdots, Q$,

$$
\begin{aligned}
a_{qi}^l(v|Z_i)dv \equiv & \mathbf{P}\{dN_{qi}^l(v) = 1|\alpha_q^l, \beta_q^l, \zeta_q^l, \eta_q, \mathcal{F}_{v-}, Z_i\} = Y_i^l(v)\rho_q(\mathbf{N}_i^l(v-); \alpha_q^l)\lambda_{q0}(\mathcal{E}_{qi}(v)) \\
& \times \exp(X_{i2}\beta_q^l + \zeta^q Z_i + W_i[\mathbf{L}(v-)]\eta_q)dv
\end{aligned}
$$

We only measure the continuous longitudinal process $W_i(v)$ at discrete time points, at any time $v \in [0, \min(\tau_i, T_i))$, where $T_i = s \wedge \tau_i$. Conditional on $\mathcal{F}_{v-}$, the observed value of LM that affects the probability of a new RCR event occurrence is measured at t_{ij}, where the predetermined t_{ij} is the closest point in time to v.

The covariate vector X_{i2} is $p_2 \times 1$, and includes possible treatment assignment and other demographic characteristics of unit i that is different from those in X_{i1} in the LM submodel. In this dynamic model for RCRs, both past event occurrences and longitudinal marker values affect the probability of a new RCR event occurrence conditional on the history. Thus,

$$\begin{aligned} A^l_{qi}(s|Z_i) = \int_0^s & Y^l_i(v)\rho_q(\mathbf{N}^l_i(v-);\alpha^l_q)\lambda_{q0}(\mathcal{E}_{qi}(v)) \\ & \times \exp(X_{i2}\beta^l_q + \zeta^q Z_i + W_i[\mathbf{L}(v-)]\eta_q)dv \end{aligned} \tag{4.5}$$

The TE Submodel

History of the LM is also accounted for in the TE submodel. In a way that is similar to the specification of the RCR submodel, the intensity process of the TE is

$$\begin{aligned} a^l_{0i}(v|Z_i)dv \equiv & \mathbf{P}\{dN^l_{0i}(v) = 1|\gamma^l_0, \beta^l_2, \zeta_0, \eta_0, \mathcal{F}_{v-}, Z_i\} = Y^l_i(v)\rho_0(\mathbf{N}^l_i(v-);\gamma^l_0)\lambda_0(v) \\ & \times \exp(X_{i2}\beta^l_2 + \zeta_0 Z_i + W_i[\mathbf{L}(v-)]\eta_0)dv \end{aligned}$$

where ζ_0 and η_0 are the one-dimensional coefficients of the random effect Z_i and longitudinal marker $W_i[\mathbf{L}_i(v-)]$. Thus,

$$\begin{aligned} A^l_{0i}(s|Z_i) = \int_0^s & Y^l_i(v)\rho_0(\mathbf{N}^l_i(v-);\gamma^l_0)\lambda_0(v) \\ & \times \exp(X_{i2}\beta^l_2 + \zeta_0 Z_i + W_i[\mathbf{L}(v-)]\eta_0)dv \end{aligned} \tag{4.6}$$

4.4 Generalized At-Risk Processes of LM/RCR/TE

We estimate the RCR baseline cumulative hazard probabilities $\Lambda^l_{q0}(s), q = 1, 2, \cdots, Q$ at calendar time s. The frailty variable $Z_i, i = 1, 2, \cdots, n$, are not observable. Given the Z_is, the doubly-indexed processes are:

$$\begin{aligned} N^l_{qi}(s,t) &= \int_0^s I(\varepsilon_{qi}(v) \le t)N^l_{qi}(dv) \\ A^l_{qi}(s,t|Z_i) &= \int_0^s I(\varepsilon_{qi}(v) \le t)A^l_{qi}(dv|Z_i) \\ M^l_{qi}(s,t|Z_i) &= \int_0^s I(\varepsilon_{qi}(v) \le t)M^l_{qi}(dv|Z_i) \end{aligned}$$

Proposition 2. *For $i = 1, 2, \cdots, n$, given $\boldsymbol{Z} = (Z_1, Z_2, \cdots, Z_n)^t$, $A^l_{qi}(s,t) = \int_0^t Y^l_{qi}(s,w)\lambda_{q0}(w)dw$, where*

$$Y^l_{qi}(s,w|Z_i) = \sum_{k=1}^{N^l_{qi}(s\bigwedge\tau_i)+1} I\{w \in (\varepsilon_{qik}(s_{ik-1}), \varepsilon_{qik}(s_{ik})]\} \frac{\kappa^l_{qi}(\varepsilon^{-1}_{qik}(w))}{\varepsilon'_{qik}(\varepsilon^{-1}_{qik}(w))}, \tag{4.7}$$

and $\kappa^l_{qi}(s|\Theta) = \rho_q(\boldsymbol{N}^l_i(s-);\alpha^l_q)\exp\{X^t_{i2}\beta^l_q + \zeta^q Z_i + W_i(\boldsymbol{L}[s-])\eta_q\}, q = 1, 2, \cdots, Q$; and $\mathcal{E}^{-1}_{qik}(.)$ is the inverse of $\mathcal{E}_{qik}(v) = \mathcal{E}_{qi}(v) I\{v \in (S_{ik-1}, S_{ik}]\}$.

4.5 The Complete Likelihood

Denote $\Theta = \Theta^l \cup \Theta_1 \cup \Theta_2$, where $\Theta_1 = \Big\{\Big(\Lambda^l_{q0}(s), q = 1, 2, \cdots, Q; \Lambda^l_0(s)\Big) : 0 < s < s^*\Big\}$ and $\Theta_2 = \Big\{(\alpha^l_q, \beta^l_q, \zeta_q, \eta_q, q = 1, 2, \cdots, Q); (\gamma^l_0, \beta^l_2, \zeta_0, \eta_0)\Big\}$. Given the random effect vector $\mathbf{Z}$, we write the complete likelihood process of the joint dynamic models as

$$\begin{aligned}\mathcal{L}^c(\mathbf{D}(s)|\mathbf{Z},\Theta) = \prod_{i=1}^{n} \Big\{ \prod_{v=0}^{s} \Big\{ \prod_{j=1}^{m_i} \Big[\mathcal{L}^l_i\Big(W_i(v)|Z_i, \Theta^l, \mathcal{F}_{v-}\Big)\Big]^{I(v=t_{ij})}\Big\} \\ \times \mathcal{L}^R_i(dv|Z_i,\Theta,\mathcal{F}_{v-})\mathcal{L}^0_i(dv|Z_i,\Theta,\mathcal{F}_{v-})\Big\}\end{aligned} \tag{4.8}$$

where $\mathbf{D}(s)$ is the data vector of all observations on $[0, s)$, and $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_n)^T$. Given $\mathbf{Z}$, the likelihood of LM at time v is

$$\begin{aligned}\mathcal{L}^l_i(W_i(v)|Z_i,\Theta^l,\mathcal{F}_{v-}) = (2\pi\sigma)^{-\frac{1}{2}} \exp -\Big\{\frac{1}{2\sigma^2}\Big(W_i(v) - \mu_i(v|Z_i,\Theta^l,\mathcal{F}_{v-}) \\ - \delta\Big[W_i(\mathbf{L}(v-)) - \mu_i(\mathbf{L}(v-)|Z_i,\Theta^l,\mathcal{F}_{v-})\Big]\Big)^2\Big\},\end{aligned}$$

where $\mu_i(v|Z_i,\Theta^l,\mathcal{F}_{v-})$, is defined in equation 4.3 for $\mu \geq 0$. Additionally, we write the likelihood of the RCR submodel given Z_i as

$$\mathcal{L}^R_i(dv|Z_i,\Theta,\mathcal{F}_{v-}) = \prod_{q=1}^{Q}[dA^l_{qi}(v)]^{N^l_{qi}(dv)}[1 - dA^l_{qi}(v)]^{1-N^l_{qi}(dv)},$$

where $A^l_{qi}(s)$ is defined as in equation (4.5), and for the RCR submodel,

$$\mathcal{L}^0_i(dv|Z_i,\Theta,\mathcal{F}_{v-}) = [dA^l_{0i}(v)]^{N^l_{0i}(dv)}[1 - dA^l_{0i}(v)]^{1-N^l_{0i}(dv)},$$

where $A^l_{0i}(s)$ is defined as in equation (4.6).

Estimation of the Baseline Cumulative Hazard Probabilities

If we know the values of the finite parameters in Θ_2, given $\mathbf{Z}$, we can develop a Nelson - Aalen type of estimators of baseline cumulative hazard probabilities for both RCRs and TE. The baseline cumulative hazard probabilities for RCR is

$$\hat{\Lambda}_{q0}(s,t|\Theta_2,\mathbf{Z}) = \int_0^t \frac{\sum_{i=1}^n N_{qi}^l(s,dw)}{S_{q0}^l(s,w|\Theta_2,\mathbf{Z})} \tag{4.9}$$

with the *aggregated generalized at-risk* process of the *qth* RCR

$$S_{q0}^l(s,w|\Theta_2,\mathbf{Z}) = \sum_{i=1}^n Y_{qi}^l(s,w|Z_i) \tag{4.10}$$

where $Y_{qi}^l(s,w|Z_i)$ is the generalized at-risk process for unit i derived in Proposition 2. The baseline cumulative hazard probabilities for TE

$$\hat{\Lambda}_0(v|\Theta_2,\mathbf{Z}) = \int_0^t \frac{\sum_{i=1}^n N_{0i}^l(dv)}{S_0^l(v|\Theta_2,\mathbf{Z})} \tag{4.11}$$

with the *aggregated generalized at-risk* process of the TE

$$S_0^l(v|\Theta_1,\mathbf{Z}) = \sum_{i=1}^n Y_i^l(v|Z_i) \tag{4.12}$$

where $Y_i^l(v|Z_i)$ is the at-risk process for unit i. Similar to Proposition 2, we define for the TE submodel

$$\kappa_{i0}^l(s|\Theta) = \rho_0(\mathbf{N}_i^l(s-);\alpha_0^l)\exp\{X_{i2}^t\beta_0^l + \zeta_0 Z_i + W_i(\mathbf{L}[s-])\eta_0\} \tag{4.13}$$

The PLEs of RCR and TE

We can also compute the PLEs of the baseline survival functions given $\mathbf{Z} = \mathbf{z}$ and values of the finite-dimensional parameters. For $q = 1, 2, \cdots, Q$,

$$\hat{\bar{F}}_{q0}^l(s,t|\mathbf{z},\alpha_q^l,\beta_q^l,\zeta_q^l,\eta_q) = \prod_{w=0}^t [1 - \hat{\Lambda}_{q0}^l(s,dw|\Theta_2,\mathbf{z})] \tag{4.14}$$

For the TE,

$$\hat{\bar{F}}_0^l(v|\mathbf{z},\gamma_0^l,\beta_2^l,\zeta_0,\eta_0) = \prod_{w=0}^t [1 - \hat{\Lambda}_0^l(dw|\Theta_2,\mathbf{z})] \tag{4.15}$$

Estimation of the Finite Dimensional Parameters

A Profile Likelihood

From equation 4.8 , only the RCR and TE submodel likelihoods involve the infinite - dimensional parameters in Θ_1. Subsituting the expressions in 4.9 and 4.11 into the complete likelihood in equation (4.8) , the finite-dimensional parameters in $\Theta^l \cup \hat{\Theta}_2$ are obtained by maximizing the following likelihood

$$\begin{aligned} L_p^l(\Theta^l \cup \Theta_2|\mathbf{D(s)},\, \mathbf{Z}) = \prod_{i=1}^{n} \Big\{ & \prod_{j=1}^{m_i} \mathcal{L}_i^l\Big(W_i(t_{ij})|Z_i, \Theta^l, \mathcal{F}_{t_{ij}-}\Big)\Big\} \\ & \times \Bigg[\prod_{q=1}^{Q} \prod_{k=1}^{N_{qi}^l(s\wedge\tau_i)} \bigg(\frac{\kappa_{qi}^l(S_{ik}|\Theta_2)}{S_{q0}^l(s, \varepsilon_{qij}(S_{ik})|\mathbf{Z}, \Theta_2)} \bigg)^{N_{qi}^l(dS_{ik})} \Bigg] \\ & \times \bigg(\frac{\kappa_{i0}^l(T_i|\Theta_2)}{S_0^l(T_i|\mathbf{Z}, \Theta_2)} \bigg)^{N_{0i}^l(dT_i)} \Bigg\} \end{aligned} \tag{4.16}$$

We maximize the likelihood $L_p^l(\Theta^l \cup \Theta_2|\mathbf{D(s)}, \mathbf{Z})$ in the above equation, which is the product of the LM submodel likelihood and *partial likelihoods* of the RCR and TE submodels.

The Estimating Equations

To estimate $\Theta^l \cup \Theta_2$, we obtain the following estimating equations. We find $\hat{\Theta}^l$ and $\hat{\Theta}_2$ by differentiating the likelihood in equation 4.16. We denote that $\hat{Z}_i = \mathbf{E}[Z_i|\mathbf{D}_i(s), \Theta]$, and $\hat{Z}_i^2 = \mathbf{E}[Z_i^2|\mathbf{D}_i(s), \Theta]$, respectively.

$$
\begin{aligned}
\hat{\nu} &= \sqrt{\frac{\sum_{i=1}^{n} \hat{Z}_i^2}{n}} \\
\hat{\sigma} &= \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}[\hat{R}(t_{ij}) - (1-\hat{\delta})\hat{\zeta}^l \hat{Z}_i]^2}{\sum_{i=1}^{n} m_i}} \\
\hat{\beta}_0^l &= \frac{1}{(1-\hat{\delta})\sum_{i=1}^{n} m_i}\Big\{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big([W_i(t_{ij}) - \hat{\delta}W_i(t_{ij-1})] - [\mathbf{N}_i^l(t_{ij}-) - \hat{\delta}\mathbf{N}_i^l(t_{ij-1}-)] \\
&\quad - (t_{ij} - \hat{\delta}t_{ij-1})\hat{\beta}_t^l - (1-\hat{\delta})(\hat{\zeta}^l \hat{Z}_i + X_{1i}^t \hat{\beta}_1^l)\Big)\Big\} \\
\hat{\beta}_t &= \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{m_i}(t_{ij} - \hat{\delta}t_{ij-1})}\Big\{\sum_{i=1}^{n}\sum_{j=1}^{m_i}[W_i(t_{ij}) - \hat{\delta}W_i(t_{ij-1})] - [\mathbf{N}_i^l(t_{ij}-) \\
&\quad - \hat{\delta}\mathbf{N}_i^l(t_{ij-1}-)] - (1-\hat{\delta})(\hat{\beta}_0^l + \hat{\zeta}^l \hat{Z}_i + X_{1i}^t \hat{\beta}_1^l))]\Big\} \\
\hat{\beta}_1^l &= \frac{1}{(1-\hat{\delta})}\Big\{\Big(\sum_{i=1}^{n}\sum_{j=1}^{m_i}[W_i(t_{iij}) - \hat{\delta}W_i(t_{ij-1})] - [\mathbf{N}_i^l(t_{ij}-) - \hat{\delta}\mathbf{N}_i^l(t_{ij-1}-)] \\
&\quad - (t_{ij} - \hat{\delta}t_{ij-1})\hat{\beta}_t^l - (1-\hat{\delta})(\hat{\beta}_0^l + \hat{\zeta}^l \hat{Z}_i)\Big)\Big(\sum_{i=1}^{n} m_i X_{1i}^{-1}\Big)\Big\} \\
\hat{\zeta}^l &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\hat{R}(t_{ij})\hat{Z}_i}{(1-\hat{\delta})\sum_{i=1}^{n}\hat{Z}_i m_i} \\
\hat{\delta} &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}[W_i(t_{ij}) - \mu_i(t_{ij}|\hat{Z}_i, \hat{\Theta}^l, \mathcal{F}_{t_{ij}-})][W_i(t_{ij-1}) - \mu_i(t_{ij-1}|\hat{Z}_i, \hat{\Theta}^l, \mathcal{F}_{t_{ij-1}-})]}{\sum_{i=1}^{n}\sum_{j=1}^{m_i}[W_i(t_{ij-1}) - \mu_i(t_{ij-1}|\hat{Z}_i, \hat{\Theta}^l, \mathcal{F}_{t_{ij-1}-})]^2} \\
\hat{\xi}^l &= \Big(\sum_{i=1}^{n}\sum_{j=1}^{m_i}[W_i(t_{ij}) - \hat{\delta}W_i(t_{ij-1})] - (t_{ij} - \hat{\delta}t_{ij-1})\hat{\beta}_t^l - (1-\hat{\delta})(\hat{\beta}_0^l + X_{1i}^t \beta_1^l \\
&\quad + \hat{\zeta}^l \hat{Z}_i)\Big)\Big([\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathbf{N}_i^l(t_{ij}-) - \hat{\delta}\mathbf{N}_i^l(t_{ij-1}-)]^{-1}\Big)
\end{aligned}
\tag{4.17}
$$

$$
\begin{aligned}
\sum_{i=1}^{n}\int_{0}^{\tau_i}\Big[\frac{\frac{\partial}{\partial\alpha_q^l}\rho_q(N_{qi}^l(v-);\alpha_q^l))}{\rho_q(N_{qi}^l(v-);\alpha_q^l))}-\frac{\frac{\partial}{\partial\alpha_q^l}S_{q0}^l(s,\mathcal{E}_{qi}(v)|\Theta_2)}{S_{q0}^l(s,\mathcal{E}_{qi}(v)|\Theta_2)}\Big]N_{qi}^l(dv)&=0\\
\sum_{i=1}^{n}\int_{0}^{\tau_i}\Big[X_{i2}-\frac{\frac{\partial}{\partial\beta_q^l}S_{q0}^l(s,\mathcal{E}_{qi}(v)|\Theta_2)}{S_{q0}^l(s,\mathcal{E}_{qi}(v)|\Theta_2)}\Big]N_{qi}^l(dv)&=0\\
\sum_{i=1}^{n}\int_{0}^{\tau_i}\Big[Z_i-\frac{\frac{\partial}{\partial\zeta_q}S_{q0}^l(s,\mathcal{E}_{qi}(v)|\Theta_2)}{S_{q0}^l(s,\mathcal{E}_{qi}(v)|\Theta_2)}\Big]N_{qi}^l(dv)&=0\\
\sum_{i=1}^{n}\int_{0}^{\tau_i}\Big[W_i(v)-\frac{\frac{\partial}{\partial\eta_q}S_{q0}^l(s,\mathcal{E}_{qi}(v)|\Theta_2)}{S_{q0}^l(s,\mathcal{E}_{qi}(v)|\Theta_2)}\Big]N_{qi}^l(dv)&=0\\
\sum_{i=1}^{n}\int_{0}^{\tau_i}\Big[\frac{\frac{\partial}{\partial\gamma_0^l}\rho_0(\mathbf{N}_i^l(v-);\gamma_0^l)}{\rho_0(\mathbf{N}_i^l(v-);\gamma_0^l)}-\frac{\frac{\partial}{\partial\gamma_0^l}S_0^l(v|\Theta_2)}{S_0^l(v|\Theta_2)}\Big]N_{0i}^l(dv)&=0 \qquad (4.18)\\
\sum_{i=1}^{n}\int_{0}^{\tau_i}\Big[X_{i2}-\frac{\frac{\partial}{\partial\beta_2^l}S_0^l(v|\Theta_2)}{S_0^l(v|\Theta_2)}\Big]N_{0i}^l(dv)&=0\\
\sum_{i=1}^{n}\int_{0}^{\tau_i}\Big[Z_i-\frac{\frac{\partial}{\partial\zeta_0}S_0^l(v|\Theta_2)}{S_{q0}^l(v|\Theta_2)}\Big]N_{0i}^l(dv)&=0\\
\sum_{i=1}^{n}\int_{0}^{\tau_i}\Big[W_i(v)-\frac{\frac{\partial}{\partial\eta_0}S_0^l(v|\Theta_2)}{S_0^l(v|\Theta_2)}\Big]N_{0i}^l(dv)&=0
\end{aligned}
$$

An EM Algorithm

We develop an EM algorithm to estimate $\Theta^l \cup \Theta_2$ due to the random effect $\mathbf{Z}$. We first find the log of the full likelihood:

$$
\begin{aligned}
\log[\mathcal{L}^{\dagger}(\mathbf{D}(s),\mathbf{Z}|\Theta)] =& \log[\mathcal{L}^c(\mathbf{D}(s)|\mathbf{Z},\Theta)]+\log[f(\mathbf{Z}|\nu)]\\
=&-\frac{n}{2}\log(\nu^2)-\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\log\sigma^2}{2}-\sum_{i=1}^{n}\Big(\frac{Z_i^2}{2\nu^2}+\sum_{j=1}^{m_i}\frac{1}{2\sigma^2}[R(t_{ij})\\
&-(1-\delta)\zeta^l Z_i]^2\Big)+\sum_{i=1}^{n}\Big(\sum_{q=1}^{Q}\sum_{k=1}^{N_{qi}^l(s\wedge\tau_i)}\zeta_q+\zeta_0 N_{0i}^l(s\wedge\tau_i)\Big)Z_i \qquad (4.19)\\
&-\sum_{i=1}^{n}\Big(\sum_{q=1}^{Q}\exp(\zeta_q Z_i)\int_0^s A_{qi}^{-}(dv)+\exp(\zeta_0 Z_i)\int_0^s A_{0i}^{-}(dv)\Big)\\
&+\sum_{i=1}^{n}\Big[\int_0^s\sum_{q=1}^{Q}\log(a_{qi}^{-}(v))N_{qi}^l(dv)+\int_0^s\log(a_{0i}^{-}(v))N_{0i}^l(dv)\Big]
\end{aligned}
$$

and we define

$$R_i(t_{ij}) = \Big(W_i(t_{ij}) - \delta W_i(t_{ij-1})\Big) - \Big(\mathbf{N}^\dagger(t_{ij-}-)\xi^l - \delta \mathbf{N}^\dagger(t_{ij-1}-)\xi^l\Big) - (t_{ij}\beta_t^l - \delta t_{ij-1}\beta_t^l) - (1-\delta)(\beta_0^l + X_{i1}^t\beta_1^l)$$

$$A_{qi}^-(s) = \int_0^s Y_i^l(v)\rho_q(\mathbf{N}^l(v-);\alpha_q^l)\lambda_{q0}(\mathcal{E}_{qi}(v))\exp(X_{i2}\beta_q^l + W_i[\mathbf{L}(v-]\eta_q)dv$$

$$a_{qi}^-(v)dv = Y_i^l(v)\rho_q(\mathbf{N}^l(v-);\alpha_q^l)\lambda_{q0}(\mathcal{E}_{qi}(v))\exp(X_{i2}\beta_q^l + W_i[\mathbf{L}(v-]\eta_q)dv$$

$$A_{i0}^-(s)dv = \int_0^s Y_i^l(v)\rho_0(\mathbf{N}^l(v-);\gamma_0^l)\lambda_0(v)\exp(X_{i2}\beta_0^l + W_i[\mathbf{L}(v-]\eta_0)dv$$

$$a_{i0}^-(v)dv = Y_i^l(v)\rho_0(\mathbf{N}^l(v-);\gamma_0^l)\lambda_0(v)\exp(X_{i2}\beta_0^l + W_i[\mathbf{L}(v-]\eta_0)dv$$

- **The E-step** For $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_n)^t$, and $\mathcal{L}^c(\mathbf{D}_i(s)|Z_i,\Theta)$ is the complete data likelihood in equation 4.8.

$$\mathbf{E}[Z_i|\mathbf{D}_i(s),\Theta] = \int_{-\infty}^{\infty} Z_i f(Z_i|\mathbf{D}_i(s),\Theta) = \frac{\int_{-\infty}^{\infty} Z_i\mathcal{L}^c(\mathbf{D}_i(s)|Z_i,\Theta)f(Z_i|\nu)dZ_i}{\int_{-\infty}^{\infty}\mathcal{L}^c(\mathbf{D}_i(s)|Z_i,\Theta)f(Z_i|\nu)dZ_i} \tag{4.20}$$

$$\mathbf{E}[Z_i^2|\mathbf{D}_i(s),\Theta] = \int_{-\infty}^{\infty} Z_i^2 f(Z_i|\mathbf{D}_i(s),\Theta) = \frac{\int_{-\infty}^{\infty} Z_i^2\mathcal{L}^c(\mathbf{D}_i(s)|Z_i,\Theta)f(Z_i|\nu)dZ_i}{\int_{-\infty}^{\infty}\mathcal{L}^c(\mathbf{D}_i(s)|Z_i,\Theta)f(Z_i|\nu)dZ_i} \tag{4.21}$$

  We approximate $\mathbf{E}[\exp(Z_i)|\mathbf{D}_i(s),\Theta]$ using the Metropolis-Hasting Algorithm.

- **The M-step** Given $\hat{\Theta}^l, \hat{\Theta}_2$ and the approximations in the E-step, we will be able to compute the estimates of the cumulative baseline hazard in equations 4.9 and 4.11. We use the notations below:

$$\widehat{Z}_i = \mathbf{E}[Z_i|\mathbf{D}_i(s),\Theta]; \quad \widehat{\exp(Z_i)} = \mathbf{E}[\exp(Z_i)|\mathbf{D}_i(s),\Theta]$$

  We describe the steps of the algorithm:

  0. Initialize $\hat{\Theta}^{l,0}, \hat{\Theta}_2^0$, and $\hat{\mathbf{Z}}^0$. Use $\hat{\Theta}_2^0$, and $\hat{\mathbf{Z}}^0$ to obtain $\hat{\Theta}_1^0 = \Big\{\Big(\hat{\Lambda}_{q0}(s), q = 1,2,\cdots,Q; \hat{\Lambda}_0(s)\Big)|0 < s < s^*\Big\}$ and set $\hat{\Theta}_2^{old} = \hat{\Theta}_2^0; \hat{\mathbf{Z}}^{old} = \hat{\mathbf{Z}}^0$.
  1. Maximize $L_p^l(\Theta|\mathbf{D(s)})$ in equation (4.16) with respect to Θ_2, and obtain $\hat{\Theta}_2^{new}$. See the estimating equations in 4.18.

2. Compute $\widehat{\mathbf{Z}^2}^0$ and $\widehat{\exp(\mathbf{Z})}^0$.
3. Obtain $\hat{\Theta}^{l,new}$ by maximizing the full likelihood in equation (4.19). See the estimating equations in 4.17 and set $\hat{\Theta}^{l,old} = \hat{\Theta}^{l,0}$.
4. Update $\hat{\mathbf{Z}}^{old}$ to $\hat{\mathbf{Z}}^{new}$ through approximations in the **E** step.
5. Update values of $\hat{\Theta}_1^{old}$ to $\hat{\Theta}_1^{new}$ using $\hat{\Theta}_2^{new}$ and $\hat{\mathbf{Z}}^{new}$.
6. Repeat step 1 - 4 until $|\hat{\Theta}^{l,new} \cup \hat{\Theta}_2^{new} - \hat{\Theta}^{l,old} \cup \hat{\Theta}_2^{old}| <$ tol, where tol is a very small number.

The computational aspect of the proposed model will be dealt with in the future: the R code that implement the inference procedure are being developed, and tested. Simulation studies that investigate large sample properties of the estimators will be performed as well.

Chapter 5

Conclusion

The joint dynamic models of recurrent competing risks (RCR) and a terminal event (TE) provides a semiparametric procedure to model possible dependence between the RCR and TE. The models contribute to existing literature under the joint model framework by accounting for the impact of past event occurrences on the intensity processes, as well as incorporating effective age processes to model possible interventions during the monitoring period. Additionally, the dynamic aspect of predicting terminal event survival probabilities of a new unit from the same population offers a possible prognostic tool for cancer research and precision medicine.

The joint dynamic models of RCR and TE with frailty case is considered to include possible associations between different event processes induced by unobserved variables. For the frailty case, more simulation studies are needed to definitively conclude the quality and performance of the estimation procedure.

When there is association between a longitudinal marker (LM) and multiple time-to event processes, simultaneously modeling a longitudinal marker, recurrent competing risks and a terminal event presents greater challenges for research in joint modeling (than compared to joint modeling only RCR and TE, for example). We propose dynamic joint models of LM, RCR and TE that take into account the dynamic aspects of the data processes considered in the joint models of just RCR and TE, and that model the dependence structure between the different data processes.

All of the joint models proposed in the dissertation can be applied to areas including but not limited to biomedical, clinical and reliability type of research. The dynamic aspect of the models, and the possible prognostic tool of predicting terminal event survival probabilities are appropriate for research in cancer and precision medicine. Future research include implementing the proposed inference procedure for the LM/RCR/TE model, and extensive simulation studies to examine the properties of the estimators. Since predicting the survival probabilities of the terminal event process is of high interest in applications, we are also interested in developing a dynamic way of prediction for the LM/RCR/TE model.

Bibliography

[1] Odd Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726, 1978.

[2] Odd O. Aalen, Ørnulf Borgan, and Håkon K. Gjessing. *Survival and event history analysis*. Statistics for Biology and Health. Springer, New York, 2008. A process point of view.

[3] Odd O Aalen and Einar Husebye. Statistical analysis of repeated events forming renewal processes. *Statistics in medicine*, 10(8):1227–1240, 1991.

[4] Paul S Albert, Lisa M McShane, and Joanna H Shih. Latent class modeling approaches for assessing diagnostic error without a gold standard: with applications to p53 immunohistochemical assays in bladder tumors. *Biometrics*, 57(2):610–619, 2001.

[5] P. Andersen, O. Borgan, R. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer-Verlag, New York, 1993.

[6] Per Kragh Andersen and Richard D Gill. Cox's regression model for counting processes: a large sample study. *The annals of statistics*, pages 1100–1120, 1982.

[7] Eleni-Rosalina Andrinopoulou, Dimitris Rizopoulos, Johanna JM Takkenberg, and Emmanuel Lesaffre. Joint modeling of two longitudinal outcomes and competing risk data. *Statistics in medicine*, 33(18):3167–3178, 2014.

[8] Paul Blanche, Cécile Proust-Lima, Lucie Loubère, Claudine Berr, Jean-François Dartigues, and Hélène Jacqmin-Gadda. Quantifying and comparing dynamic

predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics*, 71(1):102–113, 2015.

[9] Elizabeth R Brown, Joseph G Ibrahim, and Victor DeGruttola. A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73, 2005.

[10] D. R. Cox. Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B*, 34:187–220, 1972. With discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox.

[11] David Roxbee Cox and David Oakes. *Analysis of survival data*, volume 21. CRC Press, 1984.

[12] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[13] Emmanuelle Deslandes and Sylvie Chevret. Joint modeling of multivariate longitudinal data and the dropout process in a competing risk setting: application to icu data. *BMC medical research methodology*, 10(1):69, 2010.

[14] Jason P Fine and Robert J Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, 94(446):496–509, 1999.

[15] Jason P Fine, Hongyu Jiang, and Rick Chappell. On semi-competing risks data. *Biometrika*, 88(4):907–919, 2001.

[16] Lloyd D Fisher and Danyu Y Lin. Time-dependent covariates in the cox proportional-hazards regression model. *Annual review of public health*, 20(1):145–157, 1999.

[17] Thomas R Fleming and David P Harrington. *Counting processes and survival analysis*, volume 169. John Wiley & Sons, 2011.

[18] Debashis Ghosh and DY Lin. Nonparametric analysis of recurrent events and death. *Biometrics*, 56(2):554–562, 2000.

[19] Richard D Gill. Nonparametric estimation based on censored observations of a markov renewal process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 53(1):97–116, 1980.

[20] Juan R González, Edsel A Peña, and Elizabeth H Slate. Modelling intervention effects after cancer relapses. *Statistics in medicine*, 24(24):3959–3975, 2005.

[21] Xu Guo and Bradley P Carlin. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58(1):16–24, 2004.

[22] Jun Han, Elizabeth H Slate, and Edsel A Peña. Parametric latent class joint model for a longitudinal biomarker and recurrent events. *Statistics in medicine*, 26(29):5285–5302, 2007.

[23] Fushing Hsieh, Yi-Kuan Tseng, and Jane-Ling Wang. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62(4):1037–1043, 2006.

[24] Chiung-Yu Huang and Mei-Cheng Wang. Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association*, 99(468), 2004.

[25] Xin Huang, Gang Li, and Robert M Elashoff. A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements. *Statistics and its interface*, 3(2):185, 2010.

[26] Joseph G Ibrahim, Haitao Chu, and Liddy M Chen. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16):2796–2801, 2010.

[27] Bei Jiang, Michael R Elliott, Mary D Sammel, and Naisyin Wang. Joint modeling of cross-sectional health outcomes and longitudinal predictors via mixtures of means and variances. *Biometrics*, 2015.

[28] John D Kalbfleisch and Ross L Prentice. *The statistical analysis of failure time data*, volume 360. John Wiley & Sons, 2011.

[29] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[30] Sehee Kim, Donglin Zeng, Lloyd Chambless, and Yi Li. Joint models of longitudinal data and recurrent events with informative terminal event. *Statistics in biosciences*, 4(2):262–281, 2012.

[31] Aurelien Latouche, Arthur Allignol, Jan Beyersmann, Myriam Labopin, and Jason P Fine. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *Journal of clinical epidemiology*, 66(6):648–653, 2013.

[32] Aurelien Latouche, Raphaël Porcher, and Sylvie Chevret. A note on including time-dependent covariate in regression model for competing risks data. *Biometrical journal*, 47(6):807–814, 2005.

[33] Ning Li, Robert M Elashoff, Gang Li, and Jeffrey Saver. Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the ninds rt-pa stroke trial. *Statistics in medicine*, 29(5):546–557, 2010.

[34] Haiqun Lin, Bruce W Turnbull, Charles E McCulloch, and Elizabeth H Slate. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association*, 97(457):53–65, 2002.

[35] Lei Liu and Xuelin Huang. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):65–81, 2009.

[36] Lei Liu, Robert A Wolfe, and John D Kalbfleisch. A shared random effects model for censored medical costs and mortality. *Statistics in medicine*, 26(1):139–155, 2007.

[37] Audrey Mauguen, Bernard Rachet, Simone Mathoulin-Pélissier, Gaetan Mac-Grogan, Alexandre Laurent, and Virginie Rondeau. Dynamic prediction of risk of death using history of cancer recurrences in joint frailty models. *Statistics in medicine*, 32(30):5366–5380, 2013.

[38] Lisa M McCrink, Adele H Marshall, and Karen J Cairns. Advances in joint modelling: a review of recent developments with application to the survival of end stage renal disease patients. *International Statistical Review*, 81(2):249–269, 2013.

[39] Maja Miloslavsky, Sündüz Keleş, Mark J Laan, and Steve Butler. Recurrent events analysis in the presence of time-dependent covariates and dependent cen-

soring. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):239–257, 2004.

[40] Yudi Pawitan and Steve Self. Modeling disease marker processes in aids. *Journal of the American Statistical Association*, 88(423):719–726, 1993.

[41] Edsel A Peña et al. Dynamic modeling and statistical analysis of event times. *Statistical Science*, 21(4):487–500, 2006.

[42] Edsel A Peña and Myles Hollander. Models for recurrent events in reliability and survival analysis. In *Mathematical reliability: An expository perspective*, pages 105–123. Springer, 2004.

[43] Edsel A Peña, Elizabeth H Slate, and Juan R González. Semiparametric inference for a general class of models for recurrent events. *Journal of Statistical Planning and Inference*, 137(6):1727–1747, 2007.

[44] Edsel A Peña, RL Strawderman, and Myles Hollander. Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association*, 96(456):1299–1315, 2001.

[45] Edsel A Peña, Robert L Strawderman, and Myles Hollander. A weak convergence result relevant in recurrent and renewal models. In *Recent Advances in Reliability Theory*, pages 493–514. Springer, 2000.

[46] Ross L Prentice, John D Kalbfleisch, Arthur V Peterson Jr, Nancy Flournoy, VT Farewell, and NE Breslow. The analysis of failure times in the presence of competing risks. *Biometrics*, pages 541–554, 1978.

[47] Ross L Prentice, Benjamin J Williams, and Arthur V Peterson. On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379, 1981.

[48] Cécile Proust-Lima and Jeremy MG Taylor. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics*, page kxp009, 2009.

[49] Elizabeth R Brown and Joseph G Ibrahim. A bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59(2):221–228, 2003.

[50] AKM Fazlur Rahman, James D Lynch, and Edsel A Peña. Nonparametric bayes estimation of gap-time distribution with recurrent event data. *Journal of Nonparametric Statistics*, 26(3):575–598, 2014.

[51] Dimitris Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–829, 2011.

[52] Dimitris Rizopoulos, Laura A Hatfield, Bradley P Carlin, and Johanna JM Takkenberg. Combining dynamic predictions from joint models for longitudinal and time-to-event data using bayesian model averaging. *Journal of the American Statistical Association*, (just-accepted):00–00, 2014.

[53] Virginie Rondeau, Simone Mathoulin-Pelissier, Hélène Jacqmin-Gadda, Véronique Brouste, and Pierre Soubeyran. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics*, 8(4):708–721, 2007.

[54] Thomas Sellke. Weak convergence of the aalen estimator for a censored renewal process. *Statistical Decision Theory and Related Topics IV. S Gupta y J. Berger eds*, 2:183–94, 1988.

[55] Xiao Song, Marie Davidian, and Anastasios A Tsiatis. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58(4):742–753, 2002.

[56] Laura L Taylor and Edsel A Peña. Nonparametric estimation with recurrent competing risks data. *Lifetime data analysis*, 20(4):514–537, 2014.

[57] Laura L Taylor and Edsel A Peña. Parametric estimation in a recurrent competing risks model. *JIRSS: journal of the Iranian Statistical Society*, 12(1):153, 2013.

[58] AA Tsiatis, Victor Degruttola, and MS Wulfsohn. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90(429):27–37, 1995.

[59] Anastasios A Tsiatis and Marie Davidian. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88(2):447–458, 2001.

[60] Anastasios A Tsiatis and Marie Davidian. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(3):809–834, 2004.

[61] Edward F Vonesh, Tom Greene, and Mark D Schluchter. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in medicine*, 25(1):143–163, 2006.

[62] Mei-Cheng Wang and Shu-Hui Chang. Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association*, 94(445):146–153, 1999.

[63] Yan Wang and Jeremy M G Taylor. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96(455):895–905, 2001.

[64] Lee-Jen Wei, Danyu Y Lin, and L Weissfeld. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American statistical association*, 84(408):1065–1073, 1989.

[65] PR Williamson, Ruwanthi Kolamunnage-Dona, Pete Philipson, and AG Marson. Joint modelling of longitudinal and competing risks data. *Statistics in medicine*, 27(30):6426–6438, 2008.

[66] Michael S Wulfsohn and Anastasios A Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339, 1997.