Event History Analysis in Multivariate Longitudinal Data

Chaoyu Yuan

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

Abstract

Event History Analysis in Multivariate Longitudinal Data

Chaoyu Yuan

This thesis studies event history analysis in multivariate longitudinal observational databases (LODs) and its application in postmarketing surveillance to identify and measure the relationship between events of health outcomes and drug exposures. The LODs contain repeated measurements on each individual whose healthcare information is recorded electronically. Novel statistical methods are being developed to handle challenging issues arising from the scale and complexity of postmarketing surveillance LODs. In particular, the self-controlled case series (SCCS) method has been developed with two major features (1) it only uses individuals with at least one event for analysis and inference and, (2) it uses each individual to be served as his/her own control, effectively requiring a person to switch treatments during the observation period. Although this method handles heterogeneity and bias, it does not take full advantage of the observational databases. In this connection, the SCCS method may lead to a substantial loss of efficiency.

We proposed a multivariate proportional intensity modeling approach with random effect for multivariate LODs. The proposed method can explain the heterogeneity and eliminate bias in LODs. It also handles multiple types of event cases and makes full use of the observational databases. In the first part of this thesis, we present the multivariate proportional intensity model with correlated frailty. We explore the correlation structure between multiple types of clinical events and drug exposures. We introduce a multivariate Gaussian frailty to incorporate the

within-subject heterogeneity, i.e. hidden confounding factors. For parameter estimation, we adopt the Bayesian approach using the Markov chain Monte Carlo method to get a series of samples from the targeted full likelihood. We compare the new method with the SCCS method and some frailty models through simulation studies. We apply the proposed model to an electronic health record (EHR) dataset and identify event types as defined in Observational Outcomes Medical Partnership (OMOP) project. We show that the proposed method outperforms the existing methods in terms of common metrics, such as receiver operating characteristic (ROC) metrics. Finally, we extend the proposed correlated frailty model to include a dynamic random effect. We establish a general asymptotic theory for the nonparametric maximum likelihood estimators in terms of identifiability, consistency, asymptotic normality and asymptotic efficiency. A detailed illustration of the proposed method is done with the clinical event Myocardial Infarction (MI) and drug treatment of Angiotensin-converting-enzyme (ACE) inhibitors, showing the dynamic effect of unobserved heterogeneity.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor Prof. Zhiliang Ying for his continuous support of my PhD study and life. His kindness, patience and knowledge not only help in my research but also guide the direction of my life, which are my largest fortune in the past five years. I am so lucky to have him as my advisor. Besides my advisor, I would to thank Sumit Mukherjee, Cynthia G. Rush, Mengling Liu and Yang Feng for being my defense committees and providing helpful and insightful feedback.

I really appreciate that I can spend the past five years in statistics department in Columbia University. Dood Kalicharan and Anthony Cruz support me in various administrative works. Other faculty and staff also provide helpful conversations and suggestions in my study and life.

Besides, I have to show my gratitude to all good friends I make in Columbia. Both research and career benefit from the discussion with my friends and classmates. With them, I am able to get through the hardest time.

Last but not least, I would like to share my happiness with my parents, who are always encouraging and supporting me to achieve my goal in life. It is my honor to have them in my life.

Chapter 1: Introduction

1.1 Background

Event history analysis, also known as survival analysis, consists of statistical methods for the analysis of observational data on the occurrence and timing of events. In particular, an event can be defined as a qualitative change that occurs at the time during the observation period, such as death, marriage and birth. The event history analysis has found applications in many disciplines, including sociology, bio-medicine and finance. The event history analysis can be useful in longitudinal observational databases (LODs). The LODs collect repeated observations of the same study subject at the different time points. The subjects can be individuals, companies, groups and so on. The analysis based on LODs can help with exploring patterns of change and the dynamics of individual behaviors. This is important for understanding how subjects move from one situation to another. Besides, proper analysis of the LODs can provide insights into causal mechanisms between earlier history and later outcomes. Consequently, the healthcare LODs can be useful in drug safety surveillance.

Recently, increased attention has been paid to drug safety in postmarketing surveillance, whose task is to identify the unexpected adverse events (AE) caused by marketed drugs. Drug safety is an important public health challenge, which caused $4.2 - 30\%$ of hospitalizations and cost approximate 30.1 billion US dollars per year in the United States [33]. The drug safety process starts with rigorous drug preapproval procedures implemented in randomized clinical trials. The limited sample size in a clinical trial makes the detection of any adverse event difficult. The U.S. Food and Drug Administration (FDA) keeps monitoring the drug safety after regulatory market authorization, i.e. postmarketing surveillance. With much more diverse patients involved, increased attention has been brought to the postmarketing surveillance methods to solve challenging drug

safety problems.

The U.S. Food and Drug Administration (FDA) has been developing a system to identify and detect adverse events of marketed drugs and other medical products. FDA relies on submission of spontaneous reports in which clinicians or patients conclude that a drug may be the cause of some certain adverse clinical events. These records and reports are submitted voluntarily from clinicians or patients and will vary with their skills, experience, clinical history and existence of other plausible explanations. Under-reporting, reports of known reactions and false causality attribution are the common limitation of spontaneous reporting systems. Besides, spontaneous report system is a passive surveillance system and often incomplete due to the rare occurrence of uncommon adverse events and absence of control group for comparison. Some examples of spontaneous report systems are the international pharmacovigilance program of the World Health Organization (WHO), the Yellow Card Scheme of the Medicines and Healthcare Products Regulatory Agency (MHRA) and Adverse Event Reporting System (AERs) of the FDA.

Although spontaneous reporting system has its limitations, the FDA has established several analytic methods for detecting causality between adverse events and marketed drugs based on it. Disproportionality analysis is one of the most commonly used analytic methods for drug safety, mostly relying on 2×2 summary tables, e.g. Table 1.1. Disproportionality methods focus on several statistical measures of marginal association between adverse events and drug exposures. For example, the Bayesian multi-item gamma-Poisson shrinker (MGPS, DuMouchel [6]) uses the statistical measure reporting ratio (RR) (in Table 1.2), which is the number of observed occurrences with the adverse event and the drug exposure over the expected number of occurrences under independent assumption. Besides MGPS, disproportionality methods include the Bayesian confidence propagation neural network (BCPNN), the proportional reporting ratios (PRR) and the reporting odds ratios (ROR). However, these methods may lead to biased results due to ignoring the confounding factors or dynamic effects because patients may take multiple drugs at the same time and these drugs may have confounding effects.

	AE_j = Yes	AE_j = Yes	Total
$Drug_i$ = Yes	$n_{00} = 20$	$n_{01} = 100$	120
$Drug_i$ = No	$n_{10} = 100$	$n_{11} = 980$	1080
	120	1080	1200

Table 1.1: An example of 2×2 spontaneous reports

Measure	Statistical Formula
Reporting Ratio (RR)	$\frac{P(AE\|Drug)}{P(AE)}$
Proportional Reporting Ratio (PRR)	$\frac{P(AE\|Drug)}{P(AE\|notDrug)}$
Reporting Odds Ratio (ROR)	$\frac{P(AE\|Drug)/P(notAE\|Drug)}{P(AE\|notDrug)/P(notAE\|notDrug)}$
Information Component (IC)	$\log_2 \frac{P(AE\|Drug)}{P(AE)}$

Table 1.2: Common measures of association for 2×2 spontaneous reports database

The spontaneous report system is a passive surveillance system with limitations as being discussed. To overcome these limitations, a new data source for active surveillance system called longitudinal observational databases (LODs) has arisen in recent years. The information in LODs, such as insurance claim databases and electronic health record (EHR) databases, are recorded automatically instead of voluntary submission. LODs collect repeated observations of the same study subject over a period. Specifically, LODs can record information of adverse events and drug prescriptions. Based on LODs, novel machine learning and data mining algorithms have been developed for postmarketing surveillance [23, 9, 18, 26, 27, 27, 28, 32]. In particular, the self-controlled case series (SCCS) method [29, 30] was proposed for postmarketing surveillance since it is the only statistical method with a rigorous theoretical backing for dealing with LODs. We will introduce the approach with technical details in the subsequent sections.

The SCCS method is based on conditional Poisson regression to measure the causal association between adverse events and drug exposures in postmarketing surveillance. Compared with other regression methods, the SCCS method has two main features: (1) it only uses individuals with at

least one event for analysis and, (2) each individual serves as his/her own control, thus requiring each individual to have different exposure patterns.

Although SCCS related methods have achieved empirical success in postmarketing surveillance with the Observational Medical Outcomes Partnership (OMOP) project, these methods still have limitations due to model design:

- Low efficiency. The SCCS analysis only includes patients with at least one adverse event and the information from the controls is not included. In addition, each case individual needs to have different exposure patterns.

- Limited dependence structure. The SCCS methods require conditional independence assumption between multiple event occurrences. However, such an assumption is likely to be violated because the previous event occurrence can change the risk of future event. An extension of the SCCS model is given by adding a positive term to the baseline event risk [29]. This design of model may only work in very special situations and is not easy to be extended to generalized cases.

- Failure in multivariate events. In the SCCS model setting, mathematical structure fails in multivariate events unless each adverse event is treated as an independent model. Actually, this is impractical in real world problems.

To address the above issues, we propose a multivariate random effect model based on event history analysis. A central component of the multivariate random effect model is its flexible representation form to describe the casual association between multiple events and correlation structure of within-subject heterogeneity. Event history analysis, also known as survival analysis, consists of statistical methods for the analysis of observational data on the occurrence and timing of events. Below, we discuss the basic concepts and statistical models used in event history analysis.

1.2 Survival Analysis Basics

In survival analysis, we are interested in survival time which is the time to a single event for each individual, or more precisely the time elapsed from an initiating point to an event, or endpoint. Let T be a nonnegative random variable representing survival time. We have some definitions.

- **Survival function** is the probability that the time of death or endpoint is later than some specified time t:

$$S(t) = Pr(T > t).$$

- **Lifetime distribution function** $F(t)$ is the complement of the survival function and **event density function** is the rate of death or failure events per unit time:

$$\begin{aligned} F(t) =& Pr(T \leq t) = 1 - S(t), \\ f(t) =& F'(t). \end{aligned}$$

- **Hazard function** is defined as the event rate at time t conditional on survival until time t or later (that is $T \geq t$). Specifically, an item has survived at time t and we desire the probability that it will not survive for an additional time dt:

$$\begin{aligned} \lambda(t) &= \lim_{dt \to 0} \frac{1}{dt} Pr(t \leq T < t + dt | T \geq t) \\ &= \lim_{dt \to 0} \frac{Pr(t \leq t < t + dt)}{dt} \frac{1}{S(t)} \\ &= \frac{f(t)}{S(t)}. \end{aligned}$$

  Alternatively, the hazard function can be represented in terms of the cumulative hazard function:

$$\Lambda(t) = \int_0^t \lambda(s) ds = -\log S(t).$$

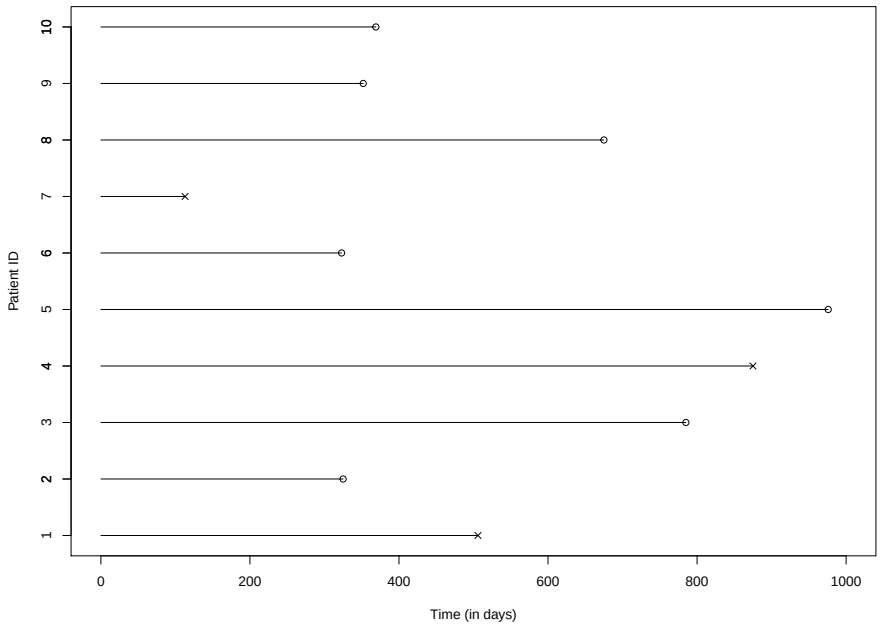

Figure 1.1: Survival times (×) and censoring times (o) in days.

1.2.1 Censoring

Censoring is one of the missing data problems in which time to event may not be fully observed for some reasons such as termination of study. Censoring is common in survival analysis which makes it special compared with ordinary linear regression or other standard statistical methods. Right-censoring is the most common type of censoring, occurring when then survival time is "incomplete" at the right side of the observation period. Event and censoring times of 10 individuals are illustrated in Figure 1.1.

In the most standard analysis and methods, censoring is non-informative which means that the time of censoring should be independent of the event time, given any observed covariates. A similar but different concept is truncation, which may occur in certain follow-up studies.

We extend survival time to right-censored case. Let T and C denote the time elapsed from an initiating event to an event and the end of the study, respectively. C is also called censoring time. Let $T^* = \min(T, C)$ be the follow up time and $\delta = I(\{T \leq C\})$ be the status, a 0/1 indicator which is 1 if survival time is observed and 0 if the observation is censored. What we actually

observe is the pair (T_i^*, δ_i) for individual i. Under the condition that the censoring is independent (noninformative), the likelihood function is

$$L = \prod_{j \text{ had event}} f(T_j^*) \prod_{k \text{ censored}} S(T_k^*) = \prod_{i=1}^{N} \lambda(T_i^*)^{\delta_i} S(T_i^*). \tag{1.1}$$

The counting process formulation replaces the pair of variables (T_i^*, δ_i) with the pair of functions $(N_i(t), Y_i(t))$, where

$$N_i(t) = \text{the number of observed events in } [0, t] \text{ for unit } i,$$

$$Y_i(t) = \begin{cases} 1 & \text{unit } i \text{ is under observation and at risk at time } t, \\ 0 & \text{otherwise.} \end{cases}$$

This formulation can be generalized to recurrent and multiple events, which will be introduced in later sections.

1.2.2 Kaplan-Meier Estimator

The Kaplan-Meier estimator is a common nonparametric method to estimate the survival function $S(t)$ from a sample of censored survival data. Assume we have n individuals and let $t_1 < t_2 < \cdots < t_k$ be the observed event times. Let d_j be the number of individuals who have an event at time t_j and $Y_i(t)$ is at risk indicator function for individual i. Then the Kaplan-Meier estimator is a nonparametric estimator of the survival function

$$\hat{S}(t) = \prod_{t_j < t} (1 - \frac{d_j}{\sum_{i=1}^{n} Y_i(t_j)}).$$

The standard errors can be estimated by using Greenwood's formula [15]

$$\hat{var}\left(\hat{S}(t)\right) = \left(\hat{S}(t)\right)^2 \prod_{t_j < t} \frac{d_j}{\sum_{i=1}^{n} Y_i(t_j)(\sum_{i=1}^{n} Y_i(t_j) - d_j)}.$$

1.2.3 Cox Proportion Hazards Regression Model

One objective of the survival analysis is to examine whether some features can affect survival times. The Cox proportional hazards model is one of the most common regression models in survival analysis, with hazard function [5]:

$$\lambda(t;x) = \lambda_0(t)e^{\beta x},$$

where $\lambda_0(t)$ is the baseline hazard, x is a covariate and β is a parameter to be estimated, representing the relative effect of the covariate on the hazard function. The conditional survival function can be estimated as

$$S(t;x) = [S_0(t)]^{e^{\beta x}},$$

where $S_0(t) = e^{-\int_0^t \lambda_0(u)du}$.

Suppose that we have n right-censored survival data $\{(T_i^*, \delta_i, X_i), i = 1, \cdots, n\}$. Recall the likelihood function (1.1), we can rewrite it as

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{n} \lambda_i(T_i^*)^{\delta_i} S(T_i^*) \\ &= \prod_{i=1}^{n} \left(\lambda_0(T_i^*)e^{\beta X_i}\right)^{\delta_i} e^{-\int_0^{T_i^*} e^{\beta X_i} d\Lambda(t)}. \end{aligned} \tag{1.2}$$

It is easy to show that the maximizer of this likelihood is the same as the maximizer of partial likelihood:

$$PL(\beta) = \prod_{t_j} \frac{Y_j(t_j)\lambda_0(t_j)e^{\beta x_j}}{\sum_{i=1}^{n} Y_i(t_j)\lambda_0(t_j)e^{\beta x_i}}.$$

1.2.4 Extension

In the above discussion, we focus on the time to the occurrence of a single event for each individual. However, there may be more complicated situations in real world problems. Sometimes the event may occur more than once for an individual, e.g. a carcinogenicity experiment on the times to the

development of mammary tumors for rats [11]. Besides, we may be interested in more than one type of events. For example, we may consider different causes of death as different types of events [34]. We will discuss these extension under the counting process formulation.

1.3 Counting Process and Martingales

In the last several decades, the theoretical foundation for survival and event history analysis has been connected to the study of counting process and martingale theory [14, 10].

1.3.1 Martingale Basics

Doob-Meyer Decomposition Theorem is one of the most important theorems of martingale application in counting process. The following version is from Theorem 2.2.3 in [10].

Theorem 1.1 (Doob-Meyer Decomposition) *Let $N = \{N(t); t \geq 0\}$ be a nonnegative right-continuous $\mathcal{F}_t$-local submartingale with localizing sequence $\{\tau_n\}$, where $\{\mathcal{F}_t; t \geq 0\}$ is a right-continuous filtration. Then there exists a unique increasing right-continuous predictable process A such that $A(0) = 0$ a.s., $Pr(A(t) < \infty) = 1$ for all $t > 0$ and $N - A$ is a right-continuous local martingale. At each t, $A(t)$ may be taken as the $a.s.\lim_{n\to\infty} A_n(t)$, where A_n is the compensator of the stopped submartingale $N(\cdot \wedge \tau_n)$.*

Theorem 1.1 implies that any local submartingale N can be decomposed as a local martingale and a predictable increasing process.

In survival analysis, the continuous compensator is

$$A(t) = \int_0^t Y(s)\lambda(s)ds.$$

Actually, the continuous compensator is frequently used in applications based on the following theorem, from Theorem 2.5.1 in [10].

Theorem 1.2 *Let N be a counting process on $[0, \infty)$ with $\mathrm{E}N(t) < \infty$ for all $t \geq 0$, and let A denote its compensator. Assume that almost all paths of A are continuous, and that $\mathrm{E}M^2(t) < \infty$ for all t, where $M = N - A$. Then the predictable variation process $< M, M >= A$, that is $M^2 - A$ is a right-continuous martingale.*

1.3.2 Counting Process

A counting process is a stochastic process $\{N(t); t \geq 0\}$ with $N(0) = 0$ and $N(t) < \infty$ a.s., and whose paths are with probability one right-continuous, piecewise constant, and have only jump discontinuities, with jumps of size +1. Suppose it is adapted to a $\sigma-$filtration $\{\mathcal{F}_t; t \geq 0\}$. The intensity $\lambda(t|\mathcal{F}_t)$ of a counting process is the rate of change in its predictable part, which is defined as

$$\lambda(t|\mathcal{F}_t) = \lim_{dt \to 0} \frac{1}{dt} \mathrm{E}(N(t + dt) - N(t)|\mathcal{F}_t).$$

From the Doob-Meyer decomposition, there exists a unique predictable process A such that $N - A$ is a martingale. Therefore, we have

$$\mathrm{E}(N(t + dt) - N(t)|\mathcal{F}_t) = \mathrm{E}(A(t + dt) - A(t)|\mathcal{F}_t) = dA(t),$$

since A is predictable. From this equation, we can see that $dA(t)$ is the rate of change for N conditional on the filtration. If we assume A is absolute continuous with

$$A(t) = \int_0^t l(s)ds,$$

for some random function l, we can model $l(s)$ instead.

In survival analysis, with survival time T, $N(t) = 1$ if $t \geq T$ and $N(t) = 0$, otherwise. It can be shown that the process given by $M(t) := N(t) - \int_0^t Y(u)\lambda(u)du$ is a martingale with respect to

$$\mathcal{F}_t = \sigma(N(s), Y(s); 0 \leq s \leq t),$$

provided independence censoring assumption.

By the uniqueness of Doob-Meyer decomposition, we can conclude that $l(s)$ is like the intensity $Y(s)\lambda(s)$. Therefore, we also call process A and l as cumulative intensity function and intensity function for counting process N respectively. In particular, the proportional intensity model gives that

$$A(t) = \int_0^t Y(s)\lambda_0(s)e^{\beta^T X(s)}ds.$$

This provide an interpretation for the approach that models the compensator directly.

1.3.3 Poisson Process

The Poisson process is a simple and widely used counting process model for counting data over time. Let $N(t)$ be a Poisson Process adopted by filtration $\mathcal{F}_t$ and cumulative intensity function $A(t)$. The $N(t)$ satisfies the following two properties:

a $N(t) - N(s) \sim$ Poisson $(A(t) - A(s))$ for $0 \leq s < t$.

b $N(t_1) - N(s_1)$ and $N(t_2) - N(s_2)$ are independent if $(s_1, t_1] \cap (s_2, t_2] = \emptyset$

If $A(t)$ is absolute continuous, there exists some random function $l(s)$ such that $A(t) = \int_0^t l(s)ds$. When $l(t)$ is independent of time, the process is called homogeneous. Otherwise, it's called inhomogeneous. Actually, for an inhomogeneous Poisson process $N(t)$ with $A(t)$, we can define a new process $\{N^*(s), s \geq 0\}$ by $N^*(s) = N(A^{-1}(s))$. Then the new defined $\{N^*(s), s \geq 0\}$ is a homogeneous Poisson process with $l(s) = 1$.

1.3.4 Recurrent Events

Survival analysis often deals with a single event during observation period in the most cases, while there are still more complicated situations in real world problems. Sometimes the event may occur more than once for an individual, e.g. a carcinogenicity experiment on the times to the development of mammary tumors for rats [11]. We call such case as recurrent events. More details and examples are in [4].

Instead of only one random variable T, we consider more than one event time within an individual. Let n be the total number of events detected in an individual and $\{t_j : j = 1, \cdots, n\}$ be a collection of times at which events occur over the time interval $[a, b]$. We can model such data by counting process $N_i(t)$, the number of events occurring in $[0, t]$ for individual i.

To reach the likelihood of recurrent events, we consider a partition $a = u_0 < u_1 < \cdots < u_M = b$ of the interval $[a, b]$. Each subinterval $[u_i, u_{i+1})$ only contains at most one event and $\{t_1, \cdots, t_n\} \subseteq \{u_0, \cdots, u_M\}$. Then

$$\begin{aligned}
&P(N(u_1) = n(u_1), \cdots, N(u_M) = n(u_M)|\mathcal{F}_a) \\
&= \prod_{i=0}^{M} P(N(u_i) = n(u_i)|\mathcal{F}_{u_i}) \\
&= \prod_{i=0}^{M} \left[P(\Delta N(u_i) = 1|\mathcal{F}_{u_i})^{\Delta N(u_i)} P(\Delta N(u_i) = 0|\mathcal{F}_{u_i})^{1-\Delta N(u_i)} \right] \\
&= \prod_{i=0}^{M} P(\Delta N(u_i) = 1|\mathcal{F}_{t_i}) \prod_{i:u_i \notin \{t_1, \cdots, t_n\}} P(\Delta N(u_i) = 0|\mathcal{F}_{u_i})^{1-\Delta N(u_i)},
\end{aligned} \tag{1.3}$$

where $\Delta N(u_i) = N(u_{i+1}) - N(u_i)$ is the number of events in $[u_i, u_{i+1})$. Recall the assumption that the events cannot occur simultaneously, we have

$$\begin{aligned}
P(\Delta N(u_i) = 1|\mathcal{F}_{u_i}) &= \lambda(u_i|\mathcal{F}_{u_i})\Delta u_i + o(\Delta u_i), \\
P(\Delta N(u_i) = 0|\mathcal{F}_{u_i}) &= 1 - \lambda(u_i|\mathcal{F}_{u_i})\Delta u_i + o(\Delta u_i).
\end{aligned}$$

Then equation (1.3) can be

$$\prod_{i:u_i \in \{t_1, \cdots, t_n\}} \left[\lambda(u_i|\mathcal{F}_{u_i})\Delta u_i + o(\Delta u_i)\right] \prod_{i:u_i \notin \{t_1, \cdots, t_n\}} \left[1 - \lambda(u_i|\mathcal{F}_{u_i})\Delta u_i + o(\Delta u_i)\right].$$

Divide the above expression by $\prod_{i:u_i \in \{t_1, \cdots, t_n\}} \Delta u_i$, we have

$$\prod_{i=1}^{n} \lambda(u_i|\mathcal{F}_{u_i}) \prod_{i:u_i \notin \{t_1, \cdots, t_n\}} \left[1 - \lambda(u_i|\mathcal{F}_{u_i})\Delta u_i + o(\Delta u_i)\right] [1 + o(1)].$$

Let $M \to \infty$ such that $\max_i \Delta u_i \to 0$, we have the likelihood

$$\prod_{j=1}^{n} \lambda(t_j|\mathcal{F}_{t_j}) \times \exp\left\{-\int_a^b \lambda(s|\mathcal{F}_s)ds\right\}. \tag{1.4}$$

See also [1] for more general and rigorous derivations of the likelihood.

1.3.5 Multitype Events

There are sometimes more than one event types in real world problems. Each event type corresponds to a counting process which jointly a multivariate counting process. Formally, a $J-$variate process $\{N_1, \cdots, N_J\}$ is called a multivariate counting process if each N_j, $j = 1, \cdots, J$ is a counting process adapted to the common filtration $\{\mathcal{F}_t\}$ and no two component process jump at the same time. Like Theorem 1.2, we also have a multivariate form of continuous compensator theorem, as Theorem 2.5.2 in [10].

Theorem 1.3 *Let $\{N_1, \cdots, N_J\}$ be a multivariate counting process and for $j = 1, \cdots, J$, let A_j be the compensator of N_j. Assume that each A_j is a continuous process. Then*

- $< M_j, M_j >= A_j$*, that is A_j is the unique predictable, right-continuous, increasing process with $A_j(0) = 0$ a.s and $A_j(t) < \infty$ a.s. for any t, such that $M_j^2 - A_j$ is a local martingale.*
- *If $i \neq j$, $< M_i, M_j > (t) = 0$ a.s., that is , $M_i M_j$ is a local martingale.*

We consider the likelihood for such multitype recurrent events of a $J-$variate process $\boldsymbol{N}(t) = \{N_1, \cdots, N_J\}$. Let $\mathcal{F}_t$ denote the filtration at time t, which should be commonly adapted to all components of the counting process. The intensity for each component is defined similar as before:

$$\lambda_j(t|\mathcal{F}_t) = \lim_{dt \to 0} \frac{1}{dt} \mathrm{E}(N_j(t + dt) - N_j(t)|\mathcal{F}_t).$$

Like recurrent events, we assume that at most one event can occur at time t, with

$$
\begin{aligned}
P(\Delta N_j(t) = 1|\mathcal{F}_t) &= \lambda_j(t|\mathcal{F}_t)\Delta t + o(\Delta t), \\
P(\Delta \boldsymbol{N}(t) = \mathbf{0}|\mathcal{F}_t) &= 1 - \sum_{j=1}^{J} \lambda_j(t|\mathcal{F}_t)\Delta t + o(\Delta t), \\
P\left(\sum_{j=1}^{J} \Delta N_j(t) \geq 2|\mathcal{F}_t\right) &= o(\Delta t).
\end{aligned}
$$

Let $t_{jk}, k = 1, \cdots, n_j$ denote the event times of type j over $[a, b]$ for $j = 1, \cdots, J$. Similarly, we consider a partition $a = u_0 < u_1 < \cdots < u_M = b$ of $[a, b]$, where each interval $[u_i, u_{i+1})$ contains at most one event time in $T = \{t_{jk}, k = 1, \cdots, n_j, j = 1, \cdots, J\}$ and $T \subseteq \{u_0, \cdots, u_M\}$.

Then

$$
\begin{aligned}
& P(\boldsymbol{N}(u_1) = \boldsymbol{n}(u_1), \cdots, \boldsymbol{N}(u_M) = \boldsymbol{n}(u_M)|\mathcal{F}_a) \\
&= \prod_{i=0}^{M} P(\boldsymbol{N}(u_i) = \boldsymbol{n}(u_i)|\mathcal{F}_{u_i}) \\
&= \prod_{i=0}^{M} \prod_{j=1}^{J} \left[P(\Delta N_j(u_i) = 1|\mathcal{F}_{u_i})^{\Delta N_j(u_i)} P(\Delta N_j(u_i) = 0|\mathcal{F}_{u_i})^{1-\Delta N_j(u_i)} \right] \\
&= \prod_{j=1}^{J} \prod_{k=1}^{n_j} P(\Delta N_j(t_{jk}) = 1|\mathcal{F}_{t_{jk}}) \prod_{j=1}^{J} \prod_{i:u_i \notin T} P(\Delta N_j(u_i) = 0|\mathcal{F}_{u_i}) \\
&= \prod_{j=1}^{J} \prod_{k=1}^{n_j} \left[\lambda_j(t_{jk}|\mathcal{F}_{t_{jk}}) + o(\Delta t_{jk}) \right] \prod_{i:u_i \notin T} \left[1 - \sum_{j=1}^{J} \lambda_j(u_i|\mathcal{F}_{u_i})\Delta u_i + o(\Delta u_i) \right].
\end{aligned}
\tag{1.5}
$$

Dividing the last expression by $\prod_{j=1}^{J} \prod_{k=1}^{n_j} \Delta t_{jk}$ and taking the limit as $M \to \infty$ such that $\max_i \Delta u_i \to 0$, we obtain

$$
L = \prod_{j=1}^{J} \prod_{k=1}^{n_j} \lambda_j(t_{jk}|\mathcal{F}_{t_{jk}}) \exp\left(- \sum_{j=1}^{J} \int_a^b \lambda_j(u|\mathcal{F}_u) du \right).
$$

We can write the likelihood in independent form:

$$L = \prod_{j=1}^{J} \left[\prod_{k=1}^{n_j} \lambda_j(t_{jk}|\mathcal{F}_{t_{jk}}) \exp\left(- \int_a^b \lambda_j(u|\mathcal{F}_u) du \right) \right].$$

We can find that each intensity function λ_j is functionally independent. If the intensities do not share the same parameters, the estimation can be performed separately by maximizing likelihood.

1.4 Regression Models

1.4.1 Covariates

In parametric and semiparametric model, we are interested in estimating if some features can affect survival times or rate of change in event occurrence. In this way, we often consider incorporating covariates in regression models, e.g. cox proportional hazards model. The covariates include fixed covariates and time-varying covariates.

- Fixed covariates are independent of time, e.g. age in a short time, gender and ethnic.
- Time-varying covariates can be external and internal.
  - External covariates is independent of the recurrent event process of interest. Examples are individuals' external measurements at different time or age in a long time period.
  - Internal covariates often depend on the past history. Examples are indicators whether one experienced the event before time t or the total number of events one experienced.

More details and examples are in Chapter 6 in [19].

1.4.2 Self Controlled Case Series Model

The self-controlled case series (SCCS) method is based on Poisson regression to measure the causal association between adverse events and drug exposures in postmarketing surveillance. Compared with other regression methods, the SCCS method has two main features: (1) it only uses individuals

with at least one event for analysis and, (2) each individual serves as his/her own control, thus requiring each individual to have different exposure patterns [7, 8, 30]. Let i denote the individual index, d denote the day index and x_{id} be the covariates for individual i at day d. Then, the number of events y_{id} follows

$$y_{id}|x_{id}, \phi_i \sim Poisson(e^{\phi_i + \beta^T x_{id}}).$$

To allow individual-level heterogeneity, individual i is assumed to have an individual baseline intensity e^{ϕ_i}. The SCCS method requires two assumptions:

1. Events at different days are conditionally independent given covariates and ϕ_i.

2. Events are conditionally independent with future covariates given the current covariates.

With the above assumptions, the likelihood in individual i is

$$L_i(\boldsymbol{y}_i|\boldsymbol{x}_i, \phi_i) = P(y_{i1}, \cdots, y_{i\tau_i}|x_{i1}, \cdots, x_{i\tau_i}, \phi_i) = \prod_{d=1}^{\tau_i} P(y_{id}|x_{id}, \phi_i).$$

Actually, we are interested in β since it determines the relationship between events and covariates, whereas the individual parameter ϕ_i gives no information and can be treated as a nuisance parameter. With the Fisher-Nyeman Factorization Theorem, we can show that the observed number of events $n_i = \sum_{d=1}^{\tau_i} y_{id}$ is sufficient for the nuisance parameter ϕ_i. This implies that we can use the conditional likelihood in which ϕ_i is removed from the likelihood expression:

$$L_i^c(\boldsymbol{y}_i|\boldsymbol{x}_i, n_i) = \prod_{d=1}^{\tau_i} \left(\frac{e^{\beta^T x_{id}}}{\sum_{d'} e^{\beta^T x_{id}}} \right)^{y_{id}}.$$

Estimation can be obtained by maximizing the conditional log-likelihood. Actually, the assumptions above are likely to be violated in real world problems. For example, if an individual experienced a myocardial infarction, he/she may be at higher risk to encounter another myocardial infarction in the next few days. A generalized method, the positive dependence self-controlled case series (PD-SCCS) method [29], is proposed by adding a positive term to the baseline risk.

The PD-SCCS allows the previous event occurrences to increase the future event risk additively based on a continuous time SCCS model with the following definition of intensity:

$$\lambda_i(t|\phi_i, \mathcal{F}(t)) = (e^{\phi_i} + \delta N_i(t-))e^{\beta^T x_i(t)}.$$

Such a model specification may only work in very special situations and is not easy to be extended to more general cases.

1.4.3 Frailty Model and Random Effect Model

In the Cox proportional hazards model, we specify the intensity of $N(t)$ with filtration $\mathcal{F}_t$:

$$\lambda(t|\mathcal{F}_t) = \lambda_0(t)e^{\beta^T X(t)},$$

where $\lambda_0(t)$ is an unspecified baseline intensity function. In this model setting, the occurrence of an event is independent of any earlier events unless $X(t)$ includes the past history. Besides, the data from real world problems are always high-dimensional, irregular and heterogeneous. The simple proportional intensity model is not enough in some cases. The random effect model is used in dealing with the complexity of the data [36]:

$$\lambda(t|\xi, \mathcal{F}_t) = \xi\lambda_0(t)e^{\beta^T X(t)},$$

where ξ is a random effect following some distributions. Several asymptotic theories have been established when ξ follows a gamma distribution [22, 21]. Furthermore, a broad class of intensity models with random effects that may accommodate nonproportional intensity is proposed [36]

$$\Lambda(t|X, Z; b) = G\left(\int_0^t \lambda_0(t)e^{\beta^T X(s)+b^T Z(s)}ds\right),$$

where β is the unknown regression parameters, b is the random effects with density function $\psi(b;\gamma)$, X and Z are covariate processes associated with fixed and random effects, and $G(\cdot)$

is a three times continuously differentiable and strictly increasing transformation function with $G(0) = 0$ and $G(\infty) = \infty$.

Chapter 2: A Multivariate Correlated Frailty Model for Postmarketing Surveillance

2.1 Introduction

A primary purpose of the postmarketing surveillance is to monitor the safety of a pharmaceutical drug or medical device after it has been released in the market. That is to say, we are interested in identifying and estimating the association between adverse events (AEs) and drug exposures along the time in Longitudinal Observational Databases (LODs), which contain time-to-event medical information such as dates of diagnoses and prescription and length of drug exposure periods.

Some methods in postmarketing surveillance have been developed recently, e.g. self-controlled case series (SCCS) method. The SCCS method [30, 29] is based on Poisson regression to measure the causal association between adverse events and drug exposures in postmarketing surveillance. Compared with other regression methods, the SCCS method has two main features: (1) it only uses individuals with at least one event for analysis and, (2) each individual serves as his/her own control, thus requiring each individual to have different exposure patterns.

Although the SCCS related methods have achieved empirical success in postmarketing surveillance with the Observational Medical Outcomes Partnership (OMOP) project, these methods still have limitation due to model design:

- Low efficiency. The SCCS analysis only includes patients with at least one adverse event and the information from the controls is not included. In addition, each case individual needs to have different exposure patterns.

- Limited dependence structure. The SCCS methods require conditional independence assumption between multiple event occurrences. However, such an assumption is likely to

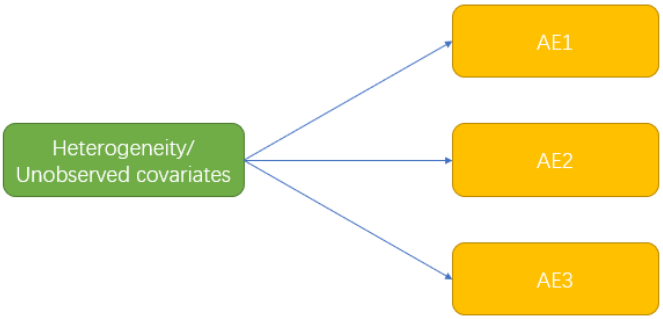

Figure 2.1: Hidden Confounding Factors

be violated because the previous event occurrence can change the risk of future event. An extension of the SCCS model is given by adding a positive term to the baseline event risk [29]. This design of model may only work in very special situations and is not easy to be extended to generalized cases.

- Failure in multivariate events. In the SCCS model setting, mathematical structure fails in multivariate events unless each adverse event is treated as an independent model. Actually, this is impractical in real world problems.

In this chapter, we discuss a multivariate random effect model based on event history analysis. A central component of the proposed model is its flexible representation form to describe the casual association between multiple events and correlation structure of within-subject heterogeneity. Event history analysis consists of statistical methods for the analysis of longitudinal data on the occurrence and timing of events. Furthermore, the event occurrence may be high-dimensional, irregular, and heterogeneous. Simple event history methods may not be enough to describe the nature of event occurrences due to complicated hidden confounding factors or unobserved covariates as in Figure 2.1.

To account for the heterogeneity, the unobserved covariates are included as random effect which allows the effects of unobserved covariates on multiple events to keep the possible correlation structure. The frailty model is a random effect model for time variables, where the random

effect (the frailty) has a multiplicative effect on the intensity:

$$\lambda_i(t|\theta_i, \mathcal{F}_t) = \theta_i e^{\beta^T X_i},$$

whereas the frailty θ_i follows distributional assumptions due to mathematical convenience. In multivariate cases, the frailty can be more complex and high dimensional. We assume K types of event and N individuals. For intensity at individual i and event k,

$$\lambda_{ik}(t|\theta_{ik}, \mathcal{F}_t) = \theta_{ik} e^{\beta_k^T X_{ik}},$$

and we assume the frailty vector within the same individual $\boldsymbol{\phi}_i = [\phi_{i1}, \cdots, \phi_{ik}]$ follows some multivariate distribution Ψ.

In univariate cases, gamma distribution is commonly used due to its compatibility with the Poisson counting process. Actually, when the variance of gamma frailty goes to infinity, the effect estimators of the frailty model will converge to it in the SCCS method. We leave the proof in section 2.6 . In multivariate cases, things become more complicated. When Ψ is degenerated to one dimension [16], we call it shared frailty model which uses one latent variable as the individual-level heterogeneity for multiple event types. The correlated PVF frailty model [35] was proposed in gamma structure which is a bivariate extension of the univariate PVF frailty model [17]. Some works [20] discussed similar models but used a compound Poisson distribution for correlation structure.

These gamma frailty models provide good views of bivariate correlation structure in event history analysis. However, two problems prevent it to be applied in general cases. The first problem is that bivariate gamma frailty only provides non-negative correlation due to its structure. Another problem is that the complexity of model increase substantially when the number of event types increases. In this chapter, we will discuss a multivariate Gaussian frailty model, which can be easily extended to high dimension and complicated cases.

Here is the outline of this chapter. In section 2.2, we construct the multivariate correlated

frailty model. We discuss the estimation methods for such model in section 2.3. In section 2.4, we compare the multivariate correlated frailty model with independent frailty and shared frailty model through simulation studies. In section 2.5, we introduce the application of our proposed model in a EHR dataset with events of interest defined by Observational Medical Outcomes Partnership (OMOP) project.

2.2 Model Setting

We introduce a multivariate correlated frailty model with multiple drug exposures and multiple adverse events. We consider our model with possibly right-censored data. Let N, J, L denote the number of individuals, number of adverse events and number of drug exposures respectively. The data consists of $\{\boldsymbol{N}_i(t), \boldsymbol{Y}_i(t), \boldsymbol{X}_i(t) : t \in [0, \tau_i], i = 1, 2, \cdots, N\}$. Let τ_i denote the length of the observation period. $\boldsymbol{N}_i(t) = (N_{i1}(t), N_{i2}(t), \cdots, N_{iJ}(t))^T$ is the occurring number of adverse events between starting time and time t and $\boldsymbol{Y}_i(t) = (Y_{i1}(t), Y_{i2}(t), \cdots, Y_{iJ}(t))^T$ is at risk status defined by $Y_{ij}(t) = \mathcal{I}(C_{ij} \geq t)$, where C_{ij} is the right-censoring time. Then we have intensity for counting process $N_{ij}(t)$

$$\lambda_{ij}(t|\phi_{ij}, \mathcal{F}_t) = Y_{ij}(t)\lambda_{0j}(t)e^{\phi_{ij}+\boldsymbol{\beta}_j^T \boldsymbol{X}_{ij}(t)}.$$

The frailty vector $\boldsymbol{\phi}_i$ follows a J dimensional Gaussian distribution

$$\boldsymbol{\phi}_i = \begin{pmatrix} \phi_{i1} \\ \phi_{i2} \\ \vdots \\ \phi_{iJ} \end{pmatrix} \sim \mathcal{N}\left(-\frac{1}{2}\text{diag}(\boldsymbol{\Sigma}), \boldsymbol{\Sigma}\right),$$

where $\boldsymbol{\Sigma}$ is a $J \times J$ positive definite matrix. The mean of this multivariate Gaussian distribution is set to guarantee the mean of frailty $e^{\phi_{ij}}$ to be 1.

The parameters $\boldsymbol{\beta}_j$ is a coefficient vector for $j-$th adverse event and $\boldsymbol{X}_{ij}(t)$ is a covariate

vector. If we focus on estimating the strength of the association between adverse events and drug exposures, we can replace $\boldsymbol{X}_{ij}(t)$ with $\boldsymbol{X}_i(t)$ since drug exposure information is independent of adverse event index j. In this way, the drug index is given by $l = 1, 2, \cdots, L$ and the observed drug exposure covariate is given by $\boldsymbol{X}_i(t) = (X_{i1}(t), X_{i2}(t), \cdots, X_{iL}(t))^T$, where $X_{il}(t) = 1$ if the individual i is exposed to the drug l at time t, and 0 otherwise.

The above model includes only drug exposures as covariates which ignores that the occurrence of an event can affect the risk of future events. For example, if an individual experienced a myocardial infarction, he may be more likely to experience another myocardial infarction in the next few days. Such phenomenon can even exist between two different events. For example, cardiovascular disease and chronic kidney disease are closely interrelated since disease of one organ can cause dysfunction of the other, ultimately leading to the failure of the both organs. Actually, these problems are common in the application of postmarketing surveillance which inspires us to incorporate the covariates with the history of adverse events, such as cumulative event numbers or indicators. In this chapter, for simplicity, we include similar indicator covariates for adverse events as drug exposures with intensity

$$\lambda_{ij}(t|\phi_{ij}, \mathcal{F}_t) = Y_{ij}(t)\lambda_{j0}(t)e^{\phi_{ij}+\boldsymbol{\beta}_j^T\boldsymbol{X}_i(t)+\boldsymbol{\eta}_j^T\boldsymbol{Z}_i(t)},$$

where $\boldsymbol{Z}_i(t) = \{Z_{i1}(t), Z_{i2}(t), \cdots, Z_{iJ}(t)\}$ is a covariate vector of adverse events and $Z_{ij}(t) = 1$ if adverse event j occurs before time t and 0 otherwise.

Based on above notations and discussion, the likelihood function for parameters $\theta = \{\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}\}$ and $\mathcal{A} = \{\Lambda_{10}, \Lambda_{20}, \cdots, \Lambda_{j0}\}$ is

$$\begin{aligned} L(\theta, \mathcal{A}; D) = \prod_{i=1}^{N} \int_{\boldsymbol{\phi}_i} \prod_{j=1}^{J} \Bigg[\prod_{t\le\tau} \left\{ Y_{ij}(t)\lambda_{j0}(t)e^{\phi_{ij}+\boldsymbol{\beta}_j^T\boldsymbol{X}_i(t)+\boldsymbol{\eta}_j^T\boldsymbol{Z}_i(t)} \right\}^{dN_{ij}(t)} \\ \times \exp\left\{ -\int_0^\tau Y_{ij}(t)e^{\phi_{ij}+\boldsymbol{\beta}_j^T\boldsymbol{X}_i(t)+\boldsymbol{\eta}_j^T\boldsymbol{Z}_i(t)} d\Lambda_{j0}(t) \right\} \Bigg] \psi(\boldsymbol{\phi}_i; \boldsymbol{\Sigma}) d\boldsymbol{\phi}_i. \end{aligned} \tag{2.1}$$

If we consider $\Lambda_{j0}(t)$ as absolute continuous functions, we can not reach a maximum of the

likelihood. Thus, we restrict $\Lambda_{j0}(t)$ in step functions with jumps at the observed event times $\{t_{ijk}, i = 1, 2, \cdots, n; j = 1, 2, \cdots, J; k = 1, 2, \cdots, n_{ij}\}$ and the baseline intensity $\lambda_{j0}(t)$ becomes the jump size of $\Lambda_{j0}(t)$ at time t, denoted by $\Lambda_{j0}\{t\}$. The modified likelihood function becomes

$$\begin{aligned} L(\theta, \mathcal{A}; D) = \prod_{i=1}^{N} \int_{\boldsymbol{\phi}_i} \prod_{j=1}^{J} \Bigg[\prod_{t \le \tau} \left\{ Y_{ij}(t) \Lambda_{j0}\{t\} e^{\phi_{ij} + \boldsymbol{\beta}_j^T \boldsymbol{X}_i(t) + \boldsymbol{\eta}_j^T \boldsymbol{Z}_i(t)} \right\}^{dN_{ij}(t)} \\ \times \exp\left\{ - \int_0^{\tau} Y_{ij}(t) e^{\phi_{ij} + \boldsymbol{\beta}_j^T \boldsymbol{X}_i(t) + \boldsymbol{\eta}_j^T \boldsymbol{Z}_i(t)} d\Lambda_{j0}(t) \right\} \Bigg] \psi(\boldsymbol{\phi}_i; \boldsymbol{\Sigma}) d\boldsymbol{\phi}_i. \end{aligned} \tag{2.2}$$

The theory about consistency and asymptotically normality is discussed in the next chapter, where we establish a general asymptotic theory for nonparametric maximum likelihood estimation in event history analysis with dynamic random effect.

2.3 Estimation

In the gamma frailty model, parameters estimation is mostly based on maximizing the integrated likelihood by simple convex optimization algorithms, e.g. gradient descent method and newton's method. This is because the probability density function of gamma distribution and the likelihood of Poisson process are in the same exponential family and we can integrate the likelihood (2.2) in a closed form. However, the multivariate Gaussian distribution can not derive such closed form. The Expectation-Maximization algorithm was discussed [36] by implementing Gaussian-quadrature approximation for Gaussian frailty for conditional expectations numerical integration in E-step. In this thesis, we will discuss the Bayesian approach applied in full likelihood.

The Bayesian approach for this model requires determination of the full probability model, including the conditional survival likelihood, the distribution of the frailty and the prior distributions of the parameters. Let D_{obs} denote the data and $\mathcal{R} = \{\boldsymbol{\phi}_i, i = 1, 2, \cdots, N\}$ denote the frailty, the full likelihood for the data and model is

$$L_{\text{full}}(D_{\text{obs}}, \mathcal{R}, \boldsymbol{\beta}, \boldsymbol{\eta}, \mathcal{A}, \boldsymbol{\Sigma}) = L_c(D_{\text{obs}} | \boldsymbol{\beta}, \boldsymbol{\eta}, \mathcal{A}, \mathcal{R}) \left[\prod_{i=1}^{n} f(\boldsymbol{\phi}_i | \boldsymbol{\Sigma}) \right] \pi(\mathcal{A}) \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\eta}) \pi(\boldsymbol{\Sigma}). \tag{2.3}$$

The Markov Chain Monte Carlo (MCMC) method allows us to obtain a series of samples from the targeted posterior distribution by a Markov chain. In some papers [12], adequate burn-in periods are necessary to approximately sample from a known posterior distribution. However, the full posterior likelihood is intractable and thus we implement Gibbs sampling algorithm which requires condition distributions of each parameter given all other parameters.

For frailty covariance parameter $\mathbf{\Sigma}$ which is positive definite, we re-parametrize the covariance matrix by $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$ to use the conjugate prior. For the prior, we use a Wishart prior which is defined over positive definite matrix with hyperparameters of a real number a and a positive definite matrix $\boldsymbol{V}_{J\times J}$:

$$\begin{aligned} \mathbf{\Omega} \sim &\quad \mathcal{W}(\mathbf{\Omega}; a, \boldsymbol{V}), \\ f(\mathbf{\Omega}) = &\quad \frac{|\mathbf{\Omega}|^{(a-J-1)/2} e^{-tr(\boldsymbol{V}^{-1}\mathbf{\Omega})/2}}{2^{\frac{Ja}{2}} |\boldsymbol{V}|^{a/2} \Gamma_J(\frac{a}{2})}. \end{aligned} \tag{2.4}$$

Usually, we set $a = J$ and $\boldsymbol{V} = n^{-1}\mathbf{\Sigma}_0$ where $\mathbf{\Sigma}_0$ is some prior guess for the covariance matrix.

There are many possible choices for baseline parameters $\mathcal{A}$, e.g. piecewise linear function or constant function. In our database, the observation period is short and there is no obvious starting point. Thus, we assume the baseline keeps constant during the observation period for simplicity. We use α_j to represent the baseline intensity for event type j.

For baseline parameters α_j and relative risk parameters $\boldsymbol{\beta}_j$ and $\boldsymbol{\eta}_j$, we use noninformative priors on them with an independent univariate Normal prior with zero mean and large variance $\tau_\alpha, \tau_\beta, \tau_\gamma$, e.g. 1000:

$$\alpha_j \sim \mathcal{N}(0, \tau_\alpha),$$

$$\beta_{jl} \sim \mathcal{N}(0, \tau_\beta),$$

$$\eta_{jl} \sim \mathcal{N}(0, \tau_\gamma).$$

With above settings, the full likelihood becomes

$$
\begin{aligned}
&L_{\text{full}}(D_{\text{obs}}, \mathcal{R}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}) \\
&\propto \prod_{i=1}^{N}\prod_{j=1}^{J}\left[\prod_{k=1}^{n_{ij}} Y_{ij}(t_{ijk}) e^{\alpha_j+\phi_{ij}+\boldsymbol{\beta}_j^T \boldsymbol{X}_i(t_{ijk})+\boldsymbol{\eta}_j^T \boldsymbol{Z}_i(t_{ijk})} \times \exp\left\{-\int_0^{\tau} Y_{ij}(t) e^{\alpha_j+\phi_{ij}+\boldsymbol{\beta}_j^T \boldsymbol{X}_i(t)+\boldsymbol{\eta}_j^T \boldsymbol{Z}_i(t)} dt\right\}\right] \\
&\quad \times \prod_{i=1}^{n} det(\Omega)^{1/2} \exp\left\{-\frac{1}{2}\boldsymbol{\phi}_i^T \Omega \boldsymbol{\phi}_i\right\} \\
&\quad \times \prod_{j=1}^{J} \exp\left\{-\frac{\alpha_j^2}{2\tau_\alpha^2}\right\} \prod_{j=1}^{J}\prod_{l=1}^{L} \exp\left\{-\frac{\beta_{jl}^2}{2\tau_\beta^2}\right\} \prod_{j=1}^{J}\prod_{j'=1}^{J} \exp\left\{-\frac{\eta_{jj'}^2}{2\tau_\gamma^2}\right\} \\
&\quad \times |\boldsymbol{\Omega}|^{(a-J-1)/2} e^{-tr(\boldsymbol{V}^{-1}\boldsymbol{\Omega})/2}.
\end{aligned}
\tag{2.5}
$$

2.4 Simulation Comparison Among Different Frailty Models

The goal of this section is to compare the performance of a variety of frailty models through simulation studies. We simulate the data from shared frailty, independent frailty and correlated frailty model, respectively. For each simulated data, we fit it with all three models and compare their performance under different parameter settings.

2.4.1 True Model Being Correlated Frailty

We simulate data from correlated frailty model first. We simulate $N = 10,000$ individuals and each of them has an observation period $\tau = 1,000$. We run such simulation $1,000$ times with the same model settings to generate $1,000$ estimators. To make the results clear for comparison, we choose $J = 2$ adverse events and $L = 1$ drugs exposure for simplicity. We also incorporate the adverse event indicators as covariates. Then the correlated frailty model becomes

$$
\begin{aligned}
\log \lambda_{i1}(t|\phi_{i1}, \mathcal{F}_t) &= \alpha_1 + \phi_{i1} + \beta_1 X_i(t) + \eta_{11} Z_{i1}(t) + \eta_{12} Z_{i2}(t), \\
\log \lambda_{i2}(t|\phi_{i2}, \mathcal{F}_t) &= \alpha_2 + \phi_{i2} + \beta_2 X_i(t) + \eta_{21} Z_{i1}(t) + \eta_{22} Z_{i2}(t),
\end{aligned}
$$

where the frailty

$$\begin{pmatrix} \phi_{i1} \\ \phi_{i2} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} -\sigma_1^2/2 \\ -\sigma_2^2/2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right).$$

We generate drug exposures by exponential distribution with mean 0.001. For each drug exposure, we extend it with $d = 30$ days since we assume each prescription period is 30 days and individuals are under drug exposure during the prescription period. Then $X_i(t)$ is a time-varying indicator of drug exposure. The covariate for adverse event $Z_{ij}(t) = 1$ if adverse event j occurs before time t within 30 days and 0, otherwise.

For relative risk parameters, we choose the values of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ as

$$\boldsymbol{\beta} = \begin{pmatrix} 1.0 \\ 1.5 \end{pmatrix} \qquad \boldsymbol{\eta} = \begin{pmatrix} 1.0 & -0.7 \\ -0.5 & 1.2 \end{pmatrix}.$$

For other parameters $\boldsymbol{\alpha}$, ρ and σ_1, σ_2, we vary them in a range of values to compare the performance of three models in different parameter settings. The baseline parameters $e^{\alpha_1} = e^{\alpha_2}$ are set as 0.0001,0.0005 and 0.001. Based on the definition of Poisson process, these values correspond to the expected number of events as 0.1, 0.5 and 1.0, respectively. Considering the strength of correlation structure, we vary ρ among 0.1, 0.5 and 0.9, representing from relatively uncorrelated structure to highly positive correlated structure. For σ_1, σ_2, we fix σ_1 as 1.0 but vary the ratio σ_2/σ_1 among 1.0, 1.5 and 2.0.

We show the estimation results of β_1 and η_{12} in table 2.1 , 2.2 and 2.3 and the estimation results of ρ in table 2.4.

Table 2.1 shows mean square error (MSE) and median absolute deviation (MAD) for $\hat{\beta}_1$ when the true model is correlated frailty model. When ρ is small ($\rho = 0.1$), the correlated frailty model and independent model give similar MSE and MAD. When the value of ρ increases, the correlated frailty model performs better than the independent frailty model. The shared frailty model is always the worst in MSE and MAD and the error grows with increasing ρ and σ_2/σ_1.

η_{12} is the relative risk parameters of effect from adverse event 2 to adverse event 1, which may be biased if the frailty structure is ignored. Actually, an example in section 2.6 illustrates that when the covariate is correlated with outcome (or frailty), the independence assumption would lead estimation bias. Table 2.2 shows mean square error (MSE) and median absolute deviation (MAD) for $\hat{\eta}_{12}$ when the true model is correlated frailty model. The correlated frailty model has the best performance and the lowest errors among three models for all settings. The shared frailty model has worst performance and the estimation errors are much larger, which can be 100 times of the error in the correlated frailty model. The performance of the independent frailty model is better than it in the shared frailty model, but much worse than it in the correlated frailty model, except the case with the small correlation ρ and small σ_2/σ_1 ratio.

Table 2.3 shows the 95% confidence interval coverage for $\hat{\beta}_1$ and $\hat{\eta}_{12}$ when the true model is correlated frailty model. The 95% confidence interval coverage is the probability that true parameter value falls in the empirical 95% confidence intervals. When ρ is at a low level ($\rho = 0.1$) or $\sigma_2/\sigma_1 = 1$, the independent frailty model and shared frailty model have similar performance as the correlated frailty model. When the correlation ρ and σ_2/σ_1 ratio increase, the coverage for the independent frailty model and shared frailty model both decrease, while the later one decreases much faster. The shared frailty model has terrible performance in the case of large σ_2/σ_1 and large e^{α}. With respect to $\hat{\eta}_{12}$, the shared frailty model has zero 95% coverage in all most cases, except with large ρ and $\sigma_1 = \sigma_2$. The independent frailty model also has terrible performance. The coverage is above 68% when the correlation coefficient ρ is small. When ρ increases, the coverage under independent frailty model drops sharply.

Table 2.4 shows the estimation result summary for ρ when the true model is correlated frailty model. The Bias $\hat{\rho}$ column shows the bias of the ρ estimators over all iterations. The absolute values of the bias are all at a low level which are less than 7.5% of the corresponding true values and 20% of the standard errors. On average, the bias with small ρ is larger than it with large ρ. The 95% coverage levels are generally close to 95% and at least 90%. The last two columns gives the proportion of times that BIC favored the correlated frailty model over the other two models.

Actually, all the values are large than 99.2% and in most cases, the criterion can choose the true model.

ρ	$e^{\alpha}(\times 10^{-3})$	σ_2/σ_1	MSE $\hat{\beta}_1(\times 10^{-3})$			MAD $\hat{\beta}_1(\times 10^{-2})$		
			Correlated	Independent	Shared	Correlated	Independent	Shared
		1.0	2.943	2.959	2.945	3.220	3.156	3.288
	0.5	1.5	2.552	2.568	3.706	3.221	3.296	4.421
		2.0	3.239	3.265	5.861	4.141	4.149	5.946
		1.0	1.326	1.349	1.365	2.401	2.367	2.466
0.1	1.0	1.5	0.975	0.987	2.937	2.200	2.270	4.434
		2.0	0.940	0.957	3.990	2.082	1.954	5.445
		1.0	0.475	0.476	1.006	1.526	1.555	2.490
	2.0	1.5	0.536	0.544	2.995	1.541	1.434	5.043
		2.0	0.533	0.547	3.579	1.541	1.642	5.386
		1.0	2.774	2.954	2.684	3.629	3.632	3.688
	0.5	1.5	3.293	3.356	4.758	3.346	3.300	4.569
		2.0	2.997	3.122	6.662	3.302	3.364	6.519
		1.0	1.221	1.319	1.316	2.270	2.093	2.309
0.5	1.0	1.5	1.425	1.737	3.224	2.607	2.888	4.094
		2.0	0.998	1.277	5.051	2.087	2.307	6.752
		1.0	0.668	0.842	0.792	1.601	1.931	2.039
	2.0	1.5	0.469	0.654	3.538	1.485	1.745	5.560
		2.0	0.609	0.828	5.278	1.837	2.108	6.724
		1.0	3.026	3.343	2.988	3.780	3.962	3.789
	0.5	1.5	3.061	4.007	4.040	4.032	4.149	5.423
		2.0	3.118	4.970	5.605	4.163	5.278	5.231
		1.0	1.433	1.744	1.450	2.355	2.347	2.606
0.9	1.0	1.5	1.085	1.548	3.757	2.411	2.763	5.112
		2.0	1.965	3.224	7.687	2.932	4.066	7.574
		1.0	0.566	1.151	0.560	1.490	2.565	1.570
	2.0	1.5	0.599	1.458	4.000	1.842	3.042	5.771
		2.0	0.717	1.621	9.370	1.807	3.619	8.622

Table 2.1: MSE and MAD for $\hat{\beta}_1$ under correlated frailty, independent frailty and shared frailty when true mode is correlated frailty model

ρ	$e^{\alpha}(\times 10^{-3})$	σ_2/σ_1	MSE $\hat{\eta}_{12}(\times 10^{-2})$			MAD $\hat{\eta}_{12}(\times 10^{-1})$		
			Correlated	Independent	Shared	Correlated	Independent	Shared
		1.0	2.087	2.202	39.562	0.968	1.074	6.096
	0.5	1.5	2.366	2.280	211.603	1.008	0.957	14.324
		2.0	2.484	3.030	478.846	1.066	1.153	21.688
		1.0	0.549	0.610	39.032	0.418	0.580	6.037
0.1	1.0	1.5	0.443	0.796	181.168	0.504	0.621	13.335
		2.0	0.698	1.307	411.063	0.417	0.841	20.110
		1.0	0.124	0.184	36.694	0.315	0.283	6.045
	2.0	1.5	0.143	0.328	154.199	0.260	0.441	12.339
		2.0	0.183	0.488	348.795	0.325	0.542	18.515
		1.0	1.493	6.300	12.650	0.751	2.328	3.316
	0.5	1.5	1.229	13.228	91.775	0.805	3.403	9.350
		2.0	1.108	21.055	268.959	0.671	4.442	16.102
		1.0	0.384	3.213	11.074	0.411	1.694	3.373
0.5	1.0	1.5	0.287	6.502	82.736	0.370	2.563	9.170
		2.0	0.367	12.027	225.910	0.311	3.491	14.869
		1.0	0.091	1.274	12.243	0.194	1.088	3.505
	2.0	1.5	0.092	3.276	73.903	0.205	1.801	8.537
		2.0	0.112	4.782	189.450	0.246	2.122	13.588
		1.0	1.223	13.426	1.978	0.822	3.459	0.884
	0.5	1.5	0.740	33.468	25.728	0.622	5.758	4.958
		2.0	0.600	54.954	120.061	0.578	7.271	10.709
		1.0	0.230	6.632	0.697	0.304	2.575	0.672
0.9	1.0	1.5	0.203	15.802	24.520	0.289	3.955	4.866
		2.0	0.192	28.785	109.001	0.304	5.311	10.198
		1.0	0.064	2.890	0.704	0.174	1.683	0.769
	2.0	1.5	0.068	7.737	24.815	0.186	2.767	4.937
		2.0	0.059	13.113	106.722	0.182	3.554	10.029

Table 2.2: MSE and MAD for $\hat{\eta}_{12}$ under correlated frailty, independent frailty and shared frailty when true mode is correlated frailty model

ρ	$e^{\alpha}(\times 10^{-3})$	σ_2/σ_1	95% Coverage $\hat{\beta}_1$(%)			95% Coverage $\hat{\eta}_{12}$(%)		
			Correlated	Independent	Shared	Correlated	Independent	Shared
		1.0	93.5	93.5	91.5	94.5	94.7	0.0
	0.5	1.5	92.8	92.8	88.8	95.8	92.3	0.0
		2.0	91.4	91.4	79.4	93.4	84.1	0.0
		1.0	92.8	92.8	91.8	93.8	93.2	0.0
0.1	1.0	1.5	99.7	100.0	74.7	96.7	84.4	0.0
		2.0	98.0	98.0	65.0	93.0	80.0	0.0
		1.0	95.1	98.1	86.1	98.1	87.1	0.0
	2.0	1.5	93.9	93.9	38.9	95.9	75.3	0.0
		2.0	95.4	95.4	34.4	97.4	68.2	0.0
		1.0	94.4	94.4	96.4	94.4	50.0	15.4
	0.5	1.5	91.1	87.1	86.1	95.1	14.7	0.0
		2.0	90.6	85.6	80.6	94.6	3.0	0.0
		1.0	92.2	93.2	94.2	94.2	19.8	0.0
0.5	1.0	1.5	94.2	88.2	74.2	96.2	1.2	0.0
		2.0	97.0	86.0	53.0	95.0	0.0	0.0
		1.0	93.5	88.5	89.5	93.5	7.1	0.0
	2.0	1.5	97.4	85.4	32.4	96.4	0.0	0.0
		2.0	93.0	83.0	23.0	95.0	0.0	0.0
		1.0	90.9	87.9	86.9	92.9	10.9	82.1
	0.5	1.5	93.4	85.4	81.4	94.4	0.0	0.0
		2.0	96.2	83.2	78.2	97.2	0.0	0.0
		1.0	91.5	85.5	85.5	92.5	0.0	73.4
0.9	1.0	1.5	99.2	84.2	76.2	95.2	0.0	0.0
		2.0	86.7	77.7	49.7	94.7	0.0	0.0
		1.0	92.6	83.6	85.6	94.6	0.0	6.1
	2.0	1.5	97.8	78.8	38.8	93.8	0.0	0.0
		2.0	91.4	77.4	10.4	97.4	0.0	0.0

Table 2.3: Coverage of 95% CIs for $\hat{\beta}_1$ and $\hat{\eta}_{12}$ under correlated frailty, independent frailty and shared frailty when true mode is correlated frailty model

ρ	$e^{\alpha}(\times 10^{-3})$	σ_2/σ_1	Bias $\hat{\rho}(\times 10^{-3})$	SE $\hat{\rho}(\times 10^{-2})$	95% Coverage $\hat{\rho}$(%)	BIC(%)	
						Independent	Shared
		1.0	-2.092	3.365	98.2	99.2	99.9
	0.5	1.5	-5.789	3.043	93.0	99.8	100.0
		2.0	-6.152	3.111	95.3	100.0	100.0
		1.0	1.715	2.336	97.1	99.8	99.8
0.1	1.0	1.5	-0.143	2.220	95.4	100.0	100.0
		2.0	3.224	2.302	96.0	100.0	100.0
		1.0	-3.254	1.766	91.0	99.4	100.0
	2.0	1.5	-1.396	1.740	94.9	99.5	100.0
		2.0	-2.949	1.825	94.6	100.0	100.0
		1.0	-7.392	2.953	90.3	99.7	99.8
	0.5	1.5	1.723	2.586	96.4	99.6	100.0
		2.0	0.594	2.606	94.3	100.0	100.0
		1.0	-3.720	1.987	93.2	99.9	99.8
0.5	1.0	1.5	-0.882	1.858	95.5	100.0	100.0
		2.0	0.535	1.912	95.4	100.0	100.0
		1.0	1.814	1.459	90.4	100.0	100.0
	2.0	1.5	-0.161	1.422	94.8	99.9	100.0
		2.0	2.308	1.472	98.5	100.0	100.0
		1.0	-3.135	1.903	90.1	99.8	100.0
	0.5	1.5	-2.926	1.507	94.6	100.0	100.0
		2.0	-0.428	1.458	94.8	100.0	100.0
		1.0	-0.924	1.121	97.3	100.0	100.0
0.9	1.0	1.5	0.153	0.950	96.6	100.0	100.0
		2.0	0.567	0.933	92.8	100.0	100.0
		1.0	-1.291	0.698	92.8	100.0	100.0
	2.0	1.5	-0.072	0.625	94.5	100.0	100.0
		2.0	0.430	0.625	94.9	100.0	100.0

Table 2.4: Empirical bias, variance estimates, and coverage of 95% CIs for correlation parameter $\hat{\rho}$ when true mode is correlated frailty model

2.4.2 True Model Being Independent Frailty

We simulate data from independent frailty model and fit the data with three models respectively. We use similar settings as previous with $N = 10,000$ and $\tau = 1,000$ and intensity is same as before except the distribution of frailty:

$$\log \lambda_{i1}(t|\phi_{i1}, \mathcal{F}) = \alpha_1 + \phi_{i1} + \beta_1 X_i(t) + \eta_{11} Z_{i1}(t) + \eta_{12} Z_{i2}(t),$$
$$\log \lambda_{i2}(t|\phi_{i2}, \mathcal{F}) = \alpha_2 + \phi_{i2} + \beta_2 X_i(t) + \eta_{21} Z_{i1}(t) + \eta_{22} Z_{i2}(t),$$

where the frailty

$$\begin{pmatrix} \phi_{i1} \\ \phi_{i2} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} -\sigma_1^2/2 \\ -\sigma_2^2/2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right).$$

We generate the data with the same parameters $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ and vary baseline parameters $e^{\alpha_1} = e^{\alpha_2}$ among 0.0001,0.0005 and 0.001 and the ratio of standard deviation σ_2/σ_1 among 1.0, 1.5 and 2.0.

We show the estimation results of β_1 and η_{12} in table 2.5, 2.6 and 2.7.

Table 2.5 shows mean square error (MSE) and median absolute deviation (MAD) for $\hat{\beta}_1$ when the true model is independent frailty model. The error of the correlated frailty model is really close to the error of independent frailty model even if the later one is the true model. Except the case of sparse baseline, the shared frailty model generally has the significantly highest errors. In the case of sparse baseline $e^{\alpha} = 0.0005$, the shared frailty model has the lowest estimation errors even if the ratio of standard deviation ratio σ_2/σ_1 is large. This may be because in sparsity baseline, the simulated data does not give enough information for the structure within the frailty.

Table 2.6 shows mean square error (MSE) and median absolute deviation (MAD) for $\hat{\eta}_{12}$ when the true model is independent frailty model. The error of the correlated frailty model is really close to the error of the independent frailty even if the later one is the true model. The estimation error under the shared frailty model is still the worst and much higher. This gives evidence again that the effect estimation of internal covariates can be largely affected by the assumption of frailty structure.

Table 2.7 shows 95% confidence interval coverage for estimator $\hat{\beta}_1$ and $\hat{\eta}_{12}$. Both correlated frailty model and independent frailty model show around 95% for the confidence interval. For shared frailty model, the confidence interval coverage of β_1 is close to 95% with small baseline and σ_2/σ_1 ratio but it drops sharply when the baseline and σ_2/σ_1 ratio increase. The coverage of η_{12} are all zero at all cases for shared frailty model.

$e^{\alpha}(\times 10^{-3})$	σ_2/σ_1	MSE $\hat{\beta}_1(\times 10^{-3})$			MAD $\hat{\beta}_1(\times 10^{-2})$		
		Correlated	Independent	Shared	Correlated	Independent	Shared
0.5	1.0	2.777	2.763	2.724	3.325	3.314	3.130
	1.5	2.683	2.678	2.367	3.310	3.329	3.197
	2.0	2.681	2.676	2.087	3.191	3.237	3.132
1.0	1.0	1.170	1.173	4.253	2.144	2.151	4.195
	1.5	1.225	1.228	3.377	2.015	2.060	4.811
	2.0	1.095	1.088	3.374	2.193	2.163	5.250
2.0	1.0	0.569	0.572	5.269	1.698	1.672	5.002
	1.5	0.486	0.486	4.438	1.518	1.509	5.457
	2.0	0.575	0.573	3.751	1.742	1.681	5.861

Table 2.5: MSE and MAD for $\hat{\beta}_1$ under correlated frailty, independent frailty and shared frailty when true model is independent frailty model

$e^{\alpha}(\times 10^{-3})$	σ_2/σ_1	MSE $\hat{\beta}_1(\times 10^{-2})$			MAD $\hat{\beta}_1(\times 10^{-1})$		
		Correlated	Independent	Shared	Correlated	Independent	Shared
0.5	1.0	2.651	2.587	53.543	0.930	0.859	7.000
	1.5	2.624	2.473	242.702	1.188	1.120	15.691
	2.0	2.715	2.602	577.115	1.232	1.099	24.045
1.0	1.0	0.570	0.556	46.671	0.422	0.422	6.764
	1.5	0.504	0.499	213.215	0.503	0.504	14.514
	2.0	0.630	0.641	487.953	0.576	0.564	21.698
2.0	1.0	0.132	0.132	45.142	0.252	0.266	6.696
	1.5	0.171	0.158	179.078	0.276	0.272	13.243
	2.0	0.259	0.233	410.209	0.332	0.319	20.219

Table 2.6: MSE and MAD for $\hat{\eta}_{12}$ under correlated frailty, independent frailty and shared frailty when true model is independent frailty model

$e^{\alpha}(\times 10^{-3})$	σ_2/σ_1	95% Coverage $\hat{\beta}_1(\%)$			95% Coverage $\hat{\eta}_{12}(\%)$		
		Correlated	Independent	Shared	Correlated	Independent	Shared
0.5	1.0	96.0	96.0	94.0	93.0	93.0	0.0
	1.5	95.1	95.1	84.1	95.1	95.1	0.0
	2.0	93.2	93.2	81.2	96.2	98.2	0.0
1.0	1.0	96.5	96.5	90.5	94.5	94.5	0.0
	1.5	93.1	92.1	74.1	96.1	95.1	0.0
	2.0	96.5	96.5	65.5	97.5	97.5	0.0
2.0	1.0	94.1	94.1	80.1	95.1	97.1	0.0
	1.5	94.2	95.2	34.2	94.2	97.2	0.0
	2.0	95.4	94.4	27.4	91.4	92.4	0.0

Table 2.7: Coverage of 95% CIs for $\hat{\beta}_1$ and $\hat{\eta}_{12}$ under correlated frailty, independent frailty and shared frailty when true model is independent frailty model

2.4.3 True Model Being Shared Frailty

We simulate data from shared frailty model and fit the data with three models respectively. We use similar settings as above with $N = 10,000$ and $\tau = 1,000$ and the intensity is same as before except the distribution of frailty:

$$\log \lambda_{i1}(t|\phi_i, \mathcal{F}) = \alpha_1 + \phi_i + \beta_1 X_i(t) + \eta_{11} Z_{i1}(t) + \eta_{12} Z_{i2}(t),$$
$$\log \lambda_{i2}(t|\phi_i, \mathcal{F}) = \alpha_2 + \phi_i + \beta_2 X_i(t) + \eta_{21} Z_{i1}(t) + \eta_{22} Z_{i2}(t),$$

where the frailty

$$\phi_i \sim \mathcal{N}(-\sigma^2/2, \sigma^2).$$

We generate the data with the same parameters $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ and vary baseline parameters $e^{\alpha_1} = e^{\alpha_2}$ among 0.0001,0.0005 and 0.001 and the frailty standard deviation σ among 1.0, 1.5 and 2.0.

We show estimation results of β_1 and η_{12} in table 2.8, 2.9 and 2.10.

Table 2.8 shows mean square error (MSE) and median absolute deviation (MAD) for $\hat{\beta}_1$ when the true model is shared frailty model. The error for the correlated frailty model is really close to the error for the shared frailty model even if the later one is the true model. The independent frailty model has relatively larger error in both MSE and MAD. Table 2.9 shows mean square error (MSE) and median absolute deviation (MAD) for $\hat{\eta}_{12}$ when the true model is shared frailty model. Similarly, the error for correlated frailty model is really close to the error for the true model and the independent frailty model gives relatively much larger error. Table 2.10 shows 95% confidence interval coverage for estimator $\hat{\beta}_1$ and $\hat{\eta}_{12}$. Both correlated frailty model and shared frailty model show around 95% for the confidence interval, except the case with large frailty standard deviation. For independent frailty model, the confidence interval coverage of β_1 is a little lower than the true model, most of which are around 90%. The coverage for η_{12} is almost zero at all cases for independent frailty model.

$e^{\alpha}(\times 10^{-3})$	σ	MSE $\hat{\beta}_1(\times 10^{-3})$			MAD $\hat{\beta}_1(\times 10^{-2})$		
		Correlated	Independent	Shared	Correlated	Independent	Shared
0.5	1.0	3.482	3.976	3.502	4.514	4.538	4.562
	1.5	3.726	5.274	3.645	3.863	3.971	3.907
	2.0	2.385	2.767	2.308	3.127	3.788	3.213
1.0	1.0	1.114	2.004	1.106	2.004	2.872	1.995
	1.5	1.113	2.150	1.168	2.295	3.679	2.357
	2.0	1.129	1.576	1.068	2.336	2.886	2.362
2.0	1.0	0.640	1.464	0.642	1.763	2.962	1.827
	1.5	0.643	1.122	0.641	1.672	2.621	1.653
	2.0	0.609	0.923	0.566	1.314	2.374	1.361

Table 2.8: MSE and MAD for $\hat{\beta}_1$ under correlated frailty, independent frailty and shared frailty when true model is shared frailty model

$e^{\alpha}(\times 10^{-3})$	σ	MSE $\hat{\beta}_1(\times 10^{-3})$			MAD $\hat{\beta}_1(\times 10^{-2})$		
		Correlated	Independent	Shared	Correlated	Independent	Shared
0.5	1.0	5.339	153.973	5.614	4.302	38.340	4.548
	1.5	2.154	62.765	2.047	3.199	23.624	3.007
	2.0	2.218	27.986	1.439	3.163	15.206	2.673
1.0	1.0	2.114	74.794	2.028	3.028	27.031	3.242
	1.5	1.099	35.506	0.848	2.030	18.500	1.803
	2.0	1.163	19.985	0.540	2.323	13.633	1.744
2.0	1.0	0.503	33.504	0.384	1.503	18.176	1.309
	1.5	0.543	19.362	0.228	1.390	13.691	0.960
	2.0	0.609	13.370	0.240	1.943	11.247	1.034

Table 2.9: MSE and MAD for $\hat{\eta}_{12}$ under correlated frailty, independent frailty and shared frailty when true model is shared frailty model

$e^{\alpha}(\times 10^{-3})$	σ	95% Coverage $\hat{\beta}_1$(%)			95% Coverage $\hat{\eta}_{12}$(%)		
		Correlated	Independent	Shared	Correlated	Independent	Shared
0.5	1.0	92.1	90.1	92.1	98.1	1.4	98.1
	1.5	95.1	89.1	94.1	98.1	0.0	99.1
	2.0	97.9	95.9	97.9	92.9	6.2	96.9
1.0	1.0	94.5	91.5	95.5	96.5	0.0	95.5
	1.5	96.6	89.6	95.6	94.6	0.0	95.6
	2.0	97.1	89.1	98.1	85.1	1.0	96.1
2.0	1.0	93.4	78.4	90.4	95.4	0.0	98.4
	1.5	92.1	87.1	92.1	86.1	0.0	96.1
	2.0	92.0	90.0	92.0	78.0	0.0	94.0

Table 2.10: Coverage of 95% CIs for $\hat{\beta}_1$ and $\hat{\eta}_{12}$ under correlated frailty, independent frailty and shared frailty when true model is shared frailty model

2.5 OMOP Data Analysis for Drug-AE Pairs

2.5.1 Background

The U.S. Food and Drug Administration (FDA) has been developing a system to detect adverse events of marketed drugs and other medical products. FDA relies on submission of spontaneous reports in which clinicians or patients conclude that a drug may be the cause of the certain adverse clinical events. These records or reports are submitted voluntarily from clinicians or patients and will vary with their skills, experience, clinical history and existence of other plausible explanations. As a result, active surveillance of medical products applied on longitudinal observational databases (LODs) has been brought to the forefront. Observational Outcomes Medical Partnership (OMOP) project is prompting such observational databases and developing active surveillance methods.

The OMOP project is a public-private partnership among database source owners, academic institutions, the FDA and pharmaceutical industries and is administered by the Foundation for the National Institutes of Health [31]. The OMOP project is initially to explore the possible methods to make use of observational data for drug safety problems in marketed prescription drugs.

Some articles use four databases (CCAE, MDCD, MDCR and MSLR) in table 2.11 from the OMOP collaborators to evaluate the performance of their proposed models. In this thesis, we focus on an electronic health record (EHR) dataset from a healthcare and insurance company and identify adverse events and drugs with ICD-9 codes and brand names. With the database, we compare our proposed model with some of existing methods: disproportional analysis (DP), observational screening (OS), SCCS and MSCCS methods. Disproportional analysis is one of the most common methods for surveillance in spontaneous report systems. The AE-drug pair is identified if the co-occurrence is more frequently than it under independence condition [38]. Observational Screen [25, 24] is based on screening rates, which give the number of adverse events that occur during the drug exposure divided by the cumulative time exposed to the drug. The measure of association in OS is the screening rate ratio, which is the ratio of the screening rate in a treatment group to it in a control group.

OMOP Collaborator	Database	Insurance Type	Available individuals (millions)
Thomson Reuters Commercial claims	CCAE	Multiple private insurers	58
Medicare supplement	MDCR	Medicare supplement	4.4
Medicaid	MDCD	Medicaid	11.1
MarketScan Lab	MSLR	Multiple	1.5

Table 2.11: Four frequently used OMOP Database for analysis and inference

2.5.2 Adverse Events and Drug Exposures

The OMOP project has established a series of definitions for adverse events and drugs. The adverse events selected by the OMOP project are angioedema, aplastic anemia, acute liver injury, bleeding, hospitalization due to gastrointestinal (GI) ulcer, hip fracture, hospitalization, acute MI, mortality after MI, and acute renal failure. These adverse events were chosen based on one of the following ways:

- Included in the always expedited spontaneous adverse-event report list.

- Appeared in a boxed warning.
- Represented by the spectrum of incidence, background rate, and time to onset, of typically drug-related adverse events.
- Likely to be the focus of ongoing drug safety surveillance .

The OMOP project also selects 10 drugs: angiotensin converting enzyme (ACE) inhibitors, amphotericin B, antibiotics, antiepileptics, benzodiazepines, beta blockers, bisphosphonates, tricyclic antidepressants, typical antipsychotics, and warfarin. Considering adverse events and drugs, there are in total 100 possible pairs of drugs and AEs. However, there are only 9 AE-drug pairs classified as "positive control" and 44 AE-drug pairs classified as "negative control".

Here the "positive control" means true associations between the adverse events and drugs, which are determined by prior published observational database research or the adverse events listed in the label of the drugs. The "negative control" does not have such evidence. Researchers always evaluate the performance of the analytical methods by their ability to identify these positive control pairs and negative control pair. Table 2.12 shows the 53 AE-Drug pairs.

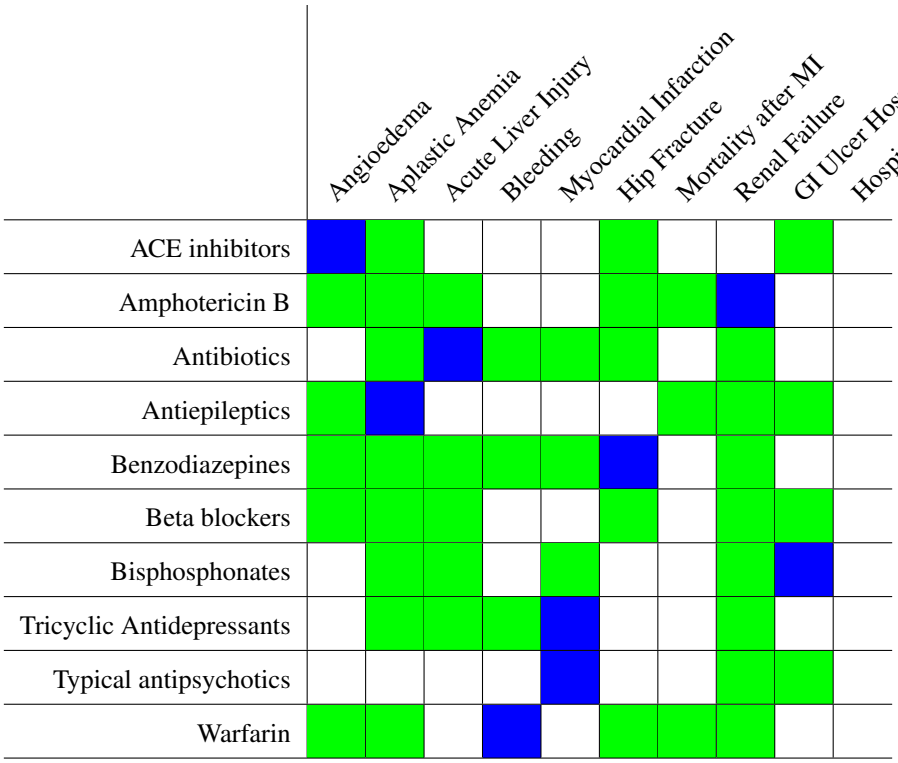

Table 2.12: Positive control and Negative control AE-Drug pairs

2.5.3 Evaluation Metrics

We introduce the rank-based (and threshold independent) performance measures to evaluate our proposed model and other existing methods. Rank-based performance metrics are based on receiver operating characteristic (ROC) curve. Before introducing these metrics, we need the knowledge of confusion matrix which is the foundation for the ROC curve.

A confusion matrix is a table used to describe the performance of a classification model where the true values are known. The basic terms of true positive, true negative, false positive and false negative are defined in table 2.13. Based on these terms, we can have more definitions:

- The true positive rate (sensitivity) is defined as the number of true positive subjects over the total number of subjects which true is positive, i.e $a/(a+c)$.
- The false positive rate (1-specificity) is the number of false positive subjects over the total number of subjects which true is negative, i.e $b/(b+d)$.

- The precision is defined by $a/(a+b)$.

Table 2.13: Confusion Matrix

	Fact is Positive	Fact is Negative
Predicted as Positive	True Positive (a)	False Positive (b)
Predicted as Negative	False Negative (c)	True Negative (d)
	9	44

We use the relative risk estimators of 9 positive pairs and 44 negative pairs in model evaluation. We treat these $M = 53$ estimators as a classification problem. Let these estimators be $e^{\hat{\beta}_{(1)}} < e^{\hat{\beta}_{(2)}} < \cdots < e^{\hat{\beta}_{(M)}}$. With a threshold level c, we say we predict the pair with a point estimator $e^{\hat{\beta}_{(m)}}$ that $m > c$ as a positive label and the pair with a point estimator $e^{\hat{\beta}_{(m)}}$ that $m \leq c$ as a negative label. In this way, each method, with threshold level $0 \leq c \leq M$, we have total $M+1 = 54$ possible prediction results. We can define the rank-based metrics based on these prediction results.

- **Area Under Curve (AUC)** is based on Receiver Operating Characteristic (ROC) cure, which is plotted by sensitivity versus 1−specificity. For each threshold level $c = 0, \cdots, M$, we have a prediction result which gives a pair (x_c, y_c) where x_c and y_c are sensitivity and 1−specificity respectively at level c. ROC curve is plotted by these points and we hope this curve has a sharp corner close to the upper left-hand corner of the plot. This is because a good classification model should have large sensitivity and small 1−specificity. Based on ROC curve, we can calculate the AUC which is the integral of this curve. Larger value of AUC gives better classification results. $AUC = 1$ gives a perfect classification model and $AUC = 0.5$ is a randomly guess of labels. Some works [3] suggested that AUC is a better summary measure of classification performance than overall accuracy.

- **Partial Area Under Curve (PAUC)** is a possible alternative for AUC. Sometimes we only focus on a limited range of 1− specificity because large value of 1−specificity is not useful and practical. We want to observe the model performance under small false positives. In this way, we define PAUC as the integral of the area up to a cut-off at a certain level of

1−specificity. In this thesis, we choose such cut-off as 30% so that the maximum value of PAUC is 0.3.

- **Mean Average Precision (MAP) score** is defined by the average of precision at some special threshold level $\{c_1, \cdots, \}$. In this case, we have 9 threshold levels $c_1, \cdots, c_9$. c_k is defined as the minimum index such that if we take c_k as threshold level, we would have k true positive predictions.

- **The sensitivity at a** 5% **false positive rate** is defined as the height of the ROC curve when c is set to make 1−specificity as 5%. The larger value means better model performance.

- **Average False Positive Rate (Average FPR)** is similar to MAP score. But it is defined by the average of false positive rate at threshold level $\{c_1, \cdots, c_9\}$. Unlike other metrics, the larger values of Average FPR corresponds to a higher proportion of false positives and worse method performance.

2.5.4 Real Data Results

We compare the proposed multivariate correlated frailty model (MCFM) with other existing methods DP, OS, SCCS and MSCCS in terms of the ROC curve and metrics define above. Figure 2.2 and 2.3 show ROC curves and 5 metrics in an electronic health record dataset. Table 2.14 gives numerical details about the values of the metrics in Figure 2.3. The term SCCS refers to simple SCCS method which models one adverse event and one drug exposure while MSCCS refers to multiple SCCS method which models one adverse event and multiple drug exposures. There are several model settings for each method, including definition of drug exposure covariates and identification of adverse events with diagnosis code. For simplicity, we use the similar parameter settings [30], where the models are set to give the optimal AUC values. Besides, in our proposed multivariate correlated frailty model (MCFM), we incorporate adverse events as indicator covariates which equal to 1 if such adverse event occurs before and 0 otherwise.

The ROC curves in Figure 2.2 for all five methods can give us a general sense for their per-

formances in AE-drug pair classification in the electronic health record database. The graph is a polygonal line with two direction, up and right. Up direction means that we identify a positive pair as positive and right direction means that we misidentify a negative pair as positive. Compared with MCFM and MSCCS, the other three models (DP, OS, SCCS) are closer to the diagonal and thus has smaller AUC metric values. Besides, MCFM can identify more positive associations at lower values of 1-specificity than the other methods, which leads a smaller value of Average FPR.

The figure 2.3 shows that our proposed approach MCFM performs the best across all five metrics in the electronic health record database. The AUC value of MCFM is around 0.84 which is an evidence that MCFM can work relatively well in the classification problem. The MSCCS approach has the second highest AUC value and this is because both MSCCS and MCFM include the confounding bystander effect and incorporate multiple drug exposures, which result in better performance in AE-drug pair classification.

For MAP Score, PAUC at 30% and AFPR, MCFM still has the best performance among all five methods while MSCCS has the second best performance. Due to the small number of positive pairs, the value for sensitivity at 5% only have 10 possible values ($\{0/9, 1/9, 2/9, \cdots, 9/9\}$) and even less possible values in a low 1-specificity level (e.g. we only have 2 values $\{1/9, 2/9\}$ in this column). It is difficult to base an assessment on this metric for different methods.

The above results are subject to the identification of AE-drug associations. The positive pairs are based on labels of drugs, previous observational research and expert consensus. And the small number of the positive pairs limits the accuracy of classification problems and model performances. The metrics are sensitive to the definitions of these positive associations. It is necessary to have more positive and negative controls to reach a more convincing evaluation. OMOP project is working on investigating more possible AE-drug pairs for model evaluation and we can extend our work in the future. Put aside the limitation, the above results gives a preliminary assessment of the multivariate correlated frailty model in postmarketing surveillance. Compared with other methods, the proposed model is always the best performer. Actually, adjusting for confounding drug exposures and implementing the correlation between adverse events can improve the accuracy

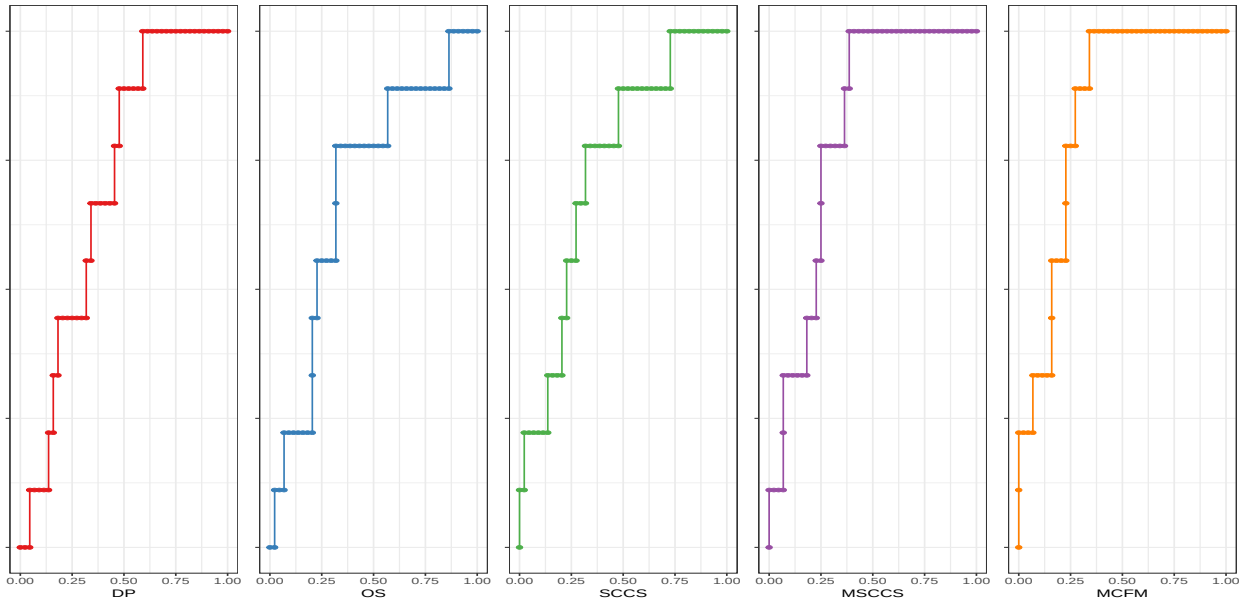

Figure 2.2: Metrics for DP, OS, SCCS, MSCCS and MCFM in the electronic health record database to classify AE-drug pairs.

Method	AUC	PAUC at 30%	MAP	Sensitivity at 5%	AFPR
DP	0.699	0.075	0.284	0.111	0.301
OS	0.689	0.086	0.318	0.111	0.311
SCCS	0.735	0.104	0.423	0.222	0.265
MSCCS	0.801	0.117	0.443	0.111	0.199
MCFM	0.838	0.143	0.538	0.222	0.162

Table 2.14: Metric values for DP, OS, SCCS, MSCCS and MCFM in the electronic health record database

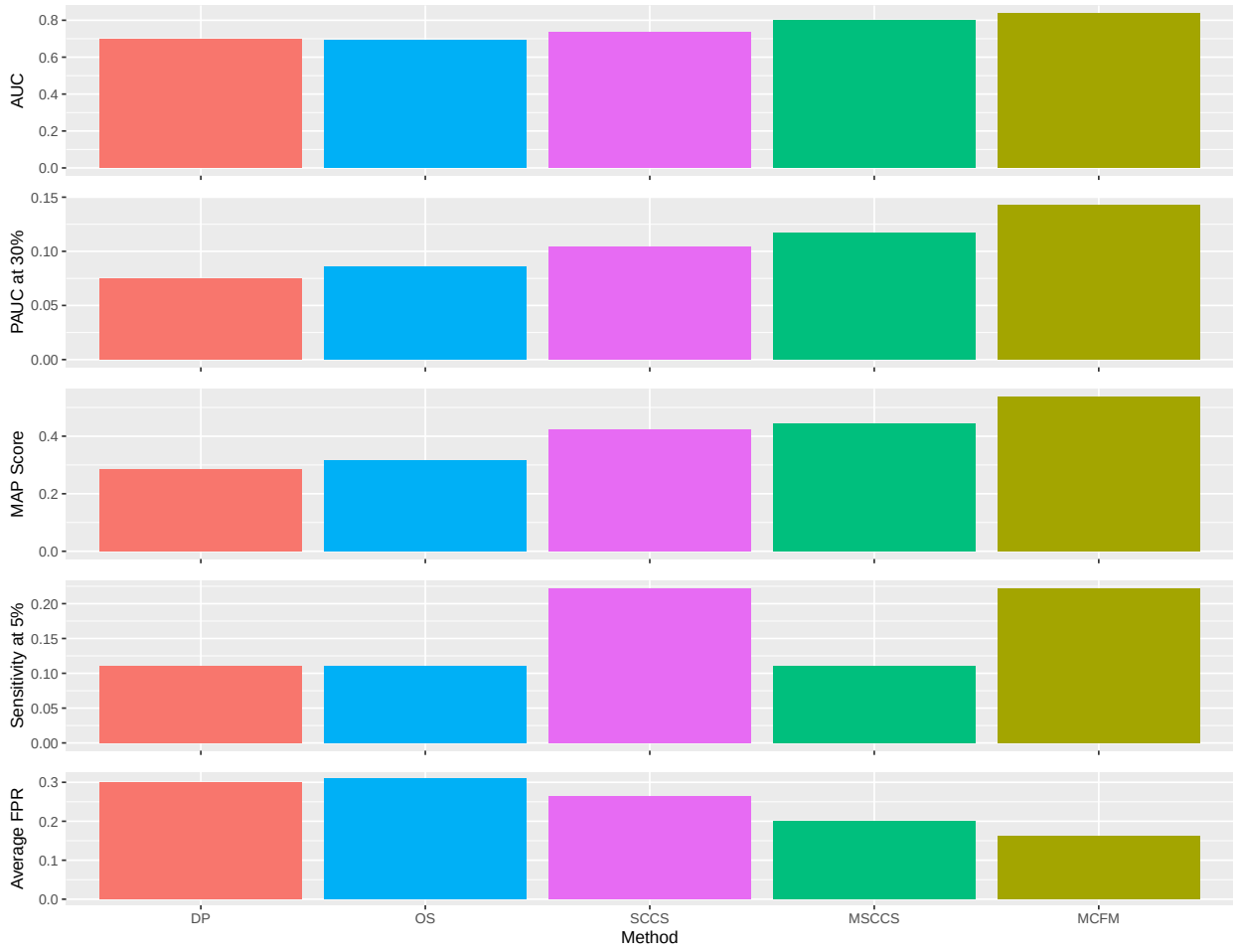

Figure 2.3: ROC curve for DP, OS, SCCS, MSCCS and MCFM in the electronic health record database

2.6 Appendix

2.6.1 Bayesian Sampling Algorithm

Recall the full likelihood:

$$
\begin{aligned}
&L_{\text{full}}(D_{\text{obs}}, \mathcal{R}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}) \\
&\propto \prod_{i=1}^{N}\prod_{j=1}^{J}\left[\prod_{k=1}^{n_{ij}} Y_{ij}(t_{ijk}) e^{\alpha_j+\phi_{ij}+\boldsymbol{\beta}_j^T \boldsymbol{X}_i(t_{ijk})+\boldsymbol{\eta}_j^T \boldsymbol{Z}_i(t_{ijk})} \times \exp\left\{-\int_0^{\tau} Y_{ij}(t) e^{\alpha_j+\phi_{ij}+\boldsymbol{\beta}_j^T \boldsymbol{X}_i(t)+\boldsymbol{\eta}_j^T \boldsymbol{Z}_i(t)} dt\right\}\right] \\
&\quad \times \prod_{i=1}^{n} det(\Omega)^{1/2} \exp\left\{-\frac{1}{2}\boldsymbol{\phi}_i^T \Omega \boldsymbol{\phi}_i\right\} \\
&\quad \times \prod_{j=1}^{J} \exp\left\{-\frac{\alpha_j^2}{2000}\right\} \prod_{j=1}^{J}\prod_{l=1}^{L} \exp\left\{-\frac{\beta_{jl}^2}{2000}\right\} \prod_{j=1}^{J}\prod_{j'=1}^{J} \exp\left\{-\frac{\eta_{jj'}^2}{2000}\right\} \\
&\quad \times |\boldsymbol{\Omega}|^{(a-J-1)/2} e^{-tr(\boldsymbol{V}^{-1}\boldsymbol{\Omega})/2}.
\end{aligned}
$$

The conditional distribution of baseline parameters α_j is

$$
\begin{aligned}
&\pi_c(\alpha_j | D_{\text{obs}}, \mathcal{R}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\alpha}_{-j}) \\
&\propto \prod_{i=1}^{N}\left[\prod_{k=1}^{n_{ij}} Y_{ij}(t_{ijk}) e^{\alpha_j+\phi_{ij}+\boldsymbol{\beta}_j^T \boldsymbol{X}_i(t_{ijk})+\boldsymbol{\eta}_j^T \boldsymbol{Z}_i(t_{ijk})} \times \exp\left\{-\int_0^{\tau} Y_{ij}(t) e^{\alpha_j+\phi_{ij}+\boldsymbol{\beta}_j^T \boldsymbol{X}_i(t)+\boldsymbol{\eta}_j^T \boldsymbol{Z}_i(t)} dt\right\}\right] \\
&\quad \times \exp\left\{-\frac{\alpha_j^2}{2000}\right\} \\
&\propto \exp\left\{\left(\sum_{i=1}^{N} n_{ij}\right)\alpha_j - \frac{\alpha_j^2}{2000} - e^{\alpha_j} \sum_{i=1}^{N}\int_0^{\tau} Y_{ij}(t) e^{\phi_{ij}+\boldsymbol{\beta}_j^T \boldsymbol{X}_i(t)+\boldsymbol{\eta}_j^T \boldsymbol{Z}_i(t)} dt\right\}.
\end{aligned}
$$

The conditional distribution of drug exposure effect parameters β_{jl} is

$$
\begin{aligned}
&\pi_c(\beta_{jl} | D_{\text{obs}}, \mathcal{R}, \boldsymbol{\beta}_{-jl}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}) \\
&\propto \exp\left\{\left(\sum_{i=1}^{N}\sum_{k=1}^{n_{ij}} X_{il}(t_{ijk})\right)\beta_{jl} - \frac{\beta_{jl}^2}{2000} - \sum_{i=1}^{N}\int_0^{\tau} Y_{ij}(t) e^{\alpha_j+\phi_{ij}+\boldsymbol{\beta}_j^T \boldsymbol{X}_i(t)+\boldsymbol{\eta}_j^T \boldsymbol{Z}_i(t)} dt\right\} \\
&\propto \exp\left\{\left(\sum_{i=1}^{N}\sum_{k=1}^{n_{ij}} X_{il}(t_{ijk})\right)\beta_{jl} - \frac{\beta_{jl}^2}{2000} - e^{\beta_{jl}} \sum_{i=1}^{N}\int_{t\in\{t:X_{il}(t)=1\}} Y_{ij}(t) e^{\alpha_j+\phi_{ij}+\boldsymbol{\beta}_{-jl}^T \boldsymbol{X}_{-il}(t)+\boldsymbol{\eta}_j^T \boldsymbol{Z}_i(t)} dt\right\}.
\end{aligned}
$$

The conditional distribution of adverse event effect parameters $\eta_{jj'}$ is

$$
\begin{aligned}
&\pi_c(\eta_{jj'}|D_{\text{obs}}, \mathcal{R}, \boldsymbol{\beta}, \boldsymbol{\eta}_{-jj'}, \boldsymbol{\alpha}, \boldsymbol{\Omega}) \\
&\propto \exp\left\{\left(\sum_{i=1}^{N}\sum_{k=1}^{n_{ij}} Z_{ij'}(t_{ijk})\right)\eta_{jj'} - \frac{\eta_{jj'}^2}{2000} - \sum_{i=1}^{N}\int_0^\tau Y_{ij}(t)e^{\alpha_j+\phi_{ij}+\boldsymbol{\beta}_j^T\boldsymbol{X}_i(t)+\boldsymbol{\eta}_j^T\mathbf{Z}_i(t)}dt\right\} \\
&\propto \exp\left\{\left(\sum_{i=1}^{N}\sum_{k=1}^{n_{ij}} Z_{ij'}(t_{ijk})\right)\eta_{jj'} - \frac{\eta_{jj'}^2}{2000} - e^{\eta_{jj'}}\sum_{i=1}^{N}\int_{t\in\{t:Z_{ij'}(t)=1\}} Y_{ij}(t)e^{\alpha_j+\phi_{ij}+\boldsymbol{\beta}_j^T\boldsymbol{X}_i(t)+\boldsymbol{\eta}_{-jj'}^T\mathbf{Z}_{-ij'}(t)}dt\right\}.
\end{aligned}
$$

The conditional distribution of frailty covariance parameters $\boldsymbol{\Omega}$ is

$$
\begin{aligned}
&\pi_c(\boldsymbol{\Omega}|D_{\text{obs}}, \mathcal{R}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}) \\
&\propto \prod_{i=1}^{n} det(\Omega)^{1/2}\exp\left\{-\frac{1}{2}\boldsymbol{\phi}_i^T\Omega\boldsymbol{\phi}_i\right\} \times |\boldsymbol{\Omega}|^{(a-J-1)/2}e^{-tr(\boldsymbol{V}^{-1}\boldsymbol{\Omega})/2} \\
&\propto |\boldsymbol{\Omega}|^{(N+a-J-1)/2}\exp\left\{-\frac{1}{2}tr\left(\left(\boldsymbol{V}^{-1}+\sum_{i=1}^{N}\boldsymbol{\phi}_i\boldsymbol{\phi}^T\right)\boldsymbol{\Omega}\right)\right\} \\
&\sim \mathcal{W}\left(\boldsymbol{\Omega}; N+a, \left(\boldsymbol{V}^{-1}+\sum_{i=1}^{N}\boldsymbol{\phi}_i\boldsymbol{\phi}^T\right)^{-1}\right).
\end{aligned}
$$

The value of ϕ_i depends on the values of the hyperparameters $\boldsymbol{\Omega}$, effect parameters $\boldsymbol{\beta}, \boldsymbol{\eta}$, baseline parameters $\boldsymbol{\alpha}$ and all event times of individual i. So the conditional distribution of the frailty $\boldsymbol{\phi}_i$ is

$$
\begin{aligned}
&\pi_c(\boldsymbol{\phi}_i|D_{\text{obs}}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}) \\
&\propto \exp\left\{\sum_{j=1}^{J} n_{ij}\phi_{ij} - \frac{1}{2}\boldsymbol{\phi}_i^T\Omega\boldsymbol{\phi}_i - \sum_{j=1}^{J}e^{\phi_{ij}}\int_0^\tau Y_{ij}(t)e^{\alpha_j+\boldsymbol{\beta}_j^T\boldsymbol{X}_i(t)+\boldsymbol{\eta}_j^T\mathbf{Z}_i(t)}dt\right\}.
\end{aligned}
$$

Adaptive Rejection Metropolis Sampling (ARMS) Algorithm Except the hyperparameter Ω, we can not sample directly from the conditional distribution of parameters $\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}$ and frailty $\boldsymbol{\phi}_i$. Thus we use the adaptive rejection sampling [13] to generate the samples from above conditional distributions.

Let $f(x)$ denote the target distribution. Adaptive rejection sampling (ARS) is an improvement of reject sampling algorithm. In reject sampling, we generate independent samples from $f(x)$ by

accepting samples generated from another proposal distribution $g(x)$ with some probability, where there is a finite constant m such that $mg(x) \geq f(x)$ a.s.. However, such proposal distribution $g(x)$ is difficult to find and in most times, the value of m is large, which means a lot of unwanted samples will be taken before we accept a sample. Instead of fixed proposal distribution $g(x)$ in reject sampling algorithm, adaptive reject sampling algorithm improves efficiency by incorporating into $g(x)$ with information about $f(x)$ obtained at each iteration. That is to say, we use a piecewise linear function for the proposal log distribution $g_n(x)$ as the "envelope" warping the target log concave distribution, where $g_n(x)$ is updated with the rejected points at each iteration. Actually, the above conditional distributions are all log concave due to the concavity of the function $f(x) = Ax - Bx^2 - Ce^x$ with positive values of A, B, C. This is because $f''(x) = -2B - Ce^x < 0$.

Let $S_n = \{x_i; i = 0, 1, \cdots, n+1\}$ denote a current set of points in ascending order, where x_0 and x_{n+1} are the possibly lower and upper limits of the domain of the target density function $f(x)$. For $1 \leq i \leq j \leq n$, let $L_{ij}(x; S_n)$ denote the straight line through points $[x_i, \log f(x_i)]$ and $[x_j, \log f(x_j)]$. For other (i, j) pairs, let $L_{01}(x; S_n)$ and $L_{n,n+1}(x; S_n)$ be vertical line at $x = x_1$ and $x = x_n$ respectively. Then we define a piecewise linear function $h_n(x)$:

$$h_n(x; S_n) = \begin{cases} L_{1,2}(x; S_n) & \text{if } x < x_1, \\ \min\left[L_{i-1,i}(x; S_n), L_{i+1,i+2}(x; S_n)\right] & \text{if } x_i \leq x < x_{i+1}, 1 \leq i \leq n-1, \\ L_{n-1,n}(x; S_n) & \text{if } x \geq x_n. \end{cases} \tag{2.6}$$

Such function $h_n(x; S_n)$ is an envelope for $\log f(x)$, i.e $h_n(x; S_n) \geq \log f(x)$ a.s., which is illustrated in Fig 2.4.

Then we can construct the proposal distribution $g_n(x)$ in reject sampling:

$$g_n(x) = \frac{1}{\int \exp h_n(x) dx} \exp h_n(x), \tag{2.7}$$

where $g_n(x)$ is a piecewise exponential distribution and can be updated in each iteration where $f(x)$ is evaluated. We have the following algorithm for ARS:

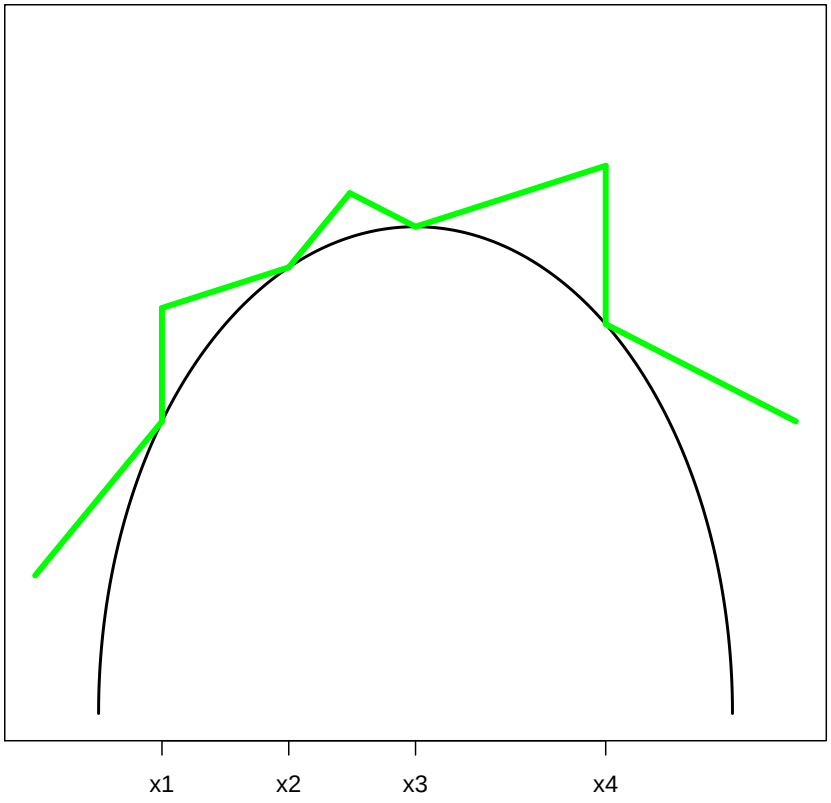

Figure 2.4: Adaptive rejection function $h_4(x)$

Algorithm 1: Adaptive Rejection Sampling

Result: Get a new sample X_A from target density $f(x)$.

Initialize n and S_n;

while X_A is not accepted **do**

  Sample X from $\exp h_n(x)$ and sample U from $U(0,1)$;

  if $U > f(X)/\exp h_n(X)$ **then**

    Reject X;

    $S_{n+1} = S_n \cup \{X\}$, keep ascending order;

    Continue to sample ;

  else

    Accept X and set $X_A = X$.

  end

end

2.6.2 Frailty Theorems

An example when frailty and covariate are correlated

We illustrate here that when the covariates are correlated with frailty or outcomes, the effect parameter estimation can be biased under independent assumption. We use a Poisson regression model:

$$Y_i|\lambda_i \sim \text{Poi}(\lambda_i),$$

$$\lambda_i = \gamma_i \exp(\beta_0 + \beta X_i).$$

Y_i is the outcome with a Poisson distribution based on covariate X_i and frailty γ_i. The frailty γ_i follows a gamma distribution $\text{Gamma}(\theta,\theta) = \frac{\theta^\theta}{\Gamma(\theta)}\gamma_i^{\theta-1}e^{-\theta\gamma_i}$. We propose a correlation setting of frailty and covariate:

$$X_i = \begin{cases} 1 & \text{if } \gamma_i > c, \\ 0 & \text{if } \gamma_i < c. \end{cases}$$

with $c > 1$. Let M_0 denote the true model and M_1 denote model with assumption that frailty and

covariate are independent. We can prove that the estimation of parameter β under M_1 is biased.

The generalized estimating equation under assumption M_1 is

$$0 = \frac{1}{N}\sum_{i=1}^{N} X_i(Y_i - \mathrm{E}_\gamma(Y_i|X_i)) = \frac{1}{N}\sum_{i=1}^{N} X_i(Y_i - e^{\beta X_i}).$$

Let β_0 and $\hat{\beta}_n$ be the true value of β and the root of generalized estimating equation, respectively. Then

$$f_n(\beta) = \frac{1}{N}\sum_{i=1}^{N} X_i(Y_i - e^{\beta X_i}) \qquad \text{and} \qquad f_n(\hat{\beta}_n) = 0.$$

By Uniform Strong Law of Large Numbers, with probability 1, $f_n(\beta)$ uniformly converge to

$$\begin{aligned} f(\beta) =& \mathrm{E}_{X,Y}\left(XY - Xe^{\beta X}\right) = \mathrm{E}_{X,Y}\left(XY\right) - \mathrm{E}_{X,Y}\left(Xe^{\beta X}\right) \\ =& \mathrm{E}_{X,\gamma}\left(\mathrm{E}_Y(XY|X,\gamma)\right) - \mathrm{E}_{X,\gamma}\left(Xe^{\beta X}\right) \\ =& \mathrm{E}_{X,\gamma}\left(X\gamma e^{\beta_0 X}\right) - \mathrm{E}_{X,\gamma}\left(Xe^{\beta X}\right). \end{aligned}$$

$f(\beta)$ is continuous and

$$\begin{aligned} f(\beta_0) =& \mathrm{E}_{X,\gamma}\left(X\gamma e^{\beta_0 X}\right) - \mathrm{E}_{X,\gamma} X e^{\beta_0 X} \\ =& \mathrm{E}_\gamma\left(\gamma \mathcal{I}(\gamma > c) e^{\beta_0}\right) - \mathrm{E}_\gamma\left(\mathcal{I}(\gamma > c) e^{\beta_0}\right) \\ >& 0, \end{aligned}$$

as $c > 1$.

Then it suffices to show that $\{\hat{\beta}_n\}_{n=1}^{\infty}$ does not converge to β_0.

Theorem 2.1 *We have functions $\{f_n\}_{n=1}^{\infty}$ uniformly converge to a continuous function f and $f_n(\hat{\beta}_n) = 0$ and $f(\beta_0) = C > 0$. Then $\{\hat{\beta}_n\}_{n=1}^{\infty}$ does not converge to β_0.*

Proof By uniformly convergence, $\forall \epsilon > 0, \exists N_1, \forall n > N_1, \sup_x |f_n(x) - f(x)| < \epsilon$. Since f is continuous at β_0, $\exists \xi > 0, |f(\hat{\beta}) - f(\beta_0)| < \epsilon$ for all $\hat{\beta}$ that $|\hat{\beta} - \beta_0| < \xi$. If $\hat{\beta}_n$ converges to β_0, then $\exists N_2, \forall n > N_2, |\hat{\beta}_n - \beta_0| < \xi$ and thus $|f(\hat{\beta}_n) - f(\beta_0)| < \epsilon$.

Choose $\epsilon = \frac{C}{4}$ and $n > \max\{N_1, N_2\}$, we have

$$\begin{aligned} C = |f_n(\hat{\beta}_n) - f(\beta_0)| &= |f_n(\hat{\beta}_n) - f(\hat{\beta}_n) + f(\hat{\beta}_n) - f(\beta_0)| \\ &\leq |f_n(\hat{\beta}_n) - f(\hat{\beta}_n)| + |f(\hat{\beta}_n) - f(\beta_0)| \\ &< 2\epsilon = \frac{C}{2}. \end{aligned}$$

It is a contradiction so $\{\hat{\beta}_n\}_{n=1}^{\infty}$ does not converge to β_0. □

Gamma Frailty model and SCCS method

We prove that when the density function of frailty satisfies some conditions, the estimators of frailty model would converge to the estimators of the SCCS method.

Lemma 2.1 *If $\{f_n\}$ and $\{g_n\}$ converge uniformly on a set E, then $\{f_n + g_n\}$ converges uniformly on E. If, in addition, $\{f_n\}$ and $\{g_n\}$ are sequences of bounded functions, then $\{f_n g_n\}$ converge uniformly on E.*

Lemma 2.2 *If $\{f_n\}$ converge uniformly to f on a set E. Let $\hat{\beta}_n$ be unique root of f_n and $\hat{\beta}$ be unique root of f. If f is a continuous and strict monotone function, we have*

$$\hat{\beta}_n \to \hat{\beta}.$$

Proof If not, then $\exists\epsilon$, for all $N > 0$, $\exists n > N$ such that $|\hat{\beta}_n - \hat{\beta}| > \epsilon$. Let $\eta = \inf_{|x-\hat{\beta}|>\epsilon} |f(x) - f(\hat{\beta})| > 0$. Since $\{f_n\}$ converge uniformly to f, then there exists N such that $|f_n(\hat{\beta}_n) - f(\hat{\beta}_n)| < \frac{\eta}{2}$. Then we have

$$\begin{aligned} 0 &= |f_n(\hat{\beta}_n) - f(\hat{\beta})| \\ &= |f_n(\hat{\beta}_n) - f(\hat{\beta}_n) + f(\hat{\beta}_n) - f(\hat{\beta})| \\ &> \frac{\eta}{2}. \end{aligned}$$

This is a contradiction and we get $\hat{\beta}_n \to \hat{\beta}$. □

Theorem 2.2 *Let $\hat{\beta}(\theta, D, X)$ be the maximum likelihood estimator of the following frailty model:*

$$\lambda_i(t|\gamma_i, \mathcal{F}_(t)) = \gamma_i e^{\beta x_i(t)},$$

where γ_i is a frailty with probability density function $f(\cdot|\theta)$ with mean 1 and variance θ. We use D and X denote the observed data and covariates.

Let $\hat{\beta}^*(D, X)$ be the maximum conditional likelihood estimator of SCCS model:

$$\lambda_i(t|\phi_i, \mathcal{F}(t)) = e^{\phi_i + \beta x_i(t)},$$

where ϕ_i is a constant individual baseline.

With the following conditions:

- $\beta \in \mathrm{B}$*, a compact set ,*
- $f(\gamma|\theta) = \frac{1}{\Gamma(\frac{1}{\theta})\theta^{\frac{1}{\theta}}} x^{\frac{1}{\theta}-1} e^{-\frac{1}{\theta}x}$,
- *The covariates $x_i(t)$ is an indicator function,*

we have

$$\lim_{\theta \to \infty} \hat{\beta}(\theta, D, X) \to \hat{\beta}^*(D, X).$$

Proof The likelihood function for frailty model is

$$\begin{aligned} L(\beta, \theta) &= \int_{R^N} \prod_{i=1}^{N} \left(\prod_{j=1}^{n_i} \gamma_i e^{\beta x_i(t_{ij})} \times \exp\left(-\gamma_i \int_{a_i}^{b_i} e^{\beta x_i(s)} ds \right) f(\gamma_i|\theta) d\gamma_i \right) \\ &= \int_{R^N} \prod_{i=1}^{N} \gamma_i^{n_i} \times \exp\left(\beta \sum_{i=1}^{N} \sum_{j=1}^{n_i} x_i(t_{ij}) - \sum_{i=1}^{N} \gamma_i (e^{\beta} T_{i1} + T_{i2}) \right) \times \prod_{i=1}^{N} f(\gamma_i|\theta) d\gamma_i. \end{aligned}$$

The MLE $\hat{\beta}_\theta$ satisfy the equation:

$$
\begin{aligned}
0 &= \frac{\partial L(\beta,\theta)}{\partial \beta} \\
&= \int_{R^N} \prod_{i=1}^{N} \gamma_i^{n_i} \times \exp\left(\beta \sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij}) - \sum_{i=1}^{N} \gamma_i (e^{\beta} T_{i1} + T_{i2})\right) \times \prod_{i=1}^{N} f(\gamma_i|\theta) d\gamma_i \\
&\quad \times \left(\sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij}) - \sum_{i=1}^{N} \gamma_i e^{\beta} T_{i1}\right) \\
&= \int_{R^N} \prod_{i=1}^{N} \gamma_i^{n_i} \times \exp\left(\beta \sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij}) - \sum_{i=1}^{N} \gamma_i (e^{\beta} T_{i1} + T_{i2})\right) \times \prod_{i=1}^{N} f(\gamma_i|\theta) d\gamma_i \\
&\quad \times \left(\sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij})\right) \\
&\quad - \int_{R^N} \prod_{i=1}^{N} \gamma_i^{n_i} \times \exp\left(\beta \sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij}) - \sum_{i=1}^{N} \gamma_i (e^{\beta} T_{i1} + T_{i2})\right) \times \prod_{i=1}^{N} f(\gamma_i|\theta) d\gamma_i \\
&\quad \times \sum_{i=1}^{N} \gamma_i e^{\beta} T_{i1},
\end{aligned}
$$

which is equivalent to

$$
\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij}) &= \frac{\int_{R^N} \prod_{i=1}^{N} \gamma_i^{n_i} \times \exp\left(-\sum_{i=1}^{N} \gamma_i (e^{\beta} T_{i1} + T_{i2})\right) \times \prod_{i=1}^{N} f(\gamma_i|\theta) d\gamma_i \times \left(\sum_{i=1}^{N} \gamma_i e^{\beta} T_{i1}\right)}{\int_{R^N} \prod_{i=1}^{N} \gamma_i^{n_i} \times \exp\left(-\sum_{i=1}^{N} \gamma_i (e^{\beta} T_{i1} + T_{i2})\right) \times \prod_{i=1}^{N} f(\gamma_i|\theta) d\gamma_i} \\
&= \sum_{i=1}^{N} \frac{\int_R \gamma_i^{n_i} \times \exp\left(-\gamma_i (e^{\beta} T_{i1} + T_{i2})\right) \times f(\gamma_i|\theta) d\gamma_i \times (\gamma_i e^{\beta} T_{i1})}{\int_R \gamma_i^{n_i} \times \exp\left(-\gamma_i (e^{\beta} T_{i1} + T_{i2})\right) \times f(\gamma_i|\theta) d\gamma_i}.
\end{aligned}
\tag{2.8}
$$

The density function of gamma distribution is

$$
\begin{aligned}
f(x|\theta) &= \frac{1}{\Gamma(\frac{1}{\theta})\theta^{\frac{1}{\theta}}} x^{\frac{1}{\theta}-1} e^{-\frac{1}{\theta}x}, \\
Var &= \theta.
\end{aligned}
$$

Then equation (2.8) becomes

$$
\begin{aligned}
0 = g_\theta(\beta) &= \sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij}) - \sum_{i=1}^{N} \frac{\int_R \gamma_i^{n_i} \times \exp\left(-\gamma_i(e^\beta T_{i1} + T_{i2})\right) \times \gamma_i^{\frac{1}{\theta}-1} e^{-\frac{1}{\theta}\gamma_i} d\gamma_i \times (\gamma_i e^\beta T_{i1})}{\int_R \gamma_i^{n_i} \times \exp\left(-\gamma_i(e^\beta T_{i1} + T_{i2})\right) \times \gamma_i^{\frac{1}{\theta}-1} e^{-\frac{1}{\theta}\gamma_i} d\gamma_i} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij}) - \sum_{i=1}^{N} e^\beta T_{i1} \frac{\int_R \gamma_i^{n_i+1+\frac{1}{\theta}-1} \times \exp\left(-\gamma_i(e^\beta T_{i1} + T_{i2} + \frac{1}{\theta})\right) d\gamma_i}{\int_R \gamma_i^{n_i+\frac{1}{\theta}-1} \times \exp\left(-\gamma_i(e^\beta T_{i1} + T_{i2} + \frac{1}{\theta})\right) d\gamma_i} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij}) - \sum_{i=1}^{N} e^\beta T_{i1} \frac{\Gamma(n_i + 1 + \frac{1}{\theta})/(e^\beta T_{i1} + T_{i2} + \frac{1}{\theta})^{n_i+1+\frac{1}{\theta}}}{\Gamma(n_i + \frac{1}{\theta})/(e^\beta T_{i1} + T_{i2} + \frac{1}{\theta})^{n_i+\frac{1}{\theta}}} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij}) - \sum_{i=1}^{N} e^\beta T_{i1} \frac{n_i + \frac{1}{\theta}}{e^\beta T_{i1} + T_{i2} + \frac{1}{\theta}}.
\end{aligned}
$$

Clearly, by the compactness of the domain of β, $\frac{n_i+\frac{1}{\theta}}{e^\beta T_{i1}+T_{i2}+\frac{1}{\theta}}$ converges uniformly to $\frac{n_i}{e^\beta T_{i1}+T_{i2}}$ as $\theta \to \infty$. And then,

$$\lim_{\theta\to\infty} g_\theta(\beta) = g(\beta) \qquad \text{Uniformly,}$$

where

$$
\begin{aligned}
g_\theta(\hat{\beta}_\theta) &= 0, \\
g(\beta) &= \sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij}) - \sum_{i=1}^{N} e^\beta T_{i1} \frac{n_i}{e^\beta T_{i1} + T_{i2}}.
\end{aligned}
$$

We come to the maximum condition likelihood estimator $\hat{\beta}^*$ of the SCCS method and the conditional log likelihood is

$$
\begin{aligned}
\log L_c &= \log \prod_{i=1}^{N}\prod_{j=1}^{n_i} \left(\frac{e^{\beta x_i(t_{ij})}}{\int_{a_i}^{b_i} e^{\beta x_i(s)} ds} \right) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{n_i} \left(\beta x_i(t_{ij}) - \log\left(e^\beta T_{i1} + T_{i2}\right) \right) \\
&= \beta \sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij}) - \sum_{i=1}^{N} n_i \log\left(e^\beta T_{i1} + T_{i2}\right).
\end{aligned}
$$

Then taking derivative with respect to β, we have

$$\sum_{i=1}^{N}\sum_{j=1}^{n_i} x_i(t_{ij}) = \sum_{i=1}^{N} \frac{n_i e^{\beta} T_{i1}}{e^{\beta} T_{i1} + T_{i2}}. \tag{2.9}$$

This can lead to $g(\hat{\beta}^*) = 0$. By Lemma 2.2, we can conclude that

$$\lim_{\theta \to \infty} \hat{\beta}(\theta, D, X) \to \hat{\beta}^*(D, X).$$

□

Chapter 3: Theoretical Results for Generalized Random Effect Model in Multivariate Counting Processes

3.1 Introduction

In chapter 2, the multivariate correlated frailty help explain the heterogeneity in LODs and construct correlation structure between multiple adverse events. Besides, the frailty, also called random effect, can be dynamic along the process of event occurrences. This design is motivated by the clinical settings in postmarketing surveillance. The random effect can be treated as hidden confounding factors which affect the risk of event occurrences. On the other hand, the event occurrences can affect these hidden confounding factors. For example, if an individual is diagnosed as MI, he is not likely to go to the hospital for another diagnose of MI in the next few days. In addition, if we treat drug prescription as an event and an individual is prescribed ACE Inhibitors for 30 days, the another prescription is not likely to occur in the next 30 days. Inspired by these ideas, we consider the dynamic random effects.

In this chapter, we propose a generalized random effect model: dynamic random effect model for multivariate counting process. Let N, J, L_1, L_2 denote the number of individuals, event types of interest, fix effect covariates and random effect covariates. The data consist of $\{\boldsymbol{N}_i(t), \boldsymbol{Y}_i(t), \boldsymbol{X}_i(t), \boldsymbol{Z}_i(t) : t \in [0, \tau_i], i = 1, 2, \cdots, N\}$. Let τ denote the length of observation period. $\boldsymbol{N}_i(t) = (N_{i1}(t), N_{i2}(t), \cdots, N_{iJ}(t))^T$ is the occurring number of events at time t and $\boldsymbol{Y}_i(t) = (Y_{i1}(t), Y_{i2}(t), \cdots, Y_{iJ}(t))^T$ is at risk status defined by $Y_{ij}(t) = \mathcal{I}(C_{ij} \geq t)$, where C_{ij} is the right-censoring time. Then we have intensity for counting process $N_{ij}(t)$

$$\lambda_{ij}(t|\phi_{ij}, \mathcal{F}_t) = Y_{ij}(t)\lambda_{0j}(t)e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_{ij}(t) + \boldsymbol{\eta}_j^T \boldsymbol{Z}_{ij}(t)\phi_{ij}},$$

where the random effect $\boldsymbol{\phi}_i = (\phi_{i1}, \phi_{i2}, \cdots, \phi_{iJ}(t))^T$ follows a J dimensional Gaussian distribution.

The covariate vector $\boldsymbol{X}_{ij}(t) \in \mathcal{R}^{L_1 \times 1}$ includes both interval and external covariates. It can be age, gender, drug exposures and functions of previous event history. The random effect covariate vector $\boldsymbol{Z}_{ij}(t) \in \mathcal{R}^{L_2 \times 1}$ denotes the change of random effect with respect to time t. To guarantee the identifiability property, additional assumptions on $\boldsymbol{\eta}$ and $\boldsymbol{Z}_{ij}(t)$ are necessary.

Based on above notations and discussion, the likelihood function for parameters $\theta = \{\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Sigma}\}$ and $\mathcal{A} = \{\Lambda_{10}, \Lambda_{20}, \cdots, \Lambda_{j0}\}$ is

$$
\begin{aligned}
L(\theta, \mathcal{A}; D) = \prod_{i=1}^{N} \int_{\boldsymbol{\phi}_i} \prod_{j=1}^{J} \Bigg[\prod_{t \le \tau} \left\{ Y_{ij}(t) \lambda_{j0}(t) e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_{ij}(t) + \boldsymbol{\eta}_j^T \boldsymbol{Z}_{ij}(t) \phi_{ij}} \right\}^{dN_{ij}(t)} \\
\times \exp \left\{ - \int_0^{\tau} Y_{ij}(t) e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_{ij}(t) + \boldsymbol{\eta}_j^T \boldsymbol{Z}_{ij}(t) \phi_{ij}} d\Lambda_{j0}(t) \right\} \Bigg] \psi(\boldsymbol{\phi}_i; \boldsymbol{\Sigma}) d\boldsymbol{\phi}_i .
\end{aligned}
\tag{3.1}
$$

If we consider $\Lambda_{j0}(t)$ as absolute continuous functions, we can not reach a maximum of the likelihood. Thus, we restrict $\Lambda_{j0}(t)$ in step functions with jumps at the observed event times $\{t_{ijk}, i = 1, 2, \cdots, n; j = 1, 2, \cdots, J; k = 1, 2, \cdots, n_{ij}\}$ and the baseline intensity $\lambda_{j0}(t)$ becomes the jump size of $\Lambda_{j0}(t)$ at time t, denoted by $\Lambda_{j0}\{t\}$. The modified likelihood function becomes

$$
\begin{aligned}
L(\theta, \mathcal{A}; D) = \prod_{i=1}^{N} \int_{\boldsymbol{\phi}_i} \prod_{j=1}^{J} \Bigg[\prod_{t \le \tau} \left\{ Y_{ij}(t) \Lambda_{j0}\{t\} e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_{ij}(t) + \boldsymbol{\eta}_j^T \boldsymbol{Z}_{ij}(t) \phi_{ij}} \right\}^{dN_{ij}(t)} \\
\times \exp \left\{ - \int_0^{\tau} Y_{ij}(t) e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_{ij}(t) + \boldsymbol{\eta}_j^T \boldsymbol{Z}_{ij}(t) \phi_{ij}} d\Lambda_{j0}(t) \right\} \Bigg] \psi(\boldsymbol{\phi}_i; \boldsymbol{\Sigma}) d\boldsymbol{\phi}_i .
\end{aligned}
\tag{3.2}
$$

3.2 Theoretical Results

In this section, we develop a series of asymptotic theories under some regularity conditions. Actually, the nonparametric maximum likelihood estimators are shown to be identifiable, consistent and asymptotic normal with covariance matrix that can be consistently estimated by the inverse information matrix. Letting $\left(\theta_0, \mathcal{A}_0 = \{\Lambda_{0,10}, \Lambda_{0,20}, \cdots, \Lambda_{0,J0}\}\right)$ be the true value of $(\theta, \mathcal{A})$, we impose the following conditions on the model and data structures.

(A 1) The true value θ_0 lies in the interior of a compact set Θ and the true functions $\Lambda_{0,j0}$ are

continuously differentiable in $[0, \tau]$ with $\Lambda'_{0,j0}(t) > 0,\ j = 1, 2, \cdots, J$.

(A 2) $\boldsymbol{X}_{ij}(t), \boldsymbol{Z}_{ij}(t)$ are almost surely bounded and left-continuous and have bounded left-derivatives and right-derivatives in $[0, \tau]$.

(A 3) With probability one, $\mathrm{P}(\cup_j \{C_{ij} \geq \tau\} | \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}) > \delta_0 > 0$. This guarantees that there is a positive expectation of number of events to be observed before the individual is censored.

(A 4) If there exists a vector ξ and a deterministic function $h(t)$ such that $h(t) + \xi^T \boldsymbol{X}_i(t) = 0$ almost surely, then we have $h(t) = 0$ and $\xi = 0$. For each $j, l = 1, \cdots, J$, if there exists a matrix B such that $Z_j^T(t) B Z_l(s) = 0$ almost surely, then $B = 0$. This guarantees the identifiability of the model.

(A 5) There exists J non-zero numbers $\{r_j\}_{j=1}^J$ such that $\eta_{j1} = r_j$ for $j = 1, \cdots, J$.

Actually, conditions (A1) and (A2) are standard assumptions of regression models in event history analysis. Condition (A3) shows that there exists some events at risk for each individual during the observational period, which is a standard assumption on the analysis of censored data. The identifiability can be derived from condition (A4) and (A5). In particular, condition (A5) is necessary when the covariance matrix in random effect is unrestricted.

We first give the identifiability of the model.

Theorem 3.1 (Identifiability) *Under assumptions (A1)-(A4), the model is identifiable. That is, if*

$$L(\theta, \mathcal{A}; D) = L(\theta', \mathcal{A}'; D), \tag{3.3}$$

for all possible $D = \{X(t), Z(t), Y_j(t), N_j(t), j = 1, 2, \cdots, J;\ t \in [0, \tau]\}$, *then* $(\theta, \mathcal{A}) = (\theta', \mathcal{A}')$.

Proof Recall the likelihood

$$L(\theta, \mathcal{A}; D) = \int_\phi \left[\prod_{j=1}^J \prod_{t \leq \tau} \lambda_{ij}(t|\theta, \mathcal{A}, \phi)^{dN_{ij}} e^{-\int_0^\tau \lambda_{ij}(s|\theta, \mathcal{A}, \phi) ds} \right] \psi(\phi; \Sigma) d\phi.$$

Choose two integers m, m' and event times $\{t_{j1}, t_{j2}, \cdots, t_{j,m+m'}\}_{j=1}^{J}$, the likelihood becomes

$$L(\theta, \mathcal{A}; D) = \int_{\phi} \left[\prod_{j=1}^{J} \prod_{k=1}^{m} \lambda_{ij}(t_{jk}|\theta, \mathcal{A}, \phi) \prod_{k=m+1}^{m+m'} \lambda_{ij}(t_{jk}|\theta, \mathcal{A}, \phi) e^{-\int_0^\tau \lambda_{ij}(s|\theta,\mathcal{A},\phi)ds} \right] \psi(\phi; \Sigma) d\phi.$$

With ideas of Taylor's series of exponential function, we integrate $t_{j1}, \cdots, t_{jm}$ for 0 to t_j and $t_{j,m+1}, \cdots, t_{j,m+m'}$ from 0 to τ. Then we multiple the equation by $\prod_{j=1}^{J} \frac{(is_j)^m}{m!} \frac{1}{m'!}$ and sum over $m = 0, 1, 2, \cdots$ and $m' = 0, 1, 2, \cdots$. The equation (3.3) becomes

$$\int_{\phi} \prod_{j=1}^{J} e^{is_j \int_0^{t_j} \lambda_j(s|\theta,\mathcal{A},\phi)ds} \psi(\phi; \Sigma) d\phi = \int_{\phi'} \prod_{j=1}^{J} e^{is_j \int_0^{t_j} \lambda_j(s|\theta',\mathcal{A}',\phi')ds} \psi(\phi'; \Sigma') d\phi'.$$

This holds for arbitrary s_j and t_j, then by the property of characteristic functions, $\{\lambda_j(s_j|\theta, \mathcal{A}, \phi)\}_{j=1}^{J}$ has the same distribution as $\{\lambda_j(s_j|\theta', \mathcal{A}', \phi')\}_{j=1}^{J}$ where $\phi \sim N(0, \Sigma)$ and $\phi' \sim N(0, \Sigma')$. We take log operation and mean operation on both sides, we have

$$\log \lambda_{j0}(s_j) + \beta_j^T X_j(s_j) = \log \lambda'_{j0}(s_j) + \beta_j'^T X_j(s_j).$$

By condition (A4), we have $\lambda_{j0} = \lambda'_{j0}$ and $\beta_j = \beta'_j$. Subtracting from the above equation, the remaining part is

$$\begin{pmatrix} \eta_1^T Z_1(s_1)\phi_1 \\ \eta_2^T Z_2(s_2)\phi_2 \\ \vdots \\ \eta_J^T Z_J(s_J)\phi_J \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \eta_1'^T Z_1(s_1)\phi_1 \\ \eta_2'^T Z_2(s_2)\phi_2 \\ \vdots \\ \eta_J'^T Z_J(s_J)\phi_J \end{pmatrix}.$$

Let $\left\{\Sigma_{jl}\right\}_{jl=1,\cdots,J}$ be elements of Σ and $\left\{\Sigma'_{jl}\right\}_{jl=1,\cdots,J}$ be elements of Σ'. Considering the covariance, for each $j, l = 1, \cdots, J$, we have

$$\begin{aligned} & Z_j^T(s_j)\eta_j \Sigma_{jl} \eta_l^T Z_l(s_l) = Z_j^T(s_j)\eta'_j \Sigma_{jl'} \eta_l'^T Z_l(s_l) \\ \Longrightarrow\ & Z_j^T(s_j) \left(\eta_j \Sigma_{jl} \eta_l^T - \eta'_j \Sigma'_{jl} \eta_l'^T \right) Z_l(s_l) = 0. \end{aligned}$$

Based on condition (A4), we have $\eta_j \Sigma_{jl} \eta_l^T - \eta'_j \Sigma'_{jl} \eta_l'^T = 0$. Recall condition (A5), we have non-zero numbers r_j such that $\eta_{j1} = \eta'_{j1} = r_j$. This leads to $r_j r_l \Sigma_{jl} = r_j r_l \Sigma'_{jl}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}'$. For identifiability of η, we have $\eta_j \eta_l^T = \eta'_j \eta_l'^T$. Consider the first row on both sides, we have

$$(\eta_{j1}\eta_{l1}, \cdots, \eta_{j1}\eta_{lJ}) = \left(\eta'_{j1}\eta'_{l1}, \cdots, \eta'_{j1}\eta'_{lJ}\right),$$

which results in $\boldsymbol{\eta}_l = \boldsymbol{\eta}'_l$. □

The identifiability property is the foundation of model estimation and extremely important in random effect model. The theorems about consistency and asymptotic property are based on the results in [37]. We need to show conditions (A1)-(A5) are sufficient for conditions (C1)-(C8) in [37].

There are some notations we need to mention first:

- $\gamma_j = (\boldsymbol{\beta}_j, \boldsymbol{\eta}_j)$.
- O_i denotes the observed data of individual i.
- We rewrite the likelihood

$$L(\boldsymbol{\theta}, \mathcal{A}; D) = \prod_{i=1}^{N} \left[\prod_{j=1}^{J} \prod_{t \le \tau} \Lambda_{j0}\{t\}^{Y_{ij}(t) dN_{ij}(t)} \right] \Psi(O_i; \boldsymbol{\theta}, \mathcal{A}),$$

$$\Psi(O_i; \boldsymbol{\theta}, \mathcal{A}) = \int_{\phi_i} \prod_{j=1}^{J} \Omega_{ij}(\phi_i; \gamma_j, \Lambda_{j0}) \psi(\phi_i; \boldsymbol{\Sigma}) d\phi_i,$$

$$\Omega_{ij}(\phi_i; \gamma_j, \Lambda_{j0}) = \prod_{t \le \tau} \left\{ Y_{ij}(t) e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_{ij}(t) + \boldsymbol{\eta}_j^T \mathbf{Z}_{ij}(t) \phi_{ij}} \right\}^{dN_{ij}(t)} e^{-q_{ij}(\tau)},$$

$$q_{ij}(t) = \int_0^t Y_{ij}(s) e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_{ij}(s) + \boldsymbol{\eta}_j^T \mathbf{Z}_{ij}(s) \phi_{ij}} d\Lambda_{j0}(s).$$

- We define derivatives of Ψ with respect to $\boldsymbol{\theta}$ and Λ_{j0}.
  - $\dot{\Psi}_{\boldsymbol{\theta}}$ denotes the derivative of Ψ with respect to $\boldsymbol{\theta}$.
  - H_j belongs to the function set where $\Lambda_{j0} + \epsilon H_j$ is increasing with bounded total variation. Let $\dot{\Psi}_j$ denote the derivative of Ψ along the path $(\Lambda_{j0} + \epsilon H_j)$ with formula

$$\dot{\Psi}_j[H_j] = \lim_{\epsilon \to 0} \frac{\Psi(O_i; \boldsymbol{\theta}, \mathcal{A} + (0, \cdots, \epsilon H_j, \cdots, 0)) - \Psi(O_i; \boldsymbol{\theta}, \mathcal{A})}{\epsilon}.$$

The following eight propositions are the conditions (C1)-(C8) listed in [37] with our settings and notations. We need to prove that these conditions can be reached based on our conditions (A1)-(A5).

Proposition 3.1 *The true value θ_0 lies in the interior of a compact set Θ and the true functions $\Lambda_{0,j0}$ are continuously differentiable in $[0, \tau]$ with $\Lambda'_{0,j0}(t) > 0,\ j = 1, 2, \cdots, J$.*

Proof This is the same as condition (A1). □

Proposition 3.2 *With probability one, $P(\inf_{s \in [0,t]} Y_{ij}(s) \geq 1 | X_{ij}, Y_{ij}) > \delta_0 > 0$ for all $t \in [0, \tau]$.*

Proof This can be derived directly from condition (A3). □

Proposition 3.3 *There exist a constant $c_1 > 0$ and a random variable $r_1(O_i) > 0$ such that $\mathrm{E}(\log r_1(O_i)) < \infty$ and for any $\theta \in \Theta$ and any finite $\Lambda_1, \cdots, \Lambda_J$,*

$$\Psi(O_i; \theta, \mathcal{A}) \leq r(O_i) \prod_{j=1}^{J} \prod_{t \leq \tau} \left\{ 1 + \int_0^t Y_{ij}(t) d\Lambda_j(t) \right\}^{-dN_{ij}(t)} \left\{ 1 + \int_0^\tau Y_{ij}(t) d\Lambda_j(t) \right\}^{-c_1}, \quad (3.4)$$

almost surely. In addition, for any constant c_2,

$$\inf \left\{ \Psi(O_i, \theta, \mathcal{A}) : \|\Lambda_1\|_{V[0,\tau]} \leq c_2, \cdots, \|\Lambda_J\|_{V[0,\tau]} \leq c_2, \theta \in \Theta \right\} > r_2(O_i) > 0, \quad (3.5)$$

where $r_2(O_i)$, which may depend on c_2, is a finite random variable with $\mathrm{E}(|\log r_2(O_i)|) < \infty$ and $\|h\|_{V[0,\tau]}$ is the total variation of $h(\cdot)$ in $[0, \tau]$.

Proof Recall that

$$\Psi(O_i;\boldsymbol{\theta},\mathcal{A}) = \int_{\phi_i}\prod_{j=1}^{J}\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0})\psi(\phi_i;\boldsymbol{\Sigma})d\phi_i,$$

$$\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0}) = \prod_{t\le\tau}\left\{Y_{ij}(t)e^{\boldsymbol{\beta}_j^T\boldsymbol{X}_{ij}(t)+\boldsymbol{\eta}_j^T\boldsymbol{Z}_{ij}(t)\phi_{ij}}\right\}^{dN_{ij}(t)}e^{-q_{ij}(\tau)},$$

$$q_{ij}(t) = \int_0^t Y_{ij}(s)e^{\boldsymbol{\beta}_j^T\boldsymbol{X}_{ij}(s)+\boldsymbol{\eta}_j^T\boldsymbol{Z}_{ij}(s)\phi_{ij}}d\Lambda_{j0}(s).$$

Due to the boundedness of parameters $\boldsymbol{\beta},\boldsymbol{\eta}$ and covariates $\boldsymbol{X}_{ij},\boldsymbol{Z}_{ij}$, we have

$$e^{-O(1)(1+|\phi_i|)} \le e^{\boldsymbol{\beta}_j^T\boldsymbol{X}_{ij}(s)+\boldsymbol{\eta}_j^T\boldsymbol{Z}_{ij}(s)\phi_{ij}} \le e^{O(1)(1+|\phi_i|)},$$

$$e^{-O(1)(1+|\phi_i|)N_{ij}(\tau)}e^{-q_{ij}(\tau)} \le \Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0}) \le e^{O(1)(1+|\phi_i|)N_{ij}(\tau)}e^{-q_{ij}(\tau)},$$

$$e^{-O(1)(1+|\phi_i|)}\left\{1+\int_0^t Y_{ij}(s)d\Lambda_j(s)\right\} \le 1+q_{ij}(t) \le e^{O(1)(1+|\phi_i|)}\left\{1+\int_0^t Y_{ij}(s)d\Lambda_j(s)\right\}.$$

For the second part of Proposition 3.3, choose a positive constant B_0 such that $P(|\boldsymbol{\phi}| \le B_0) > 0$. Then for bounded $\|\Lambda_j\|_{V[0,\tau]}$, we have

$$\begin{aligned}\Psi(O_i;\boldsymbol{\theta},\mathcal{A}) &= \int_{\phi_i}\prod_{j=1}^{J}\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0})\psi(\phi_i;\boldsymbol{\Sigma})d\phi_i\\ &\ge \int_{|\phi_i|\le B_0}\prod_{j=1}^{J}\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0})\psi(\phi_i;\boldsymbol{\Sigma})d\phi_i\\ &\ge e^{-O(1)\sum_{j=1}^{J}N_{ij}(\tau)}P(|\phi_i|\le B_0).\end{aligned}$$

We choose $r_2(O_i) = e^{-O(1)\sum_{j=1}^{J}N_{ij}(\tau)}P(|\phi_i|\le B_0) > 0$ and inequality (3.5) is satisfied.

For inequality (3.4), we note that for any positive integer m and $0 < x_1 \le \cdots \le x_m \le y$, there

exist $\mu > 0$ and $\kappa > 0$ such that

$$e^{-y}\prod_{i=1}^{m}(1+x_i) \le \mu^m(1+y)^{-\kappa}.$$

We replace $y = q_{ij}(\tau)$, $m = N_{ij}(\tau)$ and x to be $1 + q_{ij}(t)$ at jumping times. We have

$$\begin{aligned}
&\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0}) \\
&\le e^{O(1)(1+|\phi_i|)N_{ij}(\tau)}e^{-q_{ij}(\tau)} \\
&\le e^{O(1)(1+|\phi_i|)N_{ij}(\tau)}\mu^{N_{ij}(\tau)}\prod_{t\le\tau}(1+q_{ij}(t))^{-dN_{ij}(t)}(1+q_{ij}(\tau))^{-\kappa} \\
&\le e^{O(1)(1+|\phi_i|)N_{ij}(\tau)}\mu^{N_{ij}(\tau)}\prod_{t\le\tau}\left(1+\int_0^t Y_{ij}(s)d\Lambda_j(s)\right)^{-dN_{ij}(t)}\left(1+\int_0^t Y_{ij}(s)d\Lambda_j(s)\right)^{-\kappa}.
\end{aligned}$$

We choose $r_1(O_i) = \mu^{\sum_{j=1}^J N_{ij}(\tau)}\int e^{O(1)(1+|\phi_i|)\sum_{j=1}^J N_{ij}(\tau)}\psi(\phi_i)d\phi_i$ in which the inequality (3.4) holds. □

The next proposition shows some smoothness of Ψ. We denote $\dot{\Psi}$ as the derivative of $\Psi(O_i;\theta,\mathcal{A})$ with respect to θ and $\dot{\Psi}_k[H_j]$ as the derivative of $\Psi(O_i;\theta,\mathcal{A})$ along the path $(\Lambda_j + \epsilon H_j)$, where H_j is a function s.t $\Lambda_j + \epsilon H_j$ is increasing and has bounded total variation.

Proposition 3.4 *For any* $(\theta^{(1)},\theta^{(2)}) \in \Theta$, $(\Lambda_1^{(1)},\Lambda_1^{(2)}),\cdots,(\Lambda_J^{(1)},\Lambda_J^{(2)})$ *and* $(H_1^{(1)},H_1^{(2)}),\cdots,$ $(H_J^{(1)},H_J^{(2)})$ *with uniformly bounded total variations, there exist a random variable* $\mathcal{F}(O_i) \in$

$L_4(P)$ *and* J *stochastic processes* $\mu_{ij}(t; O_i) \in L_6(P)$, $j = 1, \cdots, J$ *such that*

$$
\begin{aligned}
&\left|\Psi(O_i;\theta^{(1)},\mathcal{A}^{(1)}) - \Psi(O_i;\theta^{(2)},\mathcal{A}^{(2)})| + |\dot{\Psi}_\theta(O_i;\theta^{(1)},\mathcal{A}^{(1)}) - \dot{\Psi}_\theta(O_i;\theta^{(2)},\mathcal{A}^{(2)})\right| \\
&+ \sum_{j=1}^{J}\left|\dot{\Psi}_j(O_i;\theta^{(1)},\mathcal{A}^{(1)})[H_j^{(1)}] - \dot{\Psi}_j(O_i;\theta^{(2)},\mathcal{A}^{(2)})[H_j^{(1)}]\right| \\
&+ \sum_{j=1}^{J}\left|\frac{\dot{\Psi}_j(O_i;\theta^{(1)},\mathcal{A}^{(1)})[H_j^{(1)}]}{\Psi(O_i;\theta^{(1)},\mathcal{A}^{(1)})} - \frac{\dot{\Psi}_j(O_i;\theta^{(2)},\mathcal{A}^{(2)})[H_j^{(1)}]}{\Psi(O_i;\theta^{(2)},\mathcal{A}^{(2)})}\right| \qquad (3.6) \\
&\le \mathcal{F}(O_i)\left[|\theta^{(1)} - \theta^{(2)}| + \sum_{j=1}^{J}\left\{\int_0^\tau |\Lambda_j^{(1)}(s) - \Lambda_j^{(2)}(s)| d\mu_{ij}(s;O_i)\right.\right. \\
&\quad \left.\left. + \int_0^\tau |H_j^{(1)}(s) - H_j^{(2)}(s)| d\mu_{ij}(s;O_i)\right\}\right].
\end{aligned}
$$

In addition, $\mu_{ij}(s; O_i)$ *is non-decreasing and* $\mathrm{E}[\mathcal{F}(O_i)\mu_{ij}(s; O_i)]$ *is left-continuous with uniformly bounded left- and right-derivatives for any* $s \in [0, \tau]$.

Proof Recall that

$$
\Omega_{ij}(\phi_i; \gamma_j, \Lambda_{j0}) \le e^{O(1)(1+|\phi_i|)N_{ij}(\tau)} e^{-q_{ij}(\tau)} \le e^{O(1)(1+|\phi_i|)N_{ij}(\tau)}.
$$

Besides, the derivatives are satisfied in the following conditions

$$\begin{aligned}\left|\frac{\partial}{\partial\beta_j}\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0})\right| &= \left|\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0})\left\{\int Y_{ij}(t)X_{ij}(t)dN_{ij}(t)\right.\right.\\ &\qquad \left.\left.-\int_0^\tau Y_{ij}(t)e^{\beta_j^T X_{ij}(t)+\eta_j Z_{ij}(t)\phi_{ij}}X_{ij}(t)d\Lambda_j(t)\right\}\right|\\ &\le e^{O(1)(1+|\phi_i|)(1+N_{ij}(\tau))},\end{aligned}$$

$$\left|\frac{\partial}{\partial\eta_j}\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0})\right| \le e^{O(1)(1+|\phi_i|)(1+N_{ij}(\tau))},$$

$$\begin{aligned}\left|\frac{\partial}{\partial\Lambda_j}\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0})[H_j]\right| &= \left|-\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0})\int_0^\tau Y_{ij}(t)e^{\beta_j^T X_{ij}(t)+\eta_j Z_{ij}(t)\phi_{ij}}X_{ij}(t)dH_j(t)\right|\\ &\le e^{O(1)(1+|\phi_i|)(1+N_{ij}(\tau))}.\end{aligned}$$

With these bounds of derivative and Mean Value Theorem, we have

$$
\begin{aligned}
&\left|\Omega_{ij}(\phi_i;\gamma_j^{(1)},\Lambda_{j0}) - \Omega_{ij}(\phi_i;\gamma_j^{(2)},\Lambda_{j0})\right| \\
&= \left|\frac{\partial}{\partial\gamma_j}\Omega_{ij}(\phi_i;\gamma_j^{*},\Lambda_{j0})\right|\left|\gamma_j^{(1)} - \gamma_j^{(2)}\right| \\
&\le e^{O(1)(1+|\phi_i|)(1+N_{ij}(\tau))}\left|\gamma_j^{(1)} - \gamma_j^{(2)}\right|,
\end{aligned}
$$

$$
\begin{aligned}
&\left|\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0}^{(1)}) - \Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0}^{(2)})\right| \\
&= \left|\frac{\partial}{\partial\Lambda_{j0}}\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0}^{*})\right|\left|\Lambda_{j0}^{(1)} - \Lambda_{j0}^{(2)}\right| \\
&\le e^{O(1)(1+|\phi_i|)N_{ij}(\tau)}\left|\int_0^\tau Y_{ij}(t)e^{\beta_j^T X_{ij}(t)+\eta_j Z_{ij}(t)\phi_{ij}}X_{ij}(t)d(\Lambda_{j0}^{(1)} - \Lambda_{j0}^{(2)})(t)\right| \\
&\le e^{O(1)(1+|\phi_i|)(1+N_{ij}(\tau))}\int_0^\tau\left|\Lambda_{j0}^{(1)}(t) - \Lambda_{j0}^{(2)}(t)\right|dt,
\end{aligned}
$$

$$
\begin{aligned}
&\left|\int_{\phi_i}\prod_{j=1}^{J}\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0})\psi(\phi_i;\Sigma^{(1)})d\phi_i - \int_{\phi_i}\prod_{j=1}^{J}\Omega_{ij}(\phi_i;\gamma_j,\Lambda_{j0})\psi(\phi_i;\Sigma^{(2)})d\phi_i\right| \\
&\le \left|\int e^{O(1)(1+|\phi_i|)\sum_{j=1}^{J}N_{ij}(\tau)}\frac{\partial\psi(\phi_i;\Sigma^{*})}{\partial\Sigma}\right|\left|\Sigma^{(1)} - \Sigma^{(2)}\right|.
\end{aligned}
$$

The above inequalities use integration by parts and boundedness of covariates X_{ij} and Z_{ij}. We can verify that the other three terms are satisfied the bounds with the same arguments.

□

The condition requires the smoothness of Ψ and implies that the linear functional

$$
H_j \mapsto \mathrm{E}\left[\frac{\dot{\Psi}_j(O_i;\theta,\mathcal{A})[H_j]}{\Psi(O_i;\theta,\mathcal{A})}\right] \tag{3.7}
$$

is continuous from $BV[0,\tau]$ to $\mathcal{R}$. Thus there exists a bounded function $\nu_{0j}(s;\theta,\mathcal{A})$ such that

$$
\mathrm{E}\left[\frac{\dot{\Psi}_j(O_i;\theta,\mathcal{A})[H_j]}{\Psi(O_i;\theta,\mathcal{A})}\right] = \int_0^\tau \nu_{0j}(s;\theta,\mathcal{A})dH_j(s), \tag{3.8}
$$

which will be used in Proposition 3.6.

Proposition 3.5 *The model is identifiable. That is, if*

$$L(\theta, \mathcal{A}; D) = L(\theta', \mathcal{A}'; D), \tag{3.9}$$

for all possible $D = \{X(t), Z(t), Y_j(t), N_j(t), j = 1, 2, \cdots, J;\ t \in [0, \tau]\}$*, then* $(\theta, \mathcal{A}) = (\theta', \mathcal{A}')$.

Proof This is proved in Theorem 3.1. □

Proposition 3.6 *There exist function* $\zeta_{0j}(s; \theta_0, \mathcal{A}_0) \in BV[0, \tau]$, $j = 1, \cdots, J$ *and a matrix* $\zeta_{0\theta}(\theta_0, \mathcal{A}_0)$ *such that*

$$\begin{aligned} &\left| \mathrm{E}\left[\frac{\dot{\Psi}_\theta(O_i; \theta, \mathcal{A})}{\Psi(O_i; \theta, \mathcal{A})} - \frac{\dot{\Psi}_\theta(O_i; \theta_0, \mathcal{A}_0)}{\Psi(O_i; \theta_0, \mathcal{A}_0)} \right] - \zeta_{0\theta}(\theta_0, \mathcal{A}_0)(\theta - \theta_0) \right. \\ &\left. - \sum_{j=1}^{J} \int_0^\tau \zeta_{0j}(s; \theta_0, \mathcal{A}_0) d(\Lambda_j - \Lambda_{0j}) \right| = o\left(|\theta - \theta_0| + \sum_{j=1}^{J} \|\Lambda_j - \Lambda_{0j}\|_{V[0,\tau]} \right). \end{aligned} \tag{3.10}$$

In addition, for $j = 1, \cdots, J$

$$\begin{aligned} &\sum_{j=1}^{J} \sup_{s \in [0,\tau]} \left| \left\{ \nu_{0j}(s; \theta, \mathcal{A}) - \nu_{0j}(s; \theta_0, \mathcal{A}_0) \right\} - \nu_{0j\theta}(s; \theta_0, \mathcal{A}_0)(\theta - \theta_0) \right. \\ &\left. - \int_0^\tau \sum_{m=1}^{J} \nu_{0jm}(s, t; \theta_0, \mathcal{A}_0) d(\Lambda_m - \Lambda_{0m})(t) \right| = o\left(|\theta - \theta_0| + \sum_{j=1}^{J} \|\Lambda_j - \Lambda_{0j}\|_{V[0,\tau]} \right), \end{aligned} \tag{3.11}$$

where ν_{0jm} *is a bounded bivariate function and* $\nu_{0j\theta}$ *is a* $d-$*dimensional bounded function. Furthermore, there exists a constant* c_3 *such that*

$$|\nu_{0jm}(s, t_1; \theta_0, \mathcal{A}_0) - \nu_{0jm}(s, t_2; \theta_0, \mathcal{A}_0) \le c_3 |t_1 - t_2|,$$

for any $s \in [0, \tau]$ *and any* $t_1, t_2 \in [0, \tau]$.

Proof

We calculate from definition that

$$\nu_{0j}(s;\theta,\mathcal{A}) = -\mathrm{E}\left[\int_\phi \frac{\prod_{m=1}^J \Omega_{im}(\phi;\gamma_m,\Lambda_m)\psi(\phi;\Sigma)}{\int_\phi \prod_{m=1}^J \Omega_{im}(\phi;\gamma_m,\Lambda_m)\psi(\phi;\Sigma)d\phi} Y_{ij}(s)e^{\beta_j X_{ij}(s)+\eta_j Z_{ij}(s)\phi}d\phi\right].$$

For $(\theta,\mathcal{A})$ in a neighborhood of $(\theta_0,\mathcal{A}_0)$,

$$\left|\left\{\nu_{0j}(s;\theta,\mathcal{A}) - \nu_{0j}(s;\theta_0,\mathcal{A}_0)\right\} - \frac{\partial}{\partial\theta}\nu_{0j}(s;\theta_0,\mathcal{A}_0)(\theta-\theta_0) - \sum_{m=1}^J \frac{\partial}{\partial\Lambda_m}\nu_{0j}(s,t;\theta_0,\mathcal{A}_0)[\Lambda_m-\Lambda_{0m}]\right| = o\left(|\theta-\theta_0| + \sum_{j=1}^J \|\Lambda_j-\Lambda_{0j}\|_{V[0,\tau]}\right).$$

Define $\nu_{0j\theta} = \frac{\partial\nu_{0j}}{\partial\theta}$ and ν_{0jm} to be derivative of ν_{0j} with respect to Λ_m along the direction $\Lambda_m - \Lambda_{0m}$ and we reach equation (3.11). For equation (3.10), we can derive it by verifying the Lipschitz continuity of ν_{0km}. □

Proposition 3.7 ((Second Identifiability Condition)) *If for some $v \in \mathcal{R}^d$ and $h_j \in BV[0,\tau]$, $j = 1,\cdots,J$,*

$$\sum_{j=1}^J \int h_j(t)Y_{ij}(t)dN_{ij}(t) + \frac{\dot\Psi_\theta(O_i;\theta_0,\mathcal{A}_0)^T v + \sum_{j=1}^J \dot\Psi_j(O_i;\theta_0,\mathcal{A}_0)[\int h_j d\Lambda_{0j}]}{\Psi(O_i;\theta_0,\mathcal{A}_0)} = 0, \quad (3.12)$$

almost surely, then $v = 0$ and $h_j = 0$, $j = 1,\cdots,J$.

Proof Similar arguments in Proposition 3.5. □

This proposition shows that the Fisher information matrix along any finite-dimensional submodel is non-singular.

Proposition 3.8 *There exists a neighborhood of $(\theta_0,\mathcal{A}_0)$ such that for $(\theta,\mathcal{A})$ in this neighborhood, the first and second derivatives of $\log\Psi(O_i;\theta,\mathcal{A})$ with respect to θ and along the path $\Lambda_j + \epsilon H_j$ with respect to ϵ satisfy the inequality in (3.6) .*

Proof This is similar to Proposition 3.4 based on $\frac{\partial^2\Psi(O_i;\theta,\mathcal{A})}{\partial\theta^2}$ and the first and second derivatives of $\Psi(O_i,\theta,\mathcal{A}_0+\epsilon H)$ with respect to ϵ. □

Theorem 3.2 (Consistency) *Under assumptions (A1)-(A4),*

$$|\hat{\theta}_n - \theta_0| + \sum_{j=1}^{J} \sup_{t\in[0,\tau]} |\hat{\Lambda}_{nj} - \Lambda_{0j}| \xrightarrow{a.s.} 0.$$

Proof The entire proof is complex and we leave the technical details in Appendix. We gives a sketch of the proof here. The entire proof consists of three steps.

- **Step 1:** The non-parametric maximum likelihood estimators $\hat{\Lambda}_j(\tau) < \infty$ for $j = 1, \cdots, J$ when n is large. Proposition 3.3 shows the likelihood function is bounded by

$$\prod_{i=1}^{N} r(O_i) \prod_{j=1}^{J} \left\{1 + \int_0^\tau Y_{ij}(t) d\Lambda_j(t)\right\}^{-c_1}.$$

  If $\Lambda_j(\tau) = \infty$ for some j, the condition (A2) is violated.

- **Step 2:** Show $\limsup_n \hat{\Lambda}_j(\tau) < \infty$ almost surely.

- **Step 3:** As $\hat{\Lambda}_j$ is bounded and monotone, $\hat{\Lambda}_j$ becomes weakly compact. If we apply Helly's Selection Theorem, we can select a subsequence for any subsequence, s.t the selected subsequence point-wise converges to some monotone function Λ_j^*. Similar, we define that $\hat{\theta}$ converges to θ^*. If we can prove that $\Lambda_j^* = \Lambda_{0j}$ and $\theta^* = \theta_0$, the consistency will hold.

□

Before we state the asymptotic normality of the estimators, we need to clarify some notations. Let U be any set, the space $l^\infty(U)$ is the collection of all uniformly bounded, real functions on U. For any positive real numbers a, b, $BV[a, b]$ is the collection of real functions with bounded total variations on $[a, b]$. Besides, we define $\mathcal{V} = \{v \in \mathcal{R}^d; \|v\| \leq 1\}$ and $Q = \{h : \|h\|_{V[0,\tau]} \leq 1, h(0) = 0\}$, where $\|\cdot\|_{V[0,\tau]}$ is the total variation of function h in $[0, \tau]$. Actually, $\hat{\Lambda}_{nj}$, as a bounded linear functional operator, is an element in $l^\infty(Q)$ by $\hat{\Lambda}_{nj}(h) = \int_0^\tau h(s) d\hat{\Lambda}_{nj}(s)$ for $h \in Q$. With these notations, we can identify $(\hat{\theta}_n - \theta_0, \hat{\mathcal{A}}_n - \mathcal{A}_0)$ as a random element in $l^\infty(\mathcal{V} \times Q^J)$ through the definition $(\hat{\theta}_n - \theta_0)^T v + \sum_{j=1}^J \int_0^\tau h_j(s) d(\hat{\Lambda}_{n,j0} - \Lambda_{0,j0})(s)$ for some $v \in \mathcal{V}$ and $h_j \in Q$.

Theorem 3.3 (Asymptotically Normality) *Under assumptions (A1)-(A5), we have*

$$\sqrt{n}(\hat{\theta}_n - \theta_0, \hat{\mathcal{A}}_n - \mathcal{A}_0) \xrightarrow{d} \mathcal{G},$$

in $l^\infty(\mathcal{V} \times Q^J)$, where $\mathcal{G}$ is a continuous zero-mean Gaussian process. Furthermore, the limiting covariance matrix of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ attains the semi-parametric efficiency bound.

Theorem 3.3 states that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ and $\sqrt{n}(\hat{\mathcal{A}}_n - \mathcal{A}_0)$ are joint asymptotically normal in space $l^\infty(\mathcal{V} \times Q^J)$. For asymptotic variance, let I_n be observed information matrix of log-likelihood with respect to θ and $\{\Lambda_{j0}(t_{ijk}) : i = 1, 2, \cdots, n, j = 1, 2, \cdots, J, k = 1, 2, \cdots, n_{ij}\}$. The asymptotic variance of $\sqrt{n}\left\{(\hat{\theta}_n - \theta_0)^T v + \sum_{j=1}^J \int_0^\tau h_j(s) d(\hat{\Lambda}_{j0} - \Lambda_{0j0})(s)\right\}$ is the same as the asymptotic variance of $\sqrt{n}\left\{(\hat{\theta}_n - \theta_0)^T v + \sum_{j=1}^J \sum_{i=1} \sum_{k=1}^{n_{ij}} h_j(t_{ijk}) \hat{\Lambda}_{j0}(t_{ijk})\right\}$, while the latter one can be estimated by the following corollary.

Corollary 3.1 *Under assumptions (A1)-(A5), I_n is invertible and*

$$\sup_{v \in \mathcal{V}, h_1, \cdots, h_J \in Q} \Bigg| n(v^T, \bar{h}_1^T, \cdots, \bar{h}_J^T) I_n^{-1} (v^T, \bar{h}_1^T, \cdots, \bar{h}_J^T)^T - AVar\left[\sqrt{n}\left\{(\hat{\theta}_n - \theta_0)^T v + \sum_{j=1}^J \int_0^\tau h_j(s) d(\hat{\Lambda}_{j0} - \Lambda_{0j0})(s)\right\}\right] \Bigg| \xrightarrow{P} 0,$$

where AVar denotes asymptotic variance and $\bar{h}_j$ is the vector consisting of the values of $h_j(\cdot)$ at the observed event times.

We can also estimate the asymptotic variance matrix Σ of $\sqrt{n}(\hat{\theta} - \theta_0)$ independently with profile log-likelihood function:

Theorem 3.4 *Let $PL_n(\theta)$ be the profile log-likelihood function for θ. Under the assumptions (A1)-(A5), for any $\epsilon_n = O_p(n^{-1/2})$ and any vector v,*

$$-\frac{PL_n(\hat{\theta}_n + \epsilon_n v) - 2PL_n(\hat{\theta}_n) + PL_n(\hat{\theta}_n - \epsilon_n)}{n\epsilon_n^2} \xrightarrow{P} v^T \Sigma^{-1} v.$$

Furthermore, $2\{PL_n(\hat{\theta}_n) - PL_n(\theta_0)\} \xrightarrow{d} \chi^2_d$.

Chapter 4: Real Data Analyses with Generalized Random Effect Model

We use an electronic health record dataset to investigate the application of dynamic random effect model in the last chapter. The data comes from a healthcare and insurance company. It includes patient-level prescription and diagnosis information in around 40 million people with almost a decade observation period. The diagnosis information is recorded with diagnosis date and type with ICD-9 code systems. For drug prescription, we can extract the name of the prescribed drug with brand name, the dates when the drugs are prescribed, and the length of period provided by clinicians. The unit for both dates are days. We assume the patient starts to be in drug exposure in the prescription day and keeps it until reaching the assigned number of days. If a new prescription is filled before the end of assigned number of days, we update the prescription length by the longer one.

We consider Myocardial Infarction (MI) and Angiotensin-converting-enzyme inhibitors (ACE inhibitors) as two event types. Myocardial Infarction is a medical emergency and often occurs when a blood clot blocks blood flow to the heart. Angiotensin-converting-enzyme inhibitor (ACE inhibitor) is a class of medication used primarily for the treatment of high blood pressure and heart failure. A meta analysis [2] shows evidence that an ACE inhibitor within 3 to 16 days of infarction can slow the progression of cardiovascular disease and improve the survival rate. Our objective is to illustrate the strength of casual relationship between disease diagnosis and drug prescription with dynamic random effect.

There were $142,234$ patients who have at least one MI and $1,715,301$ patients who have at least one ACE Inhibitor prescription during their observation periods. The distribution of event numbers for patients are presented in Figure 4.1 and 4.2. The average number of MI diagnosis events and ACE Inhibitor prescription events are 4.73 and 12.56, respectively, showing recurrent events structure. Among $142,234$ patients with MI events, $75,934$ of them have ACE Inhibitor

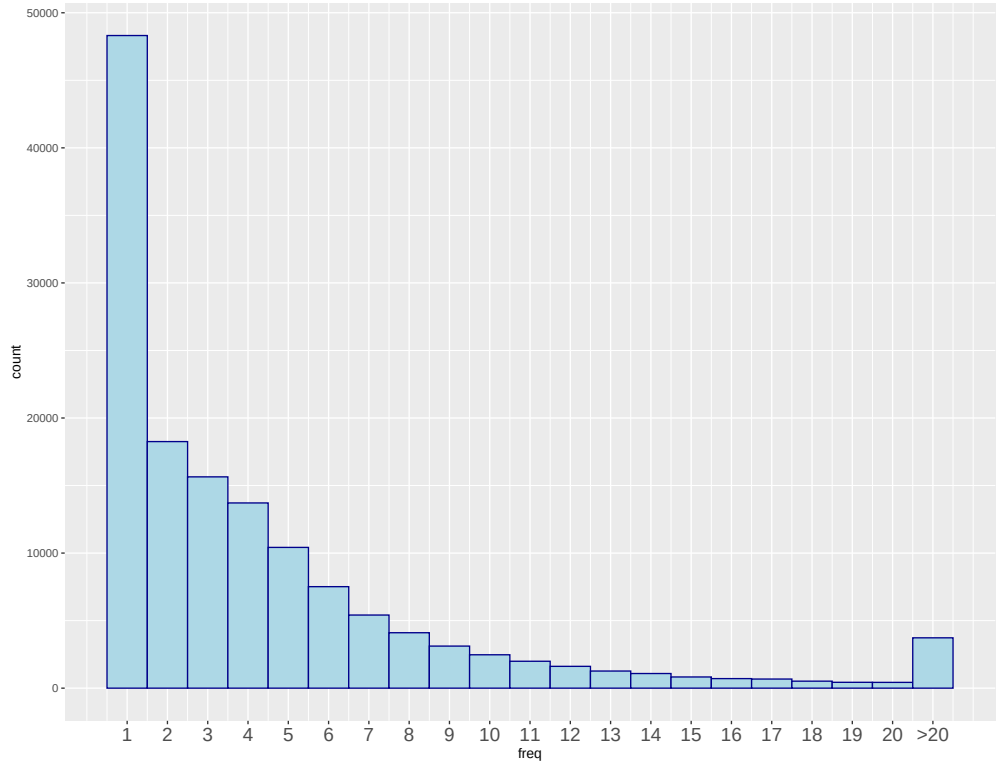

Figure 4.1: Histogram of MI frequency

prescription events with average number 14.61, which is a little higher than its population average.

Recall the intensity

$$\begin{aligned}\lambda_{i1}(t|\phi_{i1},\mathcal{F}_t) &= Y_i(t)e^{\alpha_1+\boldsymbol{\beta}_1^T\boldsymbol{X}_{i1}(t)+\boldsymbol{\eta}_1^T\mathbf{Z}_{i1}(t)\phi_{i1}},\\ \lambda_{i2}(t|\phi_{i2},\mathcal{F}_t) &= Y_i(t)e^{\alpha_2+\boldsymbol{\beta}_2^T\boldsymbol{X}_{i2}(t)+\boldsymbol{\eta}_2^T\mathbf{Z}_{i2}(t)\phi_{i2}}.\end{aligned} \tag{4.1}$$

As there is no obvious start point, we use constant baseline α_1, α_2 for both events. Let $T_{ij}(t)$ denote the latest occurring time point of event type j before time t. If no event occurs before time t, we define $T_{ij}(t) = 0$. We assume the random effect covariate $Z_{ij}(t)$ is based on the time gap between t and $T_{ij}(t)$ and fixed effect covariate $X_{ij}(t)$ is based on both the time gap between t and $T_{i1}(t)$ and the time gap between t and $T_{i2}(t)$. In other word, $Z_{ij}(t)$ is a function of $t - T_{ij}(t)$ and $X_{ij}(t)$ is a function of $(t - T_{i1}(t), t - T_{i2}(t))$. Besides time varying covariates, the fixed effect covariates also include some time-independent features, e.g. gender and age. To reach the model identifiability, we assume that if there is not event occurring before t, $\eta_j^T Z_{ij}(t) = 1$. Besides, we

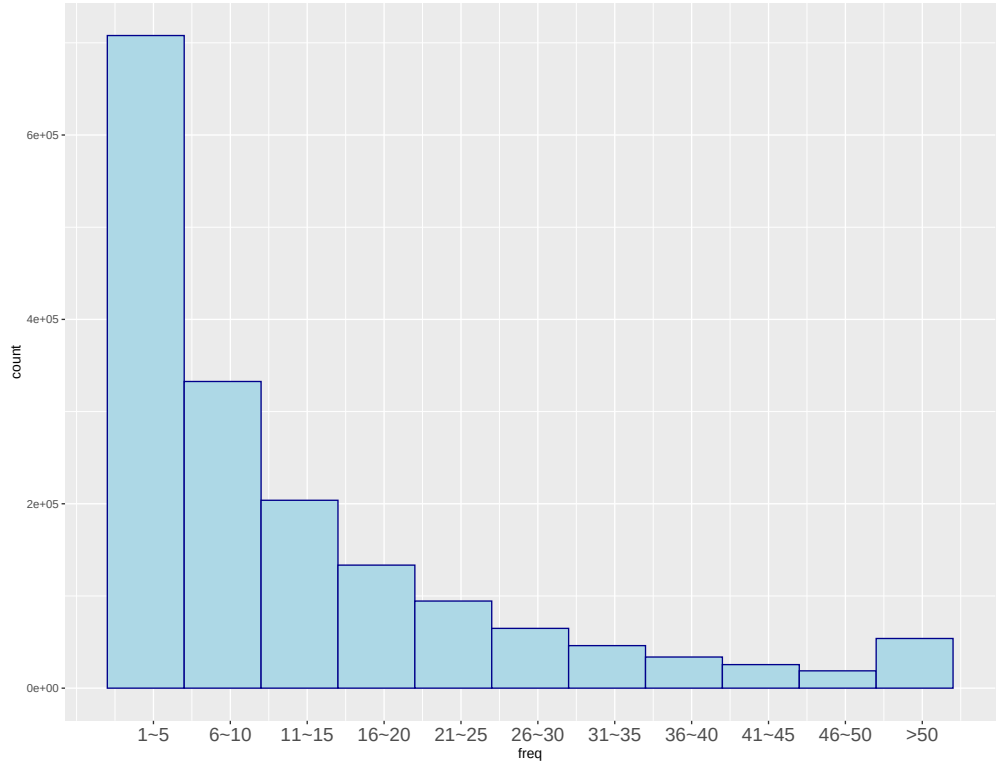

Figure 4.2: Histogram of ACE Inhibitor frequency

adopt piecewise linear functions for both $X_{ij}(t)$ and $Z_{ij}(t)$. Above all, the covariates can be

$$\begin{aligned}\boldsymbol{\beta}_1^T \boldsymbol{X}_{i1}(t) =& \beta_{gender1} \times \text{gender} + \beta_{age1} \times \text{age} + \beta_{2\to1} \mathcal{I}(T_{i2} > 0) \mathcal{I}(t - T_{i2} < W_1) \\ &+ \boldsymbol{\beta}_{1\to1}^T \begin{pmatrix} \mathcal{I}(T_{i1} > 0)(t - T_{i1}(t) - \xi_{10}^x)_+ \\ \mathcal{I}(T_{i1} > 0)(t - T_{i1}(t) - \xi_{11}^x)_+ \\ \vdots \\ \mathcal{I}(T_{i1} > 0)(t - T_{i1}(t) - \xi_{1L_1^x}^x)_+ \end{pmatrix},\end{aligned}$$

$$\begin{aligned}\boldsymbol{\beta}_2^T \boldsymbol{X}_{i2}(t) =& \beta_{gender2} \times \text{gender} + \beta_{age2} \times \text{age} + \beta_{1\to2} \mathcal{I}(T_{i1} > 0) \mathcal{I}(t - T_{i1} < W_2) \\ &+ \boldsymbol{\beta}_{2\to2}^T \begin{pmatrix} \mathcal{I}(T_{i2} > 0)(t - T_{i2}(t) - \xi_{20}^x)_+ \\ \mathcal{I}(T_{i2} > 0)(t - T_{i2}(t) - \xi_{21}^x)_+ \\ \vdots \\ \mathcal{I}(T_{i2} > 0)(t - T_{i2}(t) - \xi_{2L_2^x}^x)_+ \end{pmatrix},\end{aligned}$$

and

$$\boldsymbol{\eta}_j^T \boldsymbol{Z}_{ij}(t) = 1 + \boldsymbol{\eta}_{j\to j}^T \begin{pmatrix} \mathcal{I}(T_{ij} > 0)(t - T_{ij}(t) - \xi_{j0}^z)_+ \\ \mathcal{I}(T_{ij} > 0)(t - T_{ij}(t) - \xi_{j1}^z)_+ \\ \vdots \\ \mathcal{I}(T_{ij} > 0)(t - T_{ij}(t) - \xi_{jL_j^z}^z)_+ \end{pmatrix}.$$

The nodes for the piecewise linear function are calculated based on the quantiles of time gap between two consecutive time points of the events. We choose the number of nodes by minimizing the BIC criterion.

Table 4 and Fig 4.3 show the estimation results. In table 4, we compare three types of model and show some parameters estimation and criterion including AIC and BIC. The model III in the third column is the model we introduce above, which implement the correlated and dynamic random effect. The model I in the first column has the similar model setting except the random effect is uncorrelated but dynamic. The model II in the second column uses correlated but constant random effect. The table 4 shows the estimation of baseline parameters α_1, α_2, coefficients of time-independent covariates age and gender, the relative risk parameters of fixed effect from drug prescription event to disease diagnosis event $\beta_{2\to 1}$ and the reverse $\beta_{1\to 2}$, the hyperparameters of random effect distribution σ_1, σ_2, ρ and criterion AIC and BIC.

The coefficients of time-independent covariates age and gender are all positive, showing that older patients and male patients are more likely to have MI diagnosis and ACE Inhibitor prescription. As for the relative risk parameters of fixed effect from drug prescription event to disease diagnosis event, we can find this estimation is significantly positive in model I but significantly negative in model II and III. As we learned from medical literature, ACE inhibitor can reduce the risk of MI in some sense, which satisfy the estimation results in model II and III which implement correlated random effect. Actually, in Chapter 2, we have proved that when the random effect is correlated with covariates, the estimation will be biased under independence assumption. The estimation of correlation coefficients ρ in random effect distribution is positive. This means that the events of MI diagnosis and ACE inhibitor prescription are likely to happen together, which satisfies

the common nature. Among three models, the model III gives the best performance with respect to both AIC and BIC, which support the necessity of correlated and dynamic random effects.

The Fig 4.3 gives the estimated random effect and fixed effect functions, (a) $\boldsymbol{\eta}_{1\to1}^T \boldsymbol{Z}_{i1}(t)$ (b) $\boldsymbol{\beta}_{1\to1}^T \boldsymbol{X}_{i1}(t)$ (c) $\boldsymbol{\eta}_{2\to2}^T \boldsymbol{Z}_{i2}(t)$ (d) $\boldsymbol{\beta}_{2\to2}^T \boldsymbol{X}_{i2}(t)$,in model III. The self fixed effect of MI diagnosis $\boldsymbol{\beta}_{1\to1}^T \boldsymbol{X}_{i1}(t)$ is negative and increases at the first 21 days and then decreases to zero as time goes. The self random effect $\boldsymbol{\eta}_{1\to1}^T \boldsymbol{Z}_{i1}(t)$ is below 1 at first, which is because patients would not go to the hospital immediately after the a diagnosis. For ACE inhibitor prescription, the self fixed effect $\boldsymbol{\beta}_{2\to2}^T \boldsymbol{X}_{i2}(t)$ shows peak at 30 days which meets the fact that most prescriptions provide 30 days of medication. The self random effect $\boldsymbol{\eta}_{2\to2}^T \boldsymbol{Z}_{i2}(t)$ are below 1 all the time. This shows that after a prescription event, the heterogeneity or unobserved covariate has relative small effects on the risk of next prescription event.

	Model I	Model II	Model III
α_1	-16.11 (-16.56,-15.66)	-17.25 (-17.73,-16.77)	-16.88 (-17.33,-16.43)
α_2	-12.32 (-12.42,-12.22)	-12.91 (-13.01,-12.81)	-12.21 (-12.31,-12.11)
β_{gender1}	0.65 (0.57,0.73)	0.75 (0.66,0.84)	0.67 (0.58,0.75)
β_{age1}	0.81 (0.77,0.85)	1.02 (0.98,1.06)	0.94 (0.90,0.98)
β_{gender2}	0.28 (0.24,0.32)	0.42 (0.38,0.46)	0.29 (0.26,0.31)
β_{age2}	0.61 (0.59,0.62)	0.81 (0.80,0.82)	0.63 (0.62,0.64)
$\beta_{2\to1}$	0.59 (0.47,0.71)	-0.27 (-0.38,-0.16)	-0.22(-0.33,-0.10)
$\beta_{1\to2}$	0.91 (0.70,1.12)	0.59 (0.47,0.71)	0.57(0.46,0.68)
σ_1	1.91 (1.78,2.05)	2.03 (1.85,2.21)	2.01 (1.88,2.14)
σ_2	2.05 (2.02,2.08)	1.98 (1.92,2.04)	2.10 (2.07,2.13)
ρ	0	0.63 (0.59,0.67)	0.57 (0.52,0.62)
-$\frac{1}{n}$log-likelihood	2.7787	2.7453	2.7267
$\frac{1}{n}$BIC	5.5599	5.4922	5.4559
$\frac{1}{n}$AIC	5.5578	5.4909	5.4538

Table 4.1: Some parameters estimation under three random effects models for the electronic health record database

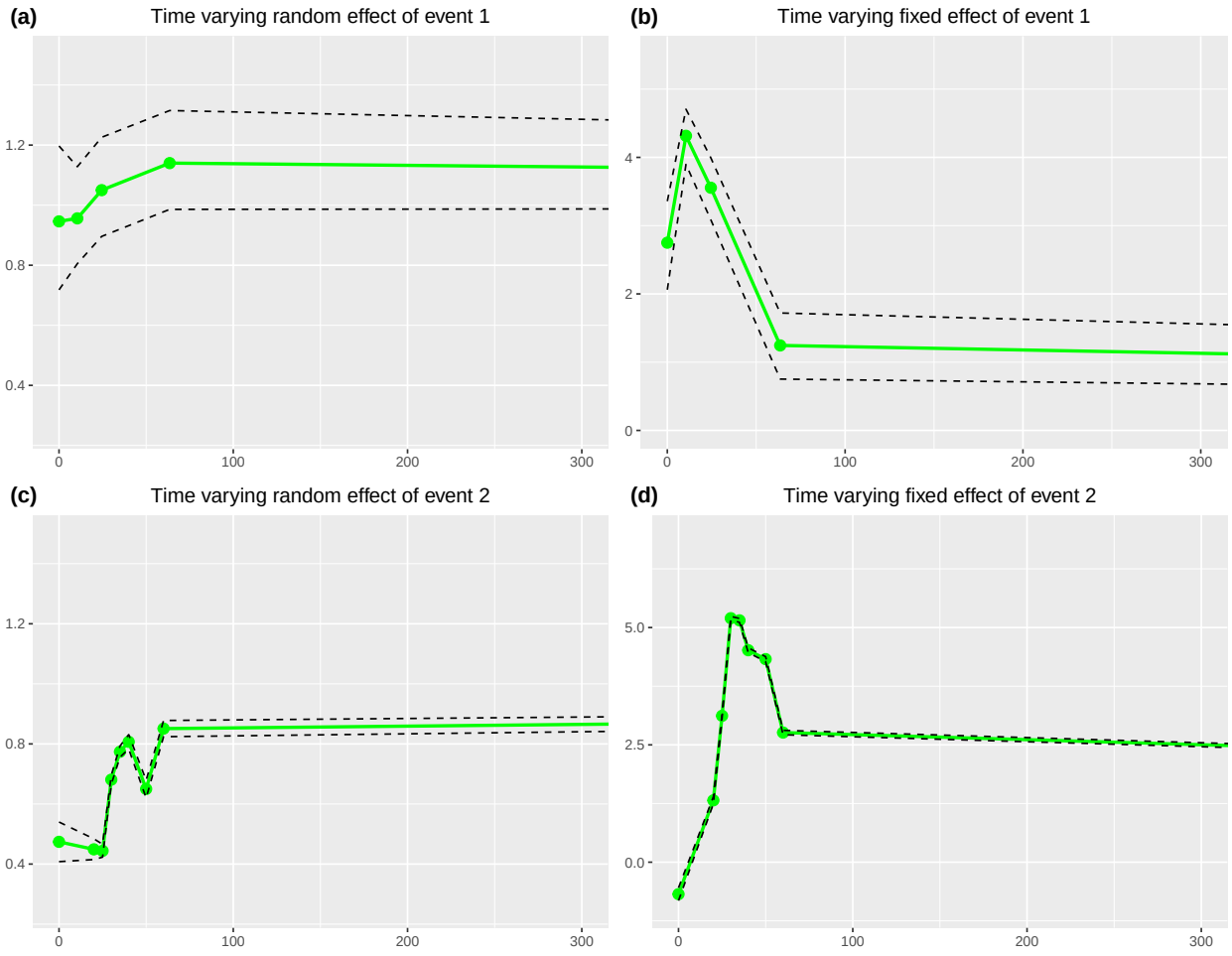

Figure 4.3: Results of time varying covariates functions for fixed effect and random effect, (a) $\boldsymbol{\eta}_{1\to1}^{T}\boldsymbol{Z}_{i1}(t)$ (b) $\boldsymbol{\beta}_{1\to1}^{T}\boldsymbol{X}_{i1}(t)$ (c) $\boldsymbol{\eta}_{2\to2}^{T}\boldsymbol{Z}_{i2}(t)$ (d) $\boldsymbol{\beta}_{2\to2}^{T}\boldsymbol{X}_{i2}(t)$

Chapter 5: Discussions

In this thesis, we discuss the event history analysis in multivariate longitudinal observational databases (LODs) and its application in postmarketing surveillance to identify and measure the relationship between events of health outcomes and drug exposures. The LODs contain a series of repeated observations of the same subjects over time and play a key role in finance, online assessment and clinical research. The scale and complexity of LODs require that the statistical analysis can explain the correlation structure of within-individual heterogeneity. If such correlation is ignored, the inferences such as statistical tests or estimation can be grossly invalid. The SCCS model [29, 30] implements an individual-level baseline constant and each individual serves as their own control. Although this method is good at explaining heterogeneity and eliminating bias, the SCCS method does not make full advantage of the observational databases. In this connection, the SCCS method may lead to loss of efficiency and fail in application of more general situations, especially in multivariate data.

We propose a multivariate proportional intensity model with random effects for multivariate longitudinal databases. A central component of the multivariate random effect model is its flexible representation to describe the association between multiple events when the independence assumption is violated. Besides, the random effect part can help explain the heterogeneity or unobserved covariates of individuals in datasets, achieved by high dimensional Log-Gaussian random effect with a more flexible form. For the estimation of the results, we adopt the Bayesian approach by Markov chain Monte Carlo method to get a series samples of target full semi-parametric likelihood. The simulation of different random effect settings illustrate the necessity of correlation structure within the unobserved covariates, especially in cases when covariates and outcomes are not independent. For real data application in Observational Outcomes Medical Partnership (OMOP) project, the performance of our proposed model and other existing model were shown by receiver

operating characteristic (ROC) metrics and our model performs best in an electronic health record (EHR) dataset.

Besides, we extend our multivariate proportional intensity model to dynamic random effect. This is due to the fact that the hidden factors of events can be influenced by the occurrence of the events. With such dynamic design, we develop the consistency and asymptotically normality theory of semi-parametric maximum likelihood estimation. With an electronic health record database, a detailed illustration of the proposed method is done with the clinical event Myocardial Infarction (MI) and drug treatment of Angiotensin-converting-enzyme (ACE) inhibitors, showing the dynamic effect of unobserved heterogeneity.

While we present the method using medical databases, the method is widely applicable. For example, in shopping website data, we are interested in their behaviors on the website and the association structure between these behaviors which gives some suggestions for our strategy. Besides, the method can be used in finance to better understand the effect of economic stocks.

In future works, we will explore the more complex structure of unobserved covariates in event history analysis. We discuss the dynamic and correlated random effect structure above. When the number of events grows, the dimension of unobserved covariates also increases. This may lead some overfitting problems. One of our future work is dealing with dimension analysis. Another future direction is to study the performance of the methods under sparsity conditions. The sparsity condition is really common in medical databases. Some adverse events, e.g rare disease and their orphan drugs, have really low frequency of occurrences. We are interested in the theory and application of random effect methods in such design.

References

[1] P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding, Statistical models based on counting processes. Springer Science & Business Media, 2012.

[2] R. Behnia, A. Molteni, and R. Igic, "Angiotensin-converting enzyme inhibitors: Mechanisms of action and implications in anesthesia practice," Current pharmaceutical design, vol. 9, no. 9, pp. 763–776, 2003.

[3] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," Pattern recognition, vol. 30, no. 7, pp. 1145–1159, 1997.

[4] R. J. Cook and J. Lawless, The statistical analysis of recurrent events. Springer Science & Business Media, 2007.

[5] D. R. Cox, "Regression models and life-tables," Journal of the Royal Statistical Society, vol. 34, no. 2, pp. 187–202, 1972.

[6] W. DuMouchel, "Bayesian data mining in large frequency tables, with an application to the fda spontaneous reporting system," The American Statistician, vol. 53, no. 3, pp. 177–190, 1999.

[7] C. Farrington, "Relative incidence estimation from case series for vaccine safety evaluation," Biometrics, pp. 228–235, 1995.

[8] C. Farrington, J Nash, and E Miller, "Case series analysis of adverse reactions to vaccines: A comparative evaluation," American journal of epidemiology, vol. 143, no. 11, pp. 1165–1173, 1996.

[9] S. Findlay, "Health policy brief: The fdas sentinel initiative," Health Affair, 2015.

[10] T. R. Fleming and D. P. Harrington, Counting processes and survival analysis. John Wiley & Sons, 2011, vol. 169.

[11] M. Gail, T. Santner, and C. Brown, "An analysis of comparative carcinogenesis experiments based on multiple times to tumor," Biometrics, pp. 255–266, 1980.

[12] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, "Introducing markov chain monte," Markov chain Monte Carlo in practice, p. 1, 1995.

[13] W. R. Gilks and P. Wild, "Adaptive rejection sampling for gibbs sampling," Journal of the Royal Statistical Society: Series C (Applied Statistics), vol. 41, no. 2, pp. 337–348, 1992.

[14] P. M. Grambsch and T. M. Therneau, "Modeling survival data: Extending the cox model," Statistics for Biology and Health, 2000.

[15] M. Greenwood, "The natural duration of cancer (report on public health and medical subjects no 33)," London: Stationery Office, 1926.

[16] R. G. Gutierrez, "Parametric frailty and shared frailty survival models," The Stata Journal, vol. 2, no. 1, pp. 22–44, 2002.

[17] P. Hougaard, "Survival models for heterogeneous populations derived from stable distributions," Biometrika, vol. 73, no. 2, pp. 387–396, 1986.

[18] G. Hripcsak et al., "Observational health data sciences and informatics (ohdsi): Opportunities for observational researchers," Studies in health technology and informatics, vol. 216, p. 574, 2015.

[19] J. D. Kalbfleisch and R. L. Prentice, The statistical analysis of failure time data. John Wiley & Sons, 2011, vol. 360.

[20] T. A. Moger and O. O. Aalen, "A distribution for multivariate frailty based on the compound poisson distribution with random scale," Lifetime Data Analysis, vol. 11, no. 1, pp. 41–59, 2005.

[21] S. A. Murphy, "Asymptotic theory for the frailty model," The Annals of Statistics, pp. 182–198, 1995.

[22] S. A. Murphy et al., "Consistency in a proportional hazards model incorporating a random effect," Annals of Statistics, vol. 22, no. 2, pp. 712–731, 1994.

[23] M. A. Robb et al., "The us food and drug administration's sentinel initiative: Expanding the horizons of medical product safety," Pharmacoepidemiology and drug safety, vol. 21, no. 1, p. 9, 2012.

[24] P Ryan and G Powell, "Exploring candidate differences between drug cohorts prior to exposure: A systematic approach using multiple observational databases pmc 14," Value in Health, vol. 11, no. 3, 2008.

[25] P. Ryan, G. Powell, K. Beach, et al., "Performance of screening multiple observational databases for active drug safety surveillance: 182.," Pharmacoepidemiology and Drug Safety, vol. 18, 2009.

[26] P. B. Ryan, D. Madigan, P. E. Stang, J Marc Overhage, J. A. Racoosin, and A. G. Hartzema, "Empirical assessment of methods for risk identification in healthcare data: Results from the experiments of the observational medical outcomes partnership," Statistics in medicine, vol. 31, no. 30, pp. 4401–4415, 2012.

[27] P. B. Ryan, M. J. Schuemie, S. Gruber, I. Zorych, and D. Madigan, "Empirical performance of a new user cohort method: Lessons for developing a risk identification and analysis system," Drug safety, vol. 36, no. 1, pp. 59–72, 2013.

[28] M. J. Schuemie, D. Madigan, and P. B. Ryan, "Empirical performance of lgps and leopard: Lessons for developing a risk identification and analysis system," Drug safety, vol. 36, no. 1, pp. 133–142, 2013.

[29] S. E. Simpson, "A positive event dependence model for self-controlled case series with applications in postmarketing surveillance," Biometrics, vol. 69, no. 1, pp. 128–136, 2013.

[30] S. E. Simpson, D. Madigan, I. Zorych, M. J. Schuemie, P. B. Ryan, and M. A. Suchard, "Multiple self-controlled case series for large-scale longitudinal observational databases," Biometrics, vol. 69, no. 4, pp. 893–902, 2013.

[31] P. E. Stang et al., "Advancing the science for active surveillance: Rationale and design for the observational medical outcomes partnership," Annals of internal medicine, vol. 153, no. 9, pp. 600–606, 2010.

[32] M. A. Suchard, I. Zorych, S. E. Simpson, M. J. Schuemie, P. B. Ryan, and D. Madigan, "Empirical performance of the self-controlled case series design: Lessons for developing a risk identification and analysis system," Drug safety, vol. 36, no. 1, pp. 83–93, 2013.

[33] J. Sultana, P. Cutroneo, and G. Trifirò, "Clinical and economic burden of adverse drug reactions," Journal of pharmacology & pharmacotherapeutics, vol. 4, no. Suppl1, S73, 2013.

[34] S. E. Vollset, A. Tverdal, and H. K. Gjessing, "Smoking and deaths between 40 and 70 years of age in women and men," Annals of internal medicine, vol. 144, no. 6, pp. 381–389, 2006.

[35] A. Yashin, A. Begun, and I. Iachine, "Genetic factors in susceptibility to death: A comparative analysis of bivariate survival models.," Journal of Epidemiology and Biostatistics, vol. 4, no. 1, pp. 53–60, 1999.

[36] D. Zeng and D. Lin, "Semiparametric transformation models with random effects for recurrent events," Journal of the American Statistical Association, vol. 102, no. 477, pp. 167–180, 2007.

[37] ——, “A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data,” Statistica Sinica, vol. 20, no. 2, p. 871, 2010.

[38] I. Zorych, D. Madigan, P. Ryan, and A. Bate, “Disproportionality methods for pharmacovigilance in longitudinal observational databases,” Statistical methods in medical research, vol. 22, no. 1, pp. 39–56, 2013.