

A Nested Frailty Model for Bivariate Recurrent Events: A Poisson Modelling Approach

by

Jiejie Wang

MSc. in Forestry, University of New Brunswick, 2017

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF**

Master of Science

In the Graduate Academic Unit of Mathematics and Statistics

Supervisor(s): Renjun Ma, Ph.D, Statistics
Guohua Yan, Ph.D, Statistics
Examining Board: Jeffrey D Picka, Ph.D, Statistics, Chair
M. Tariqul Hasan, Ph.D, Statistics
Rongxing Lu, Ph.D, Computer Science

This thesis is accepted by the

Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

July, 2018

©Jiejie Wang, 2018

Abstract

Survival analysis is used to study the time until the occurrence of an event of interest. Some events of interest can occur more than once in a subject. These events are termed recurrent events. In this thesis, we consider survival analysis of bivariate recurrent events in which each subject may experience two distinct types of recurrent events. For example, in the peritonitis dialysis study conducted in Taichung Veterans General Hospital in Taiwan, both Gram-positive and Non-Gram-positive peritonitis are observed on 575 patients over time. Each of these two types of peritonitis may occur more than once in a patient. Clearly these two types of recurrent events are clustered by subject. In addition, the recurrent events of each type are further clustered by the type of events. To characterize those clustering effects in our analysis, we incorporate two levels of nested frailties into Cox survival models to analyze bivariate recurrent events jointly.

There are many different approaches to the estimation of nested frailty Cox survival models in the literature. In this thesis, we propose a Poisson modelling approach to the estimation of our nested frailty Cox Survival models

for bivariate recurrent events. This approach enables us to develop an optimal model estimation based on orthodox best linear unbiased predictor of frailties in an auxiliary frailty Poisson model. An important feature of this approach is that the principal results depend only on the first and second moments of the unobserved frailties. Our approach deals with an unspecified baseline hazard function. In addition, the treatment of ties and stratification is explicitly incorporated in our approach in the same way as in the standard Cox model. The usefulness of our proposed method is illustrated through analysis of peritonitis dialysis data and a simulation study.

Dedication

This thesis is dedicated to my husband Xichen Zhang and my baby Luke Zhang, who have made me stronger and better than I could have ever imagined.

Acknowledgements

I would like to express my deepest appreciation to my supervisors Dr. Renjun Ma, and Dr. Guohua Yan, for their constant support, great patience and inspiration. The thesis could not be completed without their encouragement and persistent help.

I also like to thank Dr. Tariqul Hasan and Dr. Jeffery D Picka for their instruction and encouragement.

I express my gratitude to Department of Mathematics and Statistics in UNB for providing me the excellent opportunity to learn and stay.

Finally, I would like to thank all my family members and friends for their love and support.

Table of Contents

Abstract	ii
Dedication	iv
Acknowledgements	v
Table of Contents	viii
List of Tables	ix
List of Figures	x
1 Introduction	1
2 Literature Review	5
2.1 Introduction to Survival Analysis	5
2.1.1 Survival Data and Survival Analysis	5
2.1.2 General Approaches	7
2.1.2.1 Non-parametric Method	8
2.1.2.2 Semi-parametric Model	10

2.2	Multivariate / Bivariate Recurrent Events Analysis	13
2.3	Introduction to Poisson Model	16
2.4	Joint Model for Survival Analysis	18
2.4.1	Conditional Model	19
2.4.2	Shared-parameter Model	19
3	Cox Frailty Survival Model for Bivariate Recurrent Events	21
3.1	Joint Cox Frailty Survival Model for Bivariate Recurrent Events	21
3.1.1	Assumption 1	22
3.1.2	Assumption 2	22
3.1.3	Assumption 3	23
3.2	Auxiliary Joint Frailty Poisson Models for Bivariate Recurrent Event	24
3.2.1	Model Specification	25
3.2.1.1	Assumption 1 & 2	25
3.2.1.2	Assumption 3	25
3.2.2	Equivalence between Survival Model and Auxiliary Joint Poisson Model	28
3.2.3	Moment Structure	30
3.3	Orthodox Best Linear Unbiased Predictor	32
3.3.1	Prediction of Frailties	32
3.3.2	Estimation of Regression Parameters	34
3.3.3	Estimation of Dispersion Parameters	38

3.3.4	Practical Methods	39
3.3.4.1	Initial Values	39
3.3.4.2	Update Procedure	40
3.3.5	Computational Procedures	41
4	Data Analysis	43
4.1	Peritonitis Dialysis Data	44
4.1.1	Data Description	44
4.1.2	Preliminary Analysis Using K-M Curves	46
4.2	Bivariate Recurrent Event Analysis Treating Recurrent Events as Identical	49
4.3	Bivariate Recurrent Event Analysis Consider the Order of Events Occurrence	52
4.4	Simulation	57
4.4.1	Simulation Process	58
4.4.2	Simulation Performance	59
5	Discussion	61
5.1	Conclusion	61
5.2	Further Study	62
	Bibliography	64
	Vita	

List of Tables

3.1	Toy example of how recurrent event data are transformed to 0/1	26
3.2	Toy example of ordering survival time	27
3.3	Toy example with the transformed result	27
4.1	Peritoneal Dialysis Data Variables Description	45
4.2	Peritoneal Dialysis Data Sample with Two Patients	45
4.3	Estimates and SEs for PD Treating Recurrent Events as Identical	50
4.4	Model Reduction 1 for PD Treating Recurrent Events as Identical tical	51
4.5	Model Reduction 2 for PD Treating Recurrent Events as Identical tical	52
4.6	Estimates and SEs for PD Bivariate Recurrent Event Analysis	53
4.7	Model Reduction 1 for PD Bivariate Recurrent Event Analysis	55
4.8	Model Reduction 2 for PD Bivariate Recurrent Event Analysis	56
4.9	Simulation Results	60

List of Figures

4.1	The K-M curves for Gram-Positive Peritonitis	47
4.2	The K-M curves for Non-Gram-Positive Peritonitis	48

Chapter 1

Introduction

Recurrent event data arise in longitudinal follow-up studies, where each subject may repeatedly experience a particular event over time. Examples of recurrent events include a patient's repeated admissions to hospitals, cancer recurrences, multiple infections and relapses. There is an abundant literature on the analysis of recurrent events (Box-Steffensmeier, 2006; Ye et al., 2007). One type of process for recurrent events might be renewal process, in which the waiting (gap) times between successive events are statistically independent (Cox, 1972). In other words, an individual is "renewed" after each event occurrence. Extension has been made of the "renewed" process to allow episode-dependent baseline hazards, time-dependent covariates and strata (Prentice et al., 1981; Wei et al., 1989; Andersen and Gill, 1982; Kiefer, 1988; Box-Steffensmeier, 2006). Cook and Lawless (2007) provide models and thorough statistical methods for the analysis of recurrent event data.

Alternatively, each subject may experience multiple types of recurrent events, and this phenomenon is called multivariate recurrent event. Multivariate recurrent event data are encountered in many clinical and longitudinal studies. For example, chapter 4 considers a peritonitis dialysis study of patients who experience repeatedly two distinct types of peritonitis, which are Gram-positive peritonitis and Non-Gram-positive peritonitis.

Traditional methods for multivariate recurrent event data are to model each type of event independently. However, this does not give the full consideration to the multivariate nature of the data. Thus one might be interested in studying multiple types simultaneously. With the inclusion of frailty, it becomes possible to analyze bivariate or multivariate recurrent event processes (Ning et al., 2015, Abu-Libdeh et al., 1990, Ibrahim et al., 2010, Chen et al., 2004). But the exiting approaches to frailty multivariate recurrent events models generally rely on specific distribution assumption of frailties. For instance, Cook et al. (1999) and Abu-Libdeh et al. (1990) propose bivariate models in which both the hazard and frailty distribution are parametrically specified.

Incorporating random effects into Cox proportional hazard models has gained increasing attention over the past 30 years (Cox, 1972; Ibrahim, 2005). But theoretical difficulties exist when considering estimation and inference procedures (Lin and Daniel 2002). One popular approach to the frailties Cox model is using Bayesian approaches. However, they are computationally intensive and thus may not be practical for large datasets (Sargent, 1998; Sinha

et al., 1997; Ibrahim, 2005). Furthermore, these approaches usually require specific distributions for the frailties; for instance, Yau (2001) considered log-normal distribution for frailties in a survival study. In all, flexible frailties models that do not rely on specific distributional assumption of frailties are needed; in addition, these models should be computationally feasible for large datasets.

Our research is motivated by Ma et al. (2002) and Ma (2003). Specifically, we propose a frailty Cox proportional hazard model using joint Poisson modelling approach for bivariate recurrent events. Each of the bivariate recurrent event processes follows a modulated renewal process with stratum and order of event occurrence dependent baseline hazards. The bivariate processes are assumed to be conditionally independent on a pair of nested correlated frailties. Our method avoids the specification of the frailty distribution, instead requiring only its first two moments. Parameter estimation is based on orthodox best linear unbiased predictor. That is, these are the truly best linear unbiased predictors (Brockwell and Davis, 1991).

There are some benefits that come with our approach; for instance, the frailties are not limited to specific distributions; thus our framework effectively covers a variety of distributions. Furthermore, the extension to multi-level frailties is very straightforward. In addition, the Newton scoring method is more computationally effective than the EM algorithm, given that the Newton scoring algorithm does not seek the maximum at every iteration. Lastly, the joint model is more “robust” than the separate model since it takes more

information into account.

The remainder of my thesis is organized into the following four parts.

In chapter 2, we introduce the characteristics of survival data and survival analysis, especially multivariate or bivariate recurrent events data. The general approaches in survival analysis are reviewed, the Kaplan-Meier method and Cox proportional hazard regression model, corresponding to non-parametric and semi-parameter method, respectively. In this chapter, we also introduce Poisson models and joint models.

In chapter 3, we present the Cox frailty survival model and auxiliary joint frailty Poisson model. The models are formulated hierarchically through three assumptions. The connection between the frailty Cox survival model and the auxiliary frailty Poisson model is discussed. The moment structures are given as well. To predict the frailties and estimate the parameters, we use the orthodox best linear unbiased predictor approach. Then the computational procedure is presented as well.

In chapter 4, to demonstrate the proposed model, we analyze the peritonitis dialysis (PD) data. The PD data contains 300 patients who are observed for bivariate recurrent events (Gram-positive peritonitis and Non-Gram-positive peritonitis) over a period of time. Some patients may experience either type of the peritonitis more than once. Analysis results along with relevant graphs and figures are presented and discussed. In addition, we conduct a simulation study to assess the accuracy of our estimating algorithm.

Some conclusions and a discussion of further work are given in chapter 5.

Chapter 2

Literature Review

2.1 Introduction to Survival Analysis

2.1.1 Survival Data and Survival Analysis

Survival analysis is generally used for analyzing data where the outcome variable is the time until the occurrence of an event of interest. Survival analysis has gained increasing importance since the 1950s (Langova, 2008; Cox, 2008). It has been widely used in a variety of domains, for instance, social science, engineering, environmental science, medical research. The objectives for survival analysis include the analysis of patterns of survival times, the estimation and interpretation of survivor function, the examination of factors affecting the risk of event of interest, etc.

Survival data consist of event of interest, survival time (or time to event), explanatory variables, censoring, etc. The event of interest can be death, oc-

currence of a disease, machine failure, marriage, etc. The survival time is a period of time from the time observations begin until the end of the observations. It can be measured in hours, days, weeks, months, and so on (Bewick and Ball, 2004). For example, if the event of interest is epilepsy, then the survival time can be the time in days from when the epilepsy symptoms occur in that person until the epilepsy symptom disappear. Explanatory variables refer to the variables that may impact the survival time of observations.

One important issue in survival analysis is censoring. Censoring occurs when information about survival time is incomplete, and it happens for a variety of reasons; for instance, the research is terminated, or the subject does not have an event during the observation time, or some irrelevant causes. The most common type of censoring is right censoring. For example, patients are followed after a heart attack for 1 year. A patient who does not experience heart attack for the duration of the study, or drops out before the end of the study, is said to be right censored. Censoring represents a particular type of missing data. Using traditional methods, for instance, t-test or linear regression model, which fail to take censoring into account can produce unsatisfactory results, such as bias in estimates of the distribution of survival time. Thus we need to use appropriate methods when analyzing survival data, especially when censoring occurs.

We introduce and describe two quantitative terms considered in any survival analysis. These are the survivor function, denoted by $S(t)$, and the hazard function, denoted by $h(t)$. The survivor function $S(t)$ is fundamental to

survival analysis. It gives the probability that a person survives longer than some specified time t : that is, $S(t)$ gives the probability that the random variable T exceeds the specified time t .

$$S(t) = P(T \geq t) = \int_t^{\infty} f(x)dx, \quad t \geq 0 \quad (2.1)$$

The hazard function $h(t)$ is the instantaneous rate at which events occur for individuals which are surviving at time t . $h(t)$ can be shown as,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}, \quad t \geq 0 \quad (2.2)$$

2.1.2 General Approaches

There have been three popular survival analysis techniques in recent years. The first method is non-parametric, for instance, the Kaplan-Meier method, which has been widely used to estimate survival probabilities as a function of time (Akritas, 2004). The second method is semi-parametric, such as the Cox proportional hazard method (Lane et al., 1986). It is called “semi-parametric” method because the baseline hazard function, $h_0(t)$ does not need to be specified. The third one is the parametric method, which assumes that the distribution of the survival time follows a parametric probability distribution, such as exponential, Poisson, log-normal, etc. (Hosmer et al., 2011). The first two methods are discussed in the following sections. Detailed information about the parametric method can be found in Kleinbaum (1998).

2.1.2.1 Non-parametric Method

One of the most important non-parameter methods is called the Kaplan-Meier Method. In 1958, Edward L. Kaplan and Paul Meier published a paper on how to deal with incomplete data (Kaplan and Meier, 1958). Subsequently, the Kaplan-Meier estimates of survival data became a familiar way of dealing with differing survival times, especially dealing with right censoring problems. The survival function of the Kaplan-Meier method can be easily estimated with Kaplan-Meier plot, life-table, or cumulative hazard estimator (Cox and Oakes, 1984; Miller, 2011; Narendranathan and Stewart, 1991). One of the main advantages of Kaplan-Meier method is that it is simple and does not need any distributional assumptions for its hazard function. Kaplan-Meier method is also useful for summarizing survival data, getting survival function for a population, and making comparisons of survival functions among groups. When the sample size becomes large, its estimated survival function can be used as population survival function.

The Kaplan-Meier method has some limitations, one of the main ones being that it is descriptive and does not take into account covariates. In addition, the Kaplan-Meier method mainly deals with categorical predictors, and it cannot directly deal with continuous variables. For instance, time-dependent variables, which are very common in survival data, cannot be accommodated well using the Kaplan-Meier method.

We provide the following two methods of estimating survival function. The first one is called Kaplan-Meier (or Product-Limit) estimator. This method

will be selected when the sample size is large. The Kaplan-Meier estimator is effective in estimating the survival function for all values of time t . Let t_1 denote the checkpoint or cutting point for occurrence of the event. The survival function will be 1 when $t < t_1$ since no event occurs before time t_1 . But its survival function changes as long as the event happens. The estimator is shown as follows,

$$\hat{S}(t) = \begin{cases} 1, & t < t_1 \\ \prod_{t_i \leq t} \left[1 - \frac{d_i}{n_i}\right], & t \geq t_1 \end{cases} \quad (2.3)$$

where d_i denotes the number of times the events occur during the i th range of time and n_i denotes the number of subjects at risk at the beginning of this interval.

Its variance estimator has the following formula,

$$\hat{V}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)} \quad (2.4)$$

After the survivor function been estimated, the hazard function can also be derived.

The second method, the Nelson-Aalen method, is selected when the sample size is small. The cumulative hazard estimator performed by the Nelson-

Aalen estimator is shown below,

$$\hat{H}(t) = \begin{cases} 0, & t \leq t_1 \\ \sum_{t_i \leq t} \frac{d_i}{n_i^2}, & t > t_1 \end{cases} \quad (2.5)$$

The estimated variance for the cumulative hazard function is shown as:

$$\hat{V}(\hat{H}(t)) = \sum_{t_i \leq t} \frac{d_i}{n_i^2} \quad (2.6)$$

After the hazard function is calculated, the survivor function can be estimated using the following formula:

$$\hat{S}(t) = \exp(-\hat{H}(t)) \quad (2.7)$$

2.1.2.2 Semi-parametric Model

The non-parametric method has its limitations; for instance, it is descriptive and does not take into account covariates. To overcome the limitations of the non-parametric method, we introduce the semi-parametric method, the Cox proportional hazard regression model. The Cox proportional hazard model was initially introduced by Cox in 1972 for survival analysis. In the Cox proportional hazard model, both the survival time and covariates are incorporated. Therefore, for each observation, it consists of the information of (T_i, δ_i, X_i) , where T_i denotes the survival time, δ_i denotes the censorship (δ_i is 1 if censored and 0 otherwise), and x_i denotes the covariates for i th

observation. The Cox proportional hazard model provides estimates of the hazard ratio and its confidence interval. It is called a “semi-parametric” method because the baseline hazard function, $h_0(t)$ does not need to be specified.

The Cox proportional hazard regression model can be shown as follows,

$$h_i(t; x_1, \dots, x_p) = h_0(t)\exp(\beta_1 x_{i1} + \dots + \beta_p x_{ip}) \quad (2.8)$$

where $h(t)$ is the hazard function determined by a set of covariates (x_1, \dots, x_p) , and h_0 is the baseline hazard function when setting all covariates equal to zero ($x_1 = \dots = x_p = 0$).

The model is called the proportional hazard model, since the hazard ratio is proportional for two observations, and the hazard ratio is independent of time t . Suppose there are two observations, i and i^* . The hazard ratio for these two observations can be presented as,

$$\frac{h_i(t)}{h_{i^*}(t)} = \frac{h_0(t)\exp(\gamma_i)}{h_0(t)\exp(\gamma_{i^*})} = \frac{\exp(\gamma_i)}{\exp(\gamma_{i^*})} \quad (2.9)$$

where

$$\begin{aligned} \gamma_i &= \beta_1 x_{i1} + \dots + \beta_p x_{ip} \\ \gamma_{i^*} &= \beta_1 x_{i^*1} + \dots + \beta_p x_{i^*p} \end{aligned}$$

The significance test for the covariates may be required once the Cox pro-

portional hazard model is established. There are three popular methods to test the significance of covariates, which are the Wald test, the Likelihood ratio test, and the score test. Detailed information about these three test methods can be found in Kleinbaum (1998) and Klein et al. (2005).

As mentioned before, the Cox model is one of the most popular methods for analyzing survival data. There are several reasons for the popularity of the Cox model.

First of all, even though the baseline hazard is not specified, the Cox proportional hazard model is still a “robust” model. Reasonably good estimates of regression coefficients, hazard ratios of interest, and adjusted survival curves can be obtained.

Secondly, the specific form of the Cox model gives the hazard function as a product of a baseline hazard involving time t and an exponential expression involving X 's without t . The exponential expression is appealing as it ensures that the fitted model will always give non-negative estimated hazards.

Moreover, even though the baseline hazard part is unspecified, the hazard ratio can be calculated without having to estimate the baseline hazard function.

2.2 Multivariate / Bivariate Recurrent Events Analysis

Outcome events may occur more than once over the follow-up time for a given individual. Such events are called “recurrent” events. Recurrent event data occur in a wide range of circumstances including medicine, insurance, longitudinal biomedical research, and public health. Examples of recurrent event data include multiple episodes of relapses from remission, recurrent heart attacks of coronary patients, recurrence of vehicle insurance claims.

Different approaches have been used for analyzing recurrent events (Kleinbaum, 1998). For instance, when all recurrent events on the same subject are treated as identical, the counting process approach has been used. However, when recurrent events involve different categories and/or the order of events is considered important, a number of alternative approaches to analysis have been proposed that involving using stratified Cox models.

The model typically used to carry out the counting process approach is the standard Cox proportional hazard model, as shown in Equation 2.8. The main difference in the way the Cox model is used for analyzing recurrent events versus nonrecurrent data is the way several time intervals on the same subject are treated in the formation of the likelihood function maximized for the Cox model used (Kleinbaum, 1998). For nonrecurrent event data, each subject is removed from the risk set at the time of failure or censorship. In contrast, for recurrent data, a subject with more than one time interval

remains in the risk set until the last interval.

To distinguish the order in which recurrent events occur, the stratified Cox model has been used. There are three popular approaches for analyzing recurrent event data with the distinguish of order, each of which uses a stratified Cox proportional hazard model. They are called conditional 1, conditional 2, and marginal. These methods are often used to distinguish the order in which recurrent events occur. For instance, if the maximum number of failures that occur on any given subject in the dataset is 3, then time interval No.1 is assigned to stratum 1, time interval No. 2 to stratum 2, and so on.

Both the conditional approaches focus on survival time between two events, but there are differences between them. Conditional 1 uses the actual times of the two events from study entry, whereas conditional 2 starts survival time at 0 for the earlier event and stops at the later event. In contrast, the marginal approach focuses on total survival time from study entry until the occurrence of a specific event.

For recurrent events, the process of interest may consist of a single type of event. Statistical models appropriate for univariate recurrent events have been well-studied and the relevant literature has grown rapidly over the past 30 years (Lin et al. 2000; Lin et al., 2001; Cook and Lawless, 2007). For instance, Schaubel and Cai (2005) model the hospitalization rates among renal failure patients, Schaubel et al. (2006) study the hospitalization rates for post-kidney-transplantation, Lawless and Nadeau (1995) analyze the recur-

rent event of automobile warranty claims. Cook and Lawless (2007) provide a contemporary overview of the statistical analysis of recurrent events.

Sometimes there exist several types of related recurrent events, and this phenomenon is called multivariate recurrent events. Examples could be infections in bone marrow transplantation being sub-typed as bacterial, fungal and viral origins; or pulmonary exacerbation being classified according to severity. The traditional method for analyzing multivariate recurrent events is to model each event type separately. However, it is of great interest to study different types of recurrent event simultaneously. Related research includes Abu-Libdeh et al. (1990), who consider multi-type recurrent events in skin cancer involving two distinct types of lesions; Zhu et al. (2010) who perform the regression analysis on a set of bivariate recurrent event data arising from a study of leukemia patients; Cai and Schaubel (2004) who build marginal means/rates models for the multivariate recurrent events involving preschool children with asthma where hospitalizations and physician office visits are tracked. Several inference procedures have also been developed in the literature (Cai, 2004; Sun et al., 2009; Clegg et al., 1999).

Semi-parametric approaches for multivariate recurrent events have been considered. Most of these research methods consider the dependence between multivariate processes. Instead of modelling the full multivariate intensity function, they focus on the estimation of univariate rate and mean function (Ng and Cook, 1997; Ng and Cook, 1999; He et al., 2007; Cai and Schaubel, 2004). For instance, Ng and Cook (1999) consider a class of semi-

parametric marginal rates models for bivariate point processes, treating each of the marginal rates as following a proportional mean model.

An interesting alternative approach is proposed by Xue (1998), who adopts a bivariate frailty model to analyze multivariate survival data. His approach is developed based on a Poisson regression formulation and applies quasi-likelihood estimating equations to estimate parameters of interest and thus avoid using parametric assumptions for the frailty distribution.

Related research can be found in Ma et al. (2002), who propose a nested frailty Cox proportional hazards model for recurrent events. They adopt a Poisson modelling approach and their work shows the connection between the partial likelihood under the Cox proportional hazard model and the Poisson regression model. They obtain the parameter estimators of the Cox model based on the orthodox best linear unbiased predictor of the auxiliary random effects Poisson models. Their proposed methodology is preferable as it only requires one to specify the first two moments of the underlying frailty distribution. Related research can be found in Ma et al. (2003), Ma and Jorgensen (2007), Wienke (2010), O'Quigley (2008).

2.3 Introduction to Poisson Model

Poisson distribution was firstly introduced by Simeon Denis Poisson (1781 - 1840). It is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space

if these events occur with a known constant rate and independently of the time since the last event (Haight 1967). A discrete random variable x follows Poisson distribution with parameter $\lambda > 0$, the probability mass function of x is given by:

$$f(k; \lambda) = \Pr(x = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where k takes values of 0, 1, 2, ..., λ is the average number of events per interval, and it is also called rate parameter; $\lambda = E(x) = \text{Var}(x)$, in other words, the positive real number λ is equal to the expected value of x , as well as its variance.

The Poisson distribution has been used in different fields, such as biological studies, finance and insurance, astronomy studies, etc. However, there is still limited research on using the Poisson modelling approach to frailty survival models. Ma et al. (2003) proposed connecting the auxiliary random effects Poisson model to the nested random effects Cox proportional hazards model with multi-level clusters. The use of the proposed models has been illustrated through data analysis of particulate air pollution and mortality. But their research is focused on univariate response. It is a new idea to use the joint frailty Poisson model with multivariate recurrent events in survival analysis.

2.4 Joint Model for Survival Analysis

In recurrent event data analysis, it is common that multivariate types are observed. For instance, in peritonitis dialysis study, bivariate types of peritonitis have been observed in each patient, which are Gram-positive peritonitis and Non-Gram-positive peritonitis. One of the most popular methods is to model each response separately. It may be sufficient to perform separate analyses for the different event types, especially if they occur more or less independently of each other. However, models for multivariate processes are most often needed if events are related or if the occurrence of one event affects the risk of another type of event (Elashoff et al., 2008; Therneau and Grambsch, 2013). In such cases, separate models can result in biased estimates and insufficient results. Therefore, joint models are desirable, since joint models incorporate all available information simultaneously and thus provide more valid and efficient inferences.

Joint models for survival analysis have attracted increasing attention over the past two decades (Wu et al 2012; Ratcliffe et al. 2004; Tsiatis, 2004; Wulfsohn, 1997). In the following section, we will introduce two basic approaches for joint analysis in multivariate recurrent events. Suppose there are two types of recurrent events, Y_1 and Y_2 , measured on each individual. We use bivariate types of responses just for simplicity, but the extension to multivariate responses is straightforward.

2.4.1 Conditional Model

The theory behind the conditional models is that the density function can be factorized into the product of a marginal and conditional density.

$$f(y_1, y_2) = f(y_1 | y_2)f(y_2) = f(y_2 | y_1)f(y_1) \quad (2.10)$$

The main advantage of this method is that it avoids specifying a joint distribution for (y_1, y_2) . This method requires a marginally specified model for one response and a conditionally specified model for the other. A choice needs to be made of which response is modelled marginally or conditionally. Sometimes different choices may lead to totally different results.

This method has some limitations; for instance, the marginally inference may not be derived directly. Furthermore, this method becomes computationally intensive and impractical when more events are involved, since more factorization must be taken into account.

2.4.2 Shared-parameter Model

The second method is “shared-parameter” model. The theory behind the shared-parameter method is that the responses share common frailties; for instance, in the peritonitis dialysis study, two types of recurrent events are recorded on each patient. These two responses shared common frailty on the subjects’ level.

Let a be a vector of frailty. Let y_1 and y_2 be responses. The responses of

y_1 and y_2 remain the specific model and they both are dependent on a . The joint density for (y_1, y_2) can be expressed as follows:

$$f(y_1, y_2) = \int f(y_1, y_2 | a)f(a)d(a) = \int f(y_1 | a)f(y_2 | a)f(a)d(a) \quad (2.11)$$

where $f(a)$ is the marginal density of frailty. Since both of the responses y_1 and y_2 depend on a , then a is called the “shared-parameter” for y_1 and y_2 . This method has some advantages, such as the interpretation of the parameters for both the joint model and “univariate” model stays the same. Furthermore, it is straightforward to extend it to multiple types of responses, and this will not lead to computational intensity compared with the “conditional method” described above. However, this method has some limitations. For instance, it implies a strong assumption about the association between responses.

Chapter 3

Cox Frailty Survival Model for Bivariate Recurrent Events

In this chapter, we propose the Cox frailty survival model for bivariate recurrent events using the Poisson modelling approach. Random effects are estimated using the orthodox best linear unbiased predictor approach. Detailed information about our proposed models and estimation methods are given in the following sections.

3.1 Joint Cox Frailty Survival Model for Bivariate Recurrent Events

In the previous chapter, we have discussed the semi-parametric method, Cox proportional hazard model. In this section, we consider Cox model for bi-

variate recurrent events. We assume that the cohort of interest is composed of m independent clusters (or subjects) indexed by i . Then within the i th cluster, there are J_i sub-clusters (or event types) indexed by (i, j) . $J_i=2$ for bivariate event types.

3.1.1 Assumption 1

Let the cluster-level frailties be U_1, \dots, U_m . Specifically, we assume that they are positive, and independent and identically distributed with

$$E(U_i) = 1 \quad \text{and} \quad \text{Var}(U_i) = \sigma^2. \quad (3.1)$$

3.1.2 Assumption 2

We assume that there is a sub-cluster frailty U_{ij} associated with sub-cluster j from i th cluster, $j = 1, \dots, J_i$ ($J_i=2$ for bivariate recurrent events); $i = 1, \dots, m$. We further assume that, given the cluster-level frailties of $U_* = (U_1, \dots, U_m)^T$, the sub-cluster frailty U_{11}, \dots, U_{mJ_m} are positive and conditionally independent, and that the conditional distribution of U_{ij} given U_* , depends on U_i , with

$$E(U_{ij} | U_*) = U_i \quad \text{and} \quad \text{Var}(U_{ij} | U_*) = \tau_j^2 U_i. \quad (3.2)$$

Specifically, for bivariate recurrent events,

$$\begin{aligned} E(U_{i1} | U_*) &= U_i \quad \text{and} \quad \text{Var}(U_{i1} | U_*) = \tau_1^2 U_i. \\ E(U_{i2} | U_*) &= U_i \quad \text{and} \quad \text{Var}(U_{i2} | U_*) = \tau_2^2 U_i. \end{aligned}$$

3.1.3 Assumption 3

We furthermore assume that, there are l_{ij} recurrent events within each sub-cluster of (i, j) . Suppose that for each type of recurrent event, the cohort is stratified by one or more relevant covariates. Then the strata are indexed by $s = 1, 2, \dots, a$. Let $h_{ijk}^{(s)}(t)$ denotes the hazard function for the Cox model with (i, j, k) (k th recurrent event occurrence of j th event type in the i th subject) from stratum s at time t . Given both the cluster and sub-cluster frailties, we assume that the hazard functions are,

$$h_{ijk}^{(s)}(t) = h_{0jk}^{(s)}(t)u_{ij}\exp(\beta_j^T x_{ijk}^{(s)}). \quad (3.3)$$

Specifically, for bivariate recurrent events,

$$\begin{aligned} h_{i1k}^{(s)}(t) &= h_{01k}^{(s)}(t)u_{i1}\exp(\beta_1^T x_{i1k}^{(s)}). \\ h_{i2k}^{(s)}(t) &= h_{02k}^{(s)}(t)u_{i2}\exp(\beta_2^T x_{i2k}^{(s)}). \end{aligned}$$

where $k = 1, 2, \dots, l_{ij}$ with l_{ij} being the maximum number of observed recurrent occurrences for j th event type from i th subject, where $h_{0jk}^{(s)}(t)$ denotes

the unspecified baseline hazard function and $x_{ijk}^{(s)}$ denotes the covariates vector for the k th occurrence of recurrent event for j th event type in i th subject from s th stratum. The survival time within each cluster is correlated, no matter whether the observations are censored or not. The distribution of two level frailties does not depend on the regression parameter β .

Our model assumptions only require the specification of the first two moments of frailties. This is desirable since it is impractical to fully understand the random mechanism of how the unobserved frailties are generated. In addition, our assumptions (1) and (2) cover a wide range of frailty distribution, such as gamma, inverse Gaussian, log-normal, etc. For further details one can see R. Ma's PhD thesis in University of British Columbia (Ma, 1999).

3.2 Auxiliary Joint Frailty Poisson Models for Bivariate Recurrent Event

In this section, we propose an auxiliary joint frailty Poisson modelling approach for bivariate recurrent events. We illustrate how to make use of the connection between the partial likelihood under the Cox proportional hazard model and Poisson regression model. Estimates for frailties are obtained through orthodox best linear unbiased predictor. Our joint frailty Poisson modelling approach is described as follows.

3.2.1 Model Specification

Our joint frailty Poisson model is composed of m independent clusters (or subjects) that are indexed by i . Within each i th cluster, there are J_i correlated sub-clusters (or event types) indexed by (i, j) . Note that $J_1 = J_2 = \dots = J_m = 2$, since we only consider bivariate recurrent events. Within each sub-cluster (i, j) , there are l_{ij} correlated recurrent event observations.

We assume that there exist cluster-specific and sub-cluster-specific frailties. We consider the cluster-specific frailty U_i for the i th cluster and the sub-cluster-specific frailty U_{ij} for j th sub-cluster of i th cluster. Thus the vector of frailties can be written as $U = (U_1, \dots, U_m, U_{11}, \dots, U_{mJ_m})^T = (U_*, U_{**})$, where the U_* stands for $(U_1, \dots, U_m)^T$, and U_{**} stands for $(U_{11}, \dots, U_{mJ_m})^T$, respectively. Our joint frailty Poisson model is characterized by the following three assumptions.

3.2.1.1 Assumption 1 & 2

Assumptions 1 & 2 are the same for expressions (3.1) and (3.2), respectively.

3.2.1.2 Assumption 3

Let $S_{jk}^{(s)}$ be the set such that in the s th stratum, $(i, j) \in S_{jk}^{(s)}$ if and only if the j th recurrent event type of i th subject from s th strata has experienced the $(k-1)$ th recurrent event time. Let $\nu_{jk1}^{(s)}, \dots, \nu_{jkq_k}^{(s)}$ denote the distinct failure time of k th recurrent occurrence in $S_{jk}^{(s)}$, while $m_{jkh}^{(s)}$ denotes the multiplicity of failures occurring at time $\nu_{jkh}^{(s)}$. The risk set at time $\nu_{jkh}^{(s)}$ is a subset of

Table 3.1: Toy example of how recurrent event data are transformed to 0/1

ID	SurvivalTime	Count	Censorship	EventType
1	4	1	0	1
1	3	2	0	1
1	5	3	1	1
2	2	1	0	1
2	5	2	0	1

$S_{jk}^{(s)}$ such that $R(\nu_{jkh}^{(s)}) = \{(i, j, k) : t_{ijk}^{(s)} \geq \nu_{jkh}^{(s)}\}$, where $t_{ijk}^{(s)}$ denotes the k th recurrent occurrence time for j th event type in the i th subject from s th stratum. In addition, let $Y_{ijk,h}^{(s)}$ be 1 if failure occurs for (i, j, k) from s th strata at time of $\nu_{jkh}^{(s)}$, and 0 otherwise.

We use a toy example to show the details about how the data are transformed. Suppose that in a certain strata for a particular event type, there exist 2 individuals and each has four variables: Survival Time, Count (the cumulative number of lines in the data frame of recurrent events for each individual), Censorship (censorship be 1 if censored, and 0 otherwise), Event Type, see Table 3.1. Individual 1 has three lines with maximum counts of 3, and individual 2 has two lines. We show the univariate recurrent events type (with Event Type =1) example just for simplicity, and the extension to multivariate recurrent event types can be easily applied.

In Table 3.2, we order the survival times from the smallest to largest based on the ‘‘Count’’. For instance, for the first recurrent event (‘‘Count’’=1), we order the corresponding survival times from smallest to largest, which are 2 and 4. Similar procedures can be applied to the following recurrent

Table 3.2: Toy example of ordering survival time

ID	SurvivalTime	Count	Censorship	EventType
2	2	1	0	1
1	4	1	0	1
1	3	2	0	1
2	5	2	0	1
1	5	3	1	1

Table 3.3: Toy example with the transformed result

ID	Survival Time	Count	Censorship	EventType	Y
2	2	1	0	1	1
1	4	1	0	1	0
1	4	1	0	1	1
1	3	2	0	1	1
2	5	2	0	1	0
2	5	2	0	1	1

occurrences (“Count”=2 and “Count”=3).

In the 1st recurrent occurrence (with “Count”=1), the distinct failure times are 2 and 4. At the distinct failure time 2, the “Y” is 1 for individual ID=2 since the event happens at that time, and 0 for individual ID=1 since the event does not happen. Then the individual ID=2 is not contained in the risk set any more. At the distinct time of 4, the “Y” is 1 for individual ID=1 as the event happens at that time. Then individual ID=1 is not contained in the risk set after that. The similar procedure can be applied to the following recurrent events (“Count”=2 and “Count”=3). The transformed result is presented in Table 3.3

We now define an auxiliary random effects Poisson regression model. Given

the frailties of $U = (U_*, U_{**})$, the components of Y are,

$$Y_{ijk,h}^{(s)} \sim \text{Poisson}(\mu_{ijk,h}^{(s)} U_{ij}) \quad (3.4)$$

Specifically, for bivariate recurrent events,

$$\begin{aligned} Y_{i1k,h}^{(s)} &\sim \text{Poisson}(\mu_{i1k,h}^{(s)} U_{i1}) \\ Y_{i2k,h}^{(s)} &\sim \text{Poisson}(\mu_{i2k,h}^{(s)} U_{i2}) \end{aligned}$$

where $\mu_{ijk,h}^{(s)} = \exp(\alpha_{jk}^{(s)} + \beta_j^T x_{ijk,h}^{(s)})$. According to the properties of Poisson distribution, $E(Y_{ijk,h}^{(s)}) = \mu_{ijk,h}^{(s)} U_{ij}$ and $\text{Var}(Y_{ijk,h}^{(s)}) = \mu_{ijk,h}^{(s)} U_{ij}$.

3.2.2 Equivalence between Survival Model and Auxiliary Joint Poisson Model

The connections between the Cox proportional hazard model and the Poisson regression model have gained increasing attention for a long time (Whitehead, 1980; Therneau and Grambsch, 2013; Laird and Olivier, 1981; Breslow and Clayton, 1993; Lee and Nelder, 1996). Ma et al. (2002) proposes an interesting alternative for analyzing univariate recurrent events. Specifically, they make a connection between the partial likelihood under the Cox proportional hazard model and Poisson regression model. Our work is an extension of the methodology introduced by Ma et al. (2002).

Given the frailty of U , Peto's version of the conditional partial likelihood

(Cox and Oakes, 1984) is,

$$l_p(\beta; Y | U) = \prod_{j=1}^2 \prod_{s=1}^a \prod_{k=1}^{l_{ij}} \prod_{h=1}^{q_k} \frac{\prod_{(i,j) \in R(\nu_{jkh}^{(s)})} U_{ij}^{Y_{ijk,h}^{(s)}} \{\exp(\beta_j^T x_{ijk}^{(s)})\}^{Y_{ijk,h}^{(s)}} (m_{jkh}^{(s)})!}{\{\sum_{(i,j) \in R(\nu_{jkh}^{(s)})} U_{ij} \exp(\beta_j^T x_{ijk}^{(s)})\}^{m_{jkh}^{(s)}}} \quad (3.5)$$

Our auxiliary joint frailty Poisson model extends that of Whitehead (1980) for standard Cox models to frailty Cox models. The conditional likelihood for the joint frailty Poisson model, given the random effects, is:

$$l(\alpha, \beta; Y | U) = \prod_{j=1}^2 \prod_{s=1}^a \prod_{k=1}^{l_{ij}} \prod_{h=1}^{q_k} \frac{\prod_{(i,j) \in R(\nu_{jkh}^{(s)})} U_{ij}^{Y_{ijk,h}^{(s)}} \{\exp(\alpha_{jk}^{(s)} + \beta_j^T x_{ijk}^{(s)})\}^{Y_{ijk,h}^{(s)}}}{\exp\{\sum_{(i,j) \in R(\nu_{jkh}^{(s)})} U_{ij} \exp(\alpha_{jk}^{(s)} + \beta_j^T x_{ijk}^{(s)})\}} \quad (3.6)$$

It has been shown by Ma et al. (2002) that

$$l(\hat{\alpha}, \hat{\beta}; Y, U) = \prod_{j=1}^2 \prod_{s=1}^a \prod_{k=1}^{l_{ij}} \prod_{h=1}^{q_k} \left\{ \prod_{(i,j) \in R(\nu_{jkh}^{(s)})} \frac{m_{jkh}^{(s)} m_{jkh}^{(s)} \exp(-m_{jkh}^{(s)})}{m_{jkh}^{(s)}!} \right\} l_p(\hat{\beta}; Y, U) \quad (3.7)$$

where the term in parentheses on the right side from the equation (3.7) does not depend on the parameters of interest. From this we can conclude that the maximum joint Poisson likelihood estimators for the regression parameter vector β from equation (3.6) are the same as the maximum joint partial likelihood estimators for the regression parameter vector from equation (3.5). So we can estimate parameters about the frailty Cox models by fitting frailty Poisson models.

3.2.3 Moment Structure

We need to investigate the moment structures because the parameters estimators and frailties predictors are derived based on the moment structures of the model. The moment structures of our model are achieved through some algebraic calculations using conditioning technique. We use Kronecker notation of $\delta_{(i',i)}$ to help express the moment structure ($\delta_{(i',i)}$ will be 1 if $i' = i$ and 0 otherwise). Since all the derivations of the moment structures are similar, we just show several examples of them. To simplify the notations in the following moment structures, we use Y_{ijk} to represent $Y_{ijk,h}^{(s)}$, and μ_{ijk} to represent $\mu_{ijk,h}^{(s)}$.

The first and second moments of the sub-cluster-level frailties possess the following structure:

$$E(U_{ij}) = EE(U_{ij} | U_*) = E(U_i) = 1$$

$$\begin{aligned} \text{Cov}(U_{i'j'}, U_{ij}) &= \text{Cov}\{E(U_{i'j'} | U_*), E(U_{ij} | U_*)\} + E\{\text{Cov}(U_{i'j'}, U_{ij} | U_*)\} \\ &= \text{Cov}(U_{i'}, U_i) + \delta_{(i',i)}\delta_{(j',j)}E\{\text{Var}(U_{ij} | U_*)\} \\ &= \delta_{(i',i)}\text{Var}(U_i) + \delta_{(i',i)}\delta_{(j',j)}E(\tau_j^2 U_i) \\ &= \delta_{(i',i)}\sigma^2 + \delta_{(i',i)}\delta_{(j',j)}\tau_j^2 \\ &= \delta_{(i',i)}\{\sigma^2 + \delta_{(j',j)}\tau_j^2\} \end{aligned}$$

The covariance between sub-cluster specific frailties and responses is simply

given by

$$\begin{aligned}
\text{Cov}(U_{i'j'}, Y_{ijk}) &= \text{Cov}\{E(U_{i'j'} | U), E(Y_{ijk} | U)\} + E\{\text{Cov}(U_{i'j'}, Y_{ijk} | U)\} \\
&= \text{Cov}(U_{i'j'}, \mu_{ijk}U_{ij}) + 0 \\
&= \mu_{ijk}\text{Cov}(U_{i'j'}, U_{ij}) \\
&= \delta_{(i',i)}\{\sigma^2 + \delta_{(j',j)}\tau_j^2\}\mu_{ijk}
\end{aligned}$$

The first and second moments of the responses for the joint frailty Poisson models are then given by:

$$E(Y_{ijk}) = EE(Y_{ijk} | U) = E(\mu_{ijk}U_{ij}) = \mu_{ijk}$$

$$\begin{aligned}
\text{Cov}(Y_{i'j'k'}, Y_{ijk}) &= \text{Cov}\{E(Y_{i'j'k'} | U), E(Y_{ijk} | U)\} + E\{\text{Cov}(Y_{i'j'k'}, Y_{ijk} | U)\} \\
&= \text{Cov}(\mu_{i'j'k'}U_{i'j'}, \mu_{ijk}U_{ij}) + \delta_{(i',i)}\delta_{(j',j)}\delta_{(k',k)}E\{\text{Var}(Y_{ijk} | U)\} \\
&= \mu_{i'j'k'}\mu_{ijk}\text{Cov}(U_{i'j'}, U_{ij}) + \delta_{(i',i)}\delta_{(j',j)}\delta_{(k',k)}\mu_{ijk} \\
&= \delta_{(i',i)}(\sigma^2 + \delta_{(j',j)}\tau_j^2)\mu_{i'j'k'}\mu_{ijk} + \delta_{(i',i)}\delta_{(j',j)}\delta_{(k',k)}\mu_{ijk} \\
&= \delta_{(i',i)}\{\delta_{(j',j)}\delta_{(k',k)}\mu_{ijk} + \delta_{(j',j)}\tau_j^2\mu_{i'j'k'}\mu_{ijk} + \sigma^2\mu_{i'j'k'}\mu_{ijk}\}
\end{aligned}$$

We need to predict the frailties, since they are unobserved. In order to predict the frailties, the orthodox best linear unbiased predictor approach is used. To estimate the regression parameters, Newton scoring algorithm is used and iterative procedures are needed. The algorithm for the iterative

procedures consists of updating frailties and updating parameters estimates until convergence. Detailed information about the prediction of frailties, estimation of regression parameters and dispersion parameters are given in the following section.

3.3 Orthodox Best Linear Unbiased Predictor

In this section, we adopt the orthodox best linear unbiased predictor approach to predict frailties. We use the Newton scoring function and adjusted Pearson estimator to estimate the regression parameters and dispersion parameters, respectively.

3.3.1 Prediction of Frailties

Frailties can be predicted through the orthodox best linear unbiased predictor (BLUP) of U given Y , with the following expression:

$$\hat{U} = E(U) + \text{Cov}(U, Y)\text{Var}^{-1}(Y)(Y - E(Y)) \quad (3.8)$$

where $\text{Var}^{-1}(Y)$ denotes the marginal variance of Y . This is the linear unbiased predictor of U given Y . This linear unbiased predictor minimizes the mean squared distance between the frailties of U and their predictor \hat{U} within the class of linear functions of Y .

The mean squared distances between cluster-level frailty U_i and their predictors can be expressed as:

$$c_i = E(\hat{U}_i - U_i)^2 = \frac{\sigma^2}{1 + \sigma^2 \sum_{j=1}^2 \sum_{s=1}^a \sum_{k=1}^{l_{ij}} \sum_{h=1}^{q_k} \sum_{(i,j) \in R(\nu_{jkh}^{(s)})} w_{ij} \mu_{ijk,h}^{(s)}} \quad (3.9)$$

where (i, j) runs over the risk set $R(\nu_{jkh}^{(s)})$ for any given i . As shown before, the $\mu_{ijk,h}^{(s)}$ can be expressed as,

$$\mu_{ijk,h}^{(s)} = \exp(\alpha_{jk}^{(s)} + \beta_j^T x_{ijk,h}^{(s)})$$

Furthermore, the explicit expression for w_{ij} is as follows,

$$w_{ij} = \frac{1}{1 + \tau_j^2 \sum_{s=1}^a \sum_{k=1}^{l_{ij}} \sum_{h=1}^{q_k} \sum_{(i,j) \in R(\nu_{jkh}^{(s)})} \mu_{ijk,h}^{(s)}}$$

where (i, j) runs over the risk set $R(\nu_{jkh}^{(s)})$.

Similarly, the mean squared distances between components of the sub-cluster-level frailty U_{ij} and their predictors are as follows:

$$c_{ij} = E(\hat{U}_{ij} - U_{ij})^2 = w_{ij}(\tau_j^2 + c_i w_{ij}) \quad (3.10)$$

The expression of the cluster-level frailty predictors can be shown as follows,

$$\hat{U}_i = c_i \left(\frac{1}{\sigma^2} + \sum_{j=1}^2 \sum_{s=1}^a \sum_{k=1}^{l_{ij}} \sum_{h=1}^{q_k} \sum_{(i,j) \in R(\nu_{jkh}^{(s)})} w_{ij} Y_{ijk,h}^{(s)} \right) \quad (3.11)$$

where (i, j) runs over the risk set of $R(\nu_{jkh}^{(s)})$ for fixed i .

Similarly, we have the sub-cluster level frailty predictor expressed below:

$$\hat{U}_{ij} = w_{ij}\hat{U}_i + \tau_j^2 w_{ij} \sum_{s=1}^a \sum_{k=1}^{l_{ij}} \sum_{h=1}^{q_k} \sum_{(i,j) \in R(\nu_{jkh}^{(s)})} Y_{ijk,h}^{(s)} \quad (3.12)$$

where (i, j) runs over the risk set of $R(\nu_{jkh}^{(s)})$.

We can see that the right-hand side of the equations (3.11) and (3.12) involve the unknown dispersion parameters σ^2 and τ^2 . The dispersion parameters and frailties need to be estimated iteratively. At each iteration, the \hat{U}_i will be updated through evaluating the unknown dispersion parameters (σ^2 and τ^2) on the right-hand sides at their current values. In the meantime, \hat{U}_{ij} will be updated as well. We present the detailed information about how to process the iterative procedures in the remaining parts of this section.

3.3.2 Estimation of Regression Parameters

In this section, we estimate the regression parameters under the condition that dispersion parameters are known. Detailed information for estimating of unknown dispersion parameters of σ^2 and τ^2 is provided in the following section.

To estimate the regression parameters, we use the joint score function, which is yielded by differentiating the joint log-likelihood of the auxiliary Poisson model for the data and frailties. We then obtain an unbiased estimating

function for the regression parameters $\gamma = (\alpha^T, \beta^T)^T$, through replacing the random effects with their best linear unbiased predictors. The unbiased estimating function is shown as follows:

$$\begin{aligned}\psi(\gamma) &= \sum_{j=1}^2 \sum_{s=1}^a \sum_{k=1}^{l_{ij}} \sum_{h=1}^{q_k} \sum_{(i,j) \in R(\nu_{jkh}^{(s)})} x_{ijk,h}^{(s)} \{Y_{ijk,h}^{(s)} - \hat{U}_{ij}(\gamma) \mu_{ijk,h}^{(s)}(\gamma)\} \\ &= \sum_{i=1}^m \psi_i(\gamma)\end{aligned}\tag{3.13}$$

where

$$\psi_i(\gamma) = \sum_{j=1}^2 \sum_{s=1}^a \sum_{k=1}^{l_{ij}} \sum_{h=1}^{q_k} \sum_{(i,j) \in R(\nu_{jkh}^{(s)})} x_{ijk,h}^{(s)} \{Y_{ijk,h}^{(s)} - \hat{U}_{ij}(\gamma) \mu_{ijk,h}^{(s)}(\gamma)\}$$

$\psi(\gamma)$ is a vector function of the same dimension with γ . The second equality of equation (3.13) is just the rearrangement of terms. $\psi_i(\gamma)$ is obtained under the summation by clusters and it is corresponding to the i th cluster.

In the i th cluster, the frailties predictor $\hat{U}_{ij}(\gamma)$ is unbiased and it involves only the transformed data of $Y_{ijk,h}^{(s)}$. Similarly, in the i th cluster, the estimating function $\psi_i(\gamma)$ is unbiased and involves only the transformed data of $Y_{ijk,h}^{(s)}$. The transformed responses $Y_{ijk,h}^{(s)}$ from each cluster are independent; therefore, $\psi_1(\gamma), \dots, \psi_m(\gamma)$ are independent as well.

By setting the estimating function $\sum_{i=1}^m \psi_i(\gamma)$ equal to 0, we may obtain the estimators of the regression parameters. Note that the dimension of parameter γ increases with the number of clusters. The solution $\hat{\gamma}$ of the

estimating equation $\sum_{i=1}^m \psi_i(\gamma) = 0$ is shown to be asymptotically normal with asymptotic mean γ and asymptotic variance given by the inverse of the Godambe information matrix. The exact expression of Godambe information matrix is given by:

$$J(\gamma) = S(\gamma)V(\gamma)^{-1}S(\gamma)^T$$

where the sensitivity matrix $S(\gamma)$ and the variability matrix $V(\gamma)$ can be expressed as follows:

$$S(\gamma) = \sum_{i=1}^m S_i(\gamma) = \sum_{i=1}^m E_{\gamma} \left\{ \frac{\partial \psi_i(\gamma)}{\partial \gamma^T} \right\}$$

$$V(\gamma) = \sum_{i=1}^m V_i(\gamma) = \sum_{i=1}^m E_{\gamma} \{ \psi_i(\gamma) \psi_i^T(\gamma) \}$$

It has been verified that $S(\gamma) = -V(\gamma)$ for the nested frailty Poisson model; therefore, the Godambe information matrix $J(\gamma)$ is simply the sensitivity matrix multiplied by -1 for the auxiliary model (Ma et al., 2003). The estimating function $\sum_{i=1}^m \psi_i(\gamma)$ can be shown to be optimal since it attains the minimum asymptotic covariance matrix for the estimator $\hat{\gamma}$ among a certain class of linear function of Y (Ma et al., 2003; Crowder, 1986; 1987).

The estimating equation $\sum_{i=1}^m \psi_i(\gamma) = 0$ can be solved using the Newton scoring algorithm. The Newton scoring algorithm is introduced by Jørgensen

(Jørgensen, 1996). The updated value of γ can be shown as follows:

$$\gamma^* = \gamma - S^{-1}(\gamma)\psi(\gamma) \quad (3.14)$$

where γ^* denotes the updated value for γ .

The sensitivity matrix $S(\gamma)$ can be shown in the following way:

$$\begin{aligned} S(\gamma) = & \sum_{i=1}^m c_i e_i e_i^T + \sum_{i=1}^m \sum_{j=1}^2 \tau_j^2 w_{ij} f_{ij} f_{ij}^T \\ & - \sum_{j=1}^2 \sum_{s=1}^a \sum_{k=1}^{l_{ij}} \sum_{h=1}^{q_k} \sum_{(i,j) \in R(\nu_{jkh}^{(s)})} \mu_{ijk,h}^{(s)} x_{ijk,h}^{(s)} (x_{ijk,h}^{(s)})^T \end{aligned} \quad (3.15)$$

where

$$e_i = \left(\sum_{j=1}^2 \sum_{s=1}^a \sum_{k=1}^{l_{ij}} \sum_{h=1}^{q_k} \sum_{(i,j) \in R(\nu_{jkh}^{(s)})} w_{ij} \mu_{ijk,h}^{(s)} x_{ijk,h}^{(s)} \right) \quad (3.16)$$

$$f_{ij} = \left(\sum_{s=1}^a \sum_{k=1}^{l_{ij}} \sum_{h=1}^{q_k} \sum_{(i,j) \in R(\nu_{jkh}^{(s)})} \mu_{ijk,h}^{(s)} x_{ijk,h}^{(s)} \right) \quad (3.17)$$

In order to test the hypothesis $H_0: \beta_{(1)} = 0$, we use an analogue of Wald's test. The exact expression of the test statistics is shown as follows:

$$W = \hat{\beta}_{(1)}^T \{J^{11}(\hat{\gamma})\}^{-1} \hat{\beta}_{(1)} \sim \chi^2_{(k)}$$

where $\beta_{(1)}$ is a sub-vector of β , and $J^{11}(\hat{\gamma})$ is the block of the asymptotic covariance matrix of $\hat{\gamma}$ corresponding to $\beta_{(1)}$. W follows a $\chi^2(k)$ distribution,

where k denotes the size of $\hat{\beta}_{(1)}$.

3.3.3 Estimation of Dispersion Parameters

In the previous section, in order to estimate the regression parameter β , we assume that the dispersion parameters are known. In this section, we assume that the dispersion parameters of σ^2 and τ^2 are unknown. To estimate the dispersion parameter for the heterogeneity between clusters, we use the adjusted Pearson estimator. The explicit expression of the dispersion parameter estimator σ^2 can be shown as:

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^m \{(\hat{U}_i - 1)^2 + c_i\} \quad (3.18)$$

where the first term on the right side corresponds to the Pearson estimator, and the second term corresponds to the bias correlation.

Again, to estimate the dispersion parameter for subcluster specific heterogeneity between clusters, we adopt the corresponding adjusted Pearson estimator for τ_j^2 , which can be shown as follows:

$$\hat{\tau}_j^2 = \frac{1}{m} \sum_{i=1}^m \{(\hat{U}_{ij} - \hat{U}_i)^2 + c_{ij} + c_i - 2c_i w_{ij}\} \quad (3.19)$$

Similarly, the first term on the right side corresponds to the Pearson estimator, and the remaining ones correspond to the bias correlation.

3.3.4 Practical Methods

In this section, the practical procedures of how to obtain the initial values for the frailties are presented, as well as regression parameters and dispersion parameters. The procedures of how to update these values are discussed.

3.3.4.1 Initial Values

The initial values of regression parameters are obtained using the generalized linear model with Poisson family techniques. The initial values of μ are calculated as follows:

$$\mu_{ijk,h}^{(s)} = \exp(\alpha_{jk}^{(s)} + \beta_j^T x_{ijk,h}^{(s)})$$

The initial value of cluster-level frailty can be expressed as:

$$U_i^{(0)} = \frac{\frac{1}{J_i} \sum_{j=1}^{J_i} \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}}{\frac{1}{m} \sum_{i=1}^m \frac{1}{J_i} \sum_{j=1}^{J_i} \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}}$$

Note that $J_i = 2$ for bivariate recurrent event.

Similarly, the initial value of sub-cluster level frailty has the following expression:

$$U_{ij}^{(0)} = \frac{\frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}}{\frac{1}{m} \sum_{i=1}^m \frac{1}{J_i} \sum_{j=1}^{J_i} \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}}$$

The initial value for dispersion parameter σ^2 is

$$\sigma_{(0)}^2 = \frac{1}{m} \sum_{i=1}^m (\hat{U}_i - 1)^2$$

Similarly, the initial value for the dispersion parameter τ_j^2 is obtained by

$$\tau_{j(0)}^2 = \frac{1}{m} \sum_{i=1}^m (\hat{U}_{ij} - \hat{U}_i)^2$$

3.3.4.2 Update Procedure

The regression parameters estimates are updated through the Newton scoring algorithm. The sensitivity matrix can be expressed as follows:

$$S(\gamma) = - \sum_{i=1}^m \mathbf{X}_i^T \text{diag}(E(Y_i)) \text{Var}^{-1}(Y_i) \text{diag}(E(Y_i)) \mathbf{X}_i$$

The explicit expression of estimating equation can be shown below:

$$\psi(\gamma) = \sum_{i=1}^m \mathbf{X}_i^T \text{diag}(E(Y_i)) \text{Var}^{-1}(Y_i) (Y_i - E(Y_i))$$

Therefore, the updated value of γ can be expressed as follows:

$$\gamma^* = \gamma - S^{-1}(\gamma) \psi(\gamma)$$

The expressions for updating the estimates of cluster and sub-cluster level

frailties are shown below:

$$\hat{U}_i = c_i \left(\frac{1}{\sigma^2} + \sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} w_{ij} Y_{ijk} \right)$$

$$\hat{U}_{ij} = w_{ij} \hat{U}_i + \tau_j^2 w_{ij} \sum_{k=1}^{n_{ij}} Y_{ijk}$$

where

$$c_i = \frac{\sigma^2}{1 + \sigma^2 \sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} w_{ij} \mu_{ijk,h}^{(s)}}$$

$$w_{ij} = \frac{1}{1 + \tau_j^2 \sum_{k=1}^{n_{ij}} \mu_{ijk,h}^{(s)}}$$

The explicit expressions of the dispersion parameters estimators are expressed below:

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^m \{(\hat{U}_i - 1)^2 + c_i\}$$

$$\hat{\tau}_j^2 = \frac{1}{m} \sum_{i=1}^m \{(\hat{U}_{ij} - \hat{U}_i)^2 + c_{ij} + c_i - 2c_i w_{ij}\}$$

where $c_{ij} = w_{ij}(\tau_j^2 + c_i w_{ij})$, c_i and w_{ij} have the same expressions as above.

3.3.5 Computational Procedures

Iterative computation algorithms are needed due to the lack of a closed-form solution for the unknown parameters.

(1) The initial value of frailties predictions of the cluster-level \hat{U}_i are calcu-

lated by the average of responses within i th cluster divided by the average of all responses. Similarly, the initial values of frailties predictions of the sub-cluster level \hat{U}_{ij} are given by the average of the responses within (i, j) divided by the average of all responses.

(2) The initial values for the regression parameters β are obtained as the regression parameters estimation from the standard Poisson regression techniques assuming that the responses are independent.

(3) The initial values of the dispersion parameters estimates are obtained from the adjusted Pearson estimators without the bias-correction terms.

As long as the initial values are provided, the algorithm will then iterate.

The iterative procedures are shown below:

(1) Frailty predictors $(\hat{U}_i, \hat{U}_{ij})$ will be updated using the approach of orthodox BLUP (see Equation 3.11, 3.12). And the practical methods have been discussed before.

(2) The regression parameters estimates $\hat{\beta}$ will be updated through the approach of the Newton scoring algorithm (see Equation 3.14).

(3) The dispersion parameters estimates $(\hat{\sigma}^2, \hat{\tau}_j^2)$ will be updated through the approach of adjusted Pearson estimators (see Equation 3.18, 3.19). Iterative procedures continue until convergence.

Chapter 4

Data Analysis

The proposed nested frailty model for analyzing bivariate recurrent events is demonstrated by a motivating example described in Section 4.1. Then preliminary analysis has been applied using K-M curves. In Section 4.2, we analyze the PD data using our proposed model while treating the recurrent events as identical. In Section 4.3, we focus on the analysis of the bivariate recurrent events considering the order of the event occurrence. In Section 4.4, the simulation process is presented and the performance of the proposed model is evaluated.

4.1 Peritonitis Dialysis Data

4.1.1 Data Description

In our study, we use the peritoneal dialysis (PD) data, which are originally from Taichung Veterans General Hospital in Taiwan. During the period of Jan 1st 1996 and Dec 31st 2011, a total of 575 patients receiving PD were enrolled in this study. We obtained the PD data from Chen et al. (2015) but they could only provide a random sample with size 300 of the original data to the public. Therefore, our data contains 300 patients. Two types of events (Gram-positive and Non-Gram-positive peritonitis) are recorded on each patient over a time period, and some patients may experience one or both types of peritonitis more than once.

Some background information about peritonitis is introduced. Peritonitis is inflammation of the peritoneum, the lining of the inner wall of the abdomen that covers the abdominal organs (Asim, 2017; Li et al., 2016). Peritonitis is a major complication of PD. It can cause severe pain, swelling of abdomen, fever, hospitalization, technique failure, and even death (Poon et al., 2017). The clinical outcomes of peritonitis have been classified into two types: Gram-positive and Non-Gram-positive peritonitis (Eberl et al., 2016). Gram-positive peritonitis usually occurs because of touch contamination, catheter infection, poor skin care, etc (Piraino et al., 2011; Kotsanas et al., 2007). Non-Gram-positive may be due to peritoneal transport status, gastric acid inhibitors, chronic lung disease, etc (Chuang et al., 2008). Different repeated

Table 4.1: Peritoneal Dialysis Data Variables Description

Variable	Description
ID	Patients' ID (ID=1,2,...,300)
Event Type	Types of events (1: Gram-positive; 2: Non-Gram-positive)
Cannulate	Cannulate position for each patient (1 for the position in medium and 0 otherwise)
Age	The age in years of patients initiation
Start	The initial time 0 or the observed event time
End	The observed event time or the follow-up time
Count	The cumulative number of lines of each "Event Type" in each patient
Delta	Indicator of right censoring (1 if it is an event and 0 if it is right censoring)
Gender	The gender of patients (1 for male and 0 for female)
Diabetes	The diabetes status (1 if the patient has diabetes and 0 otherwise)
Vascular	The cardiovascular disease status (1: patient has cardiovascular disease and 0 otherwise)

Table 4.2: Peritoneal Dialysis Data Sample with Two Patients

ID	Event Type	Cannulate	Age	Start	End	Count	Delta	Gender	Diabetes	Vascular	Survival Time
1	1	0	38.06	0.000	0.049	1	1	1	0	1	0.049-0=0.049
1	1	0	38.06	0.049	0.140	2	1	1	0	1	0.140-0.049=0.091
1	1	0	38.06	0.140	0.236	3	1	1	0	1	0.236-0.140=0.096
1	1	0	38.06	0.236	0.403	4	0	1	0	1	0.403-0.236=0.167
1	2	0	38.06	0.000	0.403	1	0	1	0	1	0.403-0=0.403
2	1	0	60.17	0.000	2.033	1	0	1	1	1	2.033-0=2.033
2	2	0	60.17	0.000	2.033	1	0	1	1	1	2.033-0=2.033

events of peritonitis may have different implications in implementation of control strategies.

The primary goal is to present proper joint frailty survival model for bivariate recurrent peritonitis events and identify risk factors affecting peritonitis.

There are a total of 11 variables in the data. The detailed description of each variable can be found in Table 4.1.

The "Survival time" of each recurrent event on each patient has been calculated, indicating the time to each recurrent event. Specifically, "Survival time" equals "End" minus the corresponding "Start". Table 4.2 shows a sample of our peritoneal dialysis data on two patients, with ID=1 and ID=2, and the correspond "Survival time" is calculated.

4.1.2 Preliminary Analysis Using K-M Curves

Kaplan-Meier curves and estimates of survival data are the most familiar ways of dealing with differing survival times (times-to-event), especially when censorship occurs. In Figure 4.1 and Figure 4.2, we show the K-M survival curves comparing different groups for gram-positive and non-gram-positive peritonitis, respectively.

On the top of Figure 4.1 we can see that survival probabilities are close for diabetes and non-diabetes patients before a survival time of around 6. In the diabetes group, the survival probability reduces to 0 around a time of 7.74. In the non-diabetes group, after a survival time of 11.05, the survival probability does not change because of the censorship. The log rank test of the two curves shows them not to be significantly different (p -value=0.5).

The middle of Figure 4.1 shows the K-M curves of gram-positive peritonitis between vascular and non-vascular patients. The survival probability is larger for vascular patients than non-vascular before survival time of 7.74. The survival probability drops to 0 at time 7.74 for vascular patients. The log rank test indicates that vascular and non-vascular patients are significantly different (p -value=0.05).

The bottom of Figure 4.1 shows the K-M curves of gram-positive peritonitis between the cannulate position in the medium and non-medium. It graphically tells us that the survival probabilities for these two groups are close before a survival time of around 8. Then the survival probability drops to 0 at a time of 11.05 for the cannulate position in non-medium. The hori-

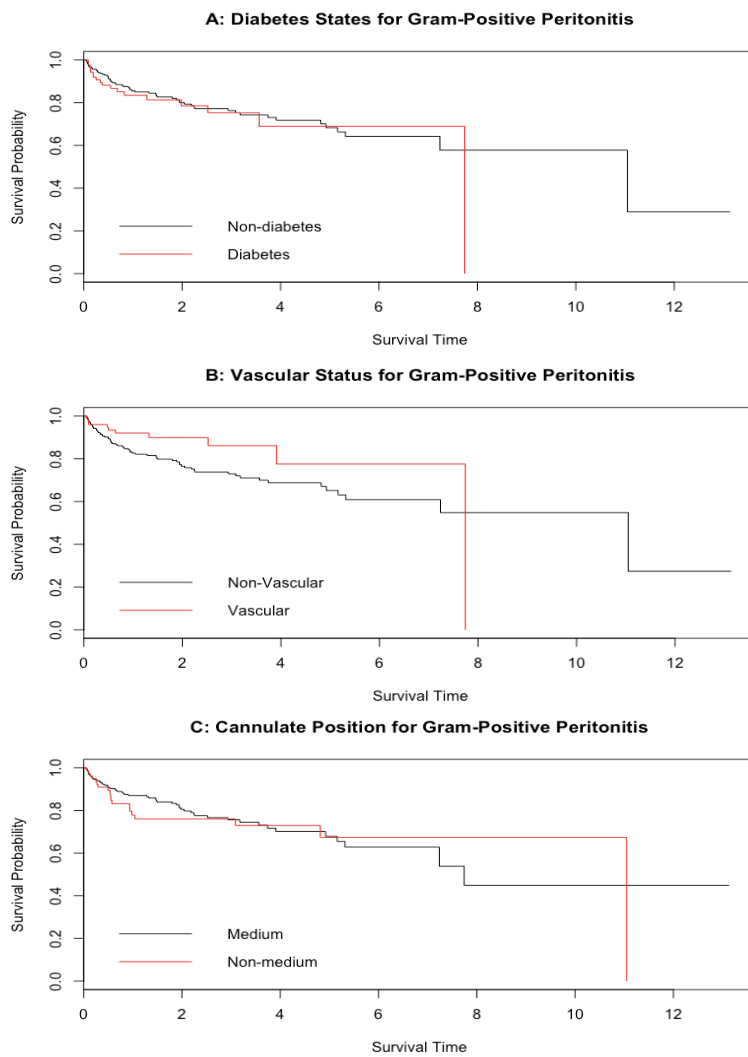


Figure 4.1: The K-M curves for Gram-Positive Peritonitis

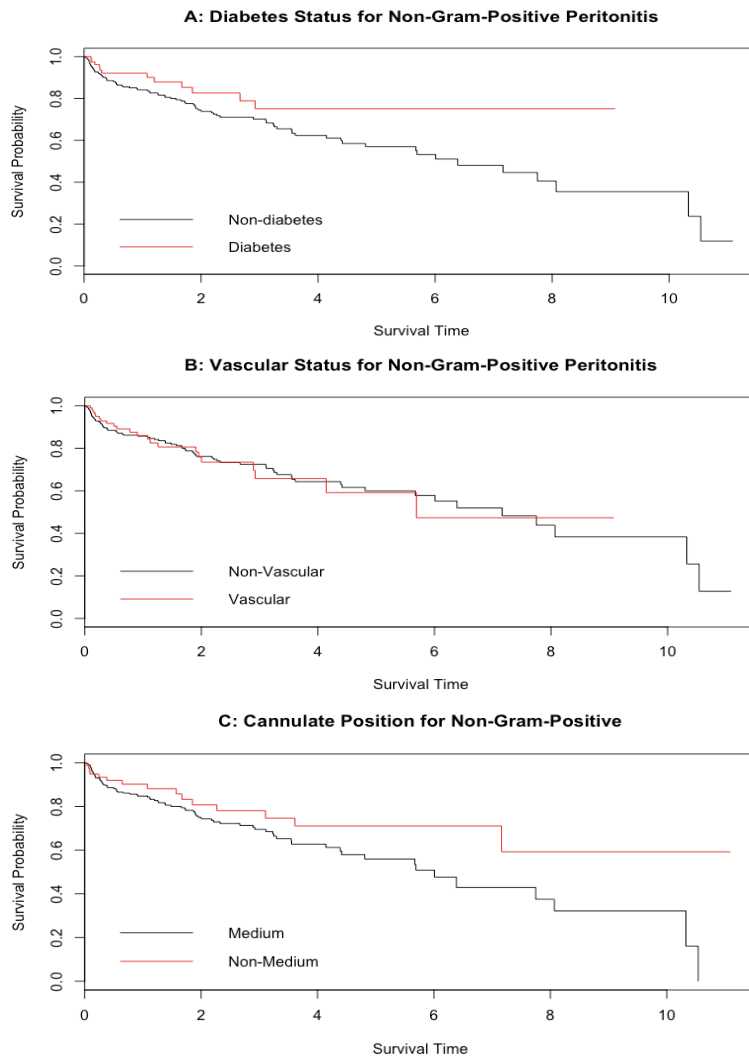


Figure 4.2: The K-M curves for Non-Gram-Positive Peritonitis

zontal line for the cannulate position in medium after a time of 8 represents censorship. The log rank test shows these two groups not to be significantly different (p -value=0.7).

The top of Figure 4.2 shows the K-M curves for non-gram-positive peritonitis between diabetes patients and non-diabetes. It graphically shows that the diabetes group has a higher survival probability than the non-diabetes group. The log rank test indicates them to be significantly different at the 10% level (p -value=0.08).

The middle of Figure 4.2 shows that the survival probabilities for non-gram-positive peritonitis are close between vascular and non-vascular patients. The log rank test confirms that the two groups are not significantly different (p -value=0.9).

The bottom of Figure 4.2 represents the K-M curves for non-gram-positive peritonitis between the cannulate position in medium and non-medium. It graphically shows the survival probability is lower for the cannulate position in medium than non-medium. The log rank test shows them to be significantly different at the 10% level (p -value=0.1).

4.2 Bivariate Recurrent Event Analysis Treating Recurrent Events as Identical

In this section, we propose a nested frailty Cox model involving the same assumptions and steps but regards the recurrent events as identical. In other

Table 4.3: Estimates and SEs for PD Treating Recurrent Events as Identical

Parameter	Estimate	SE	Z	P
Cannulate (a)	-0.003	0.312	-0.008	0.993
Diabetes (a)	0.146	0.325	0.449	0.653
Vascular (a)	-0.615	0.371	-1.660	0.097
Cannulate (b)	-0.188	0.311	-0.605	0.545
Diabetes (b)	-0.696	0.330	-2.111	0.035
Vascular (b)	-0.157	0.293	-0.538	0.591
σ^2	0.110			
τ^2 (a)	0.199			
$\tau^2(b)$	0.093			

Note: (a)s correspond to the parameters in Gram-positive peritonitis and (b)s correspond to Non-Gram-positive peritonitis.

words, the order of event occurrences is not considered to be important. The parameters estimates are listed in Table 4.3.

The p -values indicate that for the Gram-Positive event, none of the covariates are significant at the 5% level, but the effect of vascular is statistically significant at the 10% level. As for the Non-Gram-Positive, p -values indicate that the effect of diabetes is statistically significant at the 5% level, but effects of cannulate and vascular are not. We remove the covariates of cannulate in Gram-Positive peritonitis and vascular in the Non-Gram-Positive peritonitis since they have large p -values. The results for the model reduction are listed in Table 4.4.

The covariates of diabetes in Gram-Positive peritonitis and cannulate in Non-Gram-Positive peritonitis do not have significant effects (p -values are 0.655

Table 4.4: Model Reduction 1 for PD Treating Recurrent Events as Identical

Parameter	Estimate	SE	Z	P
Diabetes (a)	0.146	0.325	0.447	0.655
Vascular (a)	-0.609	0.368	-1.656	0.098
Cannulate(b)	-0.155	0.305	-0.508	0.611
Diabetes (b)	-0.700	0.330	-2.123	0.034
σ^2	0.111			
τ^2 (a)	0.198			
τ^2 (b)	0.094			

Note: (a)s correspond to the parameters in Gram-positive peritonitis and (b)s correspond to Non-Gram-positive peritonitis.

and 0.611 respectively); therefore they are removed from the model. The results for the further model reduction are listed in Table 4.5

The results in Table 4.5 show that the effect of vascular is significant for the Gram-Positive peritonitis (p -value=0.054). The hazard for patients with vascular disease is $\exp(-0.586)=0.557$ times the hazard for patients without vascular disease for the Gram-Positive peritonitis events. Non-vascular patients are somewhat more likely to have the Gram-Positive peritonitis.

The effect of diabetes has significant effect on Non-Gram-Positive peritonitis (p -value=0.017). The hazard for patients with diabetes is $\exp(-0.703)=0.495$ times the hazard for patients without diabetes. Patients with diabetes are somewhat less likely to have the Non-Gram-Positive peritonitis than non-diabetes patients.

Table 4.5: Model Reduction 2 for PD Treating Recurrent Events as Identical

Parameter	Estimate	SE	Z	P
Vascular (a)	-0.586	0.365	-1.608	0.054
Diabetes (b)	-0.703	0.330	-2.128	0.017
σ^2	0.113			
τ^2 (a)	0.205			
τ^2 (b)	0.094			

Note: (a)s correspond to the parameters in Gram-positive peritonitis and (b)s correspond to Non-Gram-positive peritonitis.

4.3 Bivariate Recurrent Event Analysis Consider the Order of Events Occurrence

In this section, we analyze the PD data using our proposed model. Two recurrent event types (Gram-positive and Non-Gram-positive peritonitis) are recorded on 300 patients. Each patient may experience a certain type of event multiple times, so there are 775 recurrent event observations in the PD data. For each event type, our data are stratified by 15-year age categories and gender. Within each stratum, the distinct failure time and risk set of k th recurrent event occurrence can be obtained. Therefore, Y of 0 and 1 can be generated for each observation in the k th event occurrence from a certain stratum. Detailed information of data transformation can be found in Section 3.2.1.

After transformation, there are 988 elements in the response variable. The α matrix of 0 and 1 is generated to track the stratum and order of event

Table 4.6: Estimates and SEs for PD Bivariate Recurrent Event Analysis

Parameter	Estimate	SE	Z	P
Cannulate (a)	-0.062	0.326	-0.189	0.850
Diabetes (a)	0.217	0.340	0.639	0.523
Vascular (a)	-0.714	0.380	-1.881	0.060
Cannulate (b)	-0.342	0.322	-1.060	0.289
Diabetes (b)	-0.739	0.337	-2.193	0.028
Vascular (b)	-0.206	0.296	-0.695	0.487
σ^2	0.209			
τ^2 (a)	0.296			
τ^2 (b)	0.114			

Note: (a)s correspond to the parameters in Gram-positive peritonitis and (b)s correspond to Non-Gram-positive peritonitis.

occurrence for the transformed elements. We use the R software to fit the proposed models and results are summarized in Table 4.6. The estimates and standard errors for regression parameters are presented, as well as the dispersion parameters estimation.

The p -values indicate that for the Gram-Positive event, none of the covariates are significant at the 5% level, but the effect of vascular is statistically significant at the 10% level (p -value=0.060). The hazard for patients with vascular disease is $\exp(-0.714)=0.490$ times the hazard for patients without vascular disease. Our results reveal that cardiovascular patients are less likely to have Gram-Positive peritonitis compared with non-cardiovascular patients. Possible explanations could be that to reduce the risk of heart attack or stroke on cardiovascular patients, medicines such as canakinumab are frequently

used (Ishani et al., 2004; Huet et al., 2017; Swirski et al., 2013; Ridker et al., 2008). The mechanism for treating cardiovascular disease is by reducing the number of inflammatory chemicals in the blood, and this may reduce the hazard of other inflammation such as Gram-Positive peritonitis (Kaptoge et al., 2013; Chiu and Mehrotra, 2010). Another possible reason could be our small data with 300 random patients, which may not fully represent the true case.

Chen et al. (2015) propose a flexible marginal model and use a two-stage estimation procedure for estimating unknown parameters using the same peritoneal dialysis data which consisted of 575 PD patients. Their research indicates the same results, that non-cardiovascular patients are more likely to have Gram-Positive peritonitis than cardiovascular patients ($\hat{\beta}=-0.4950$, although p -value=0.1183). Their explanation for this phenomenon is that patients with cardiovascular disease have higher risk for the terminal event (receiving permanent hemodialysis therapy or death), which precludes a further occurrence of Gram-Positive peritonitis. Thus, peritonitis rates for cardiovascular disease patients may be smaller than for non-cardiovascular patients. As for the Non-Gram-Positive, p -values indicate that the effect of diabetes is statistically significant at the 5% level (p -value = 0.028), but effects of cannulate and vascular are not. Results show that the hazard for patients with diabetes is $\exp(-0.739)=0.478$ times the hazard for patients without diabetes when other covariates are holding fixed. Our results reveal that non-diabetes patients are more likely to have Non-Gram-Positive peritonitis compared

Table 4.7: Model Reduction 1 for PD Bivariate Recurrent Event Analysis

Parameter	Estimate	SE	Z	P
Diabetes (a)	0.217	0.339	0.639	0.523
Vascular (a)	-0.694	0.376	-1.844	0.065
Cannulate(b)	-0.293	0.316	-0.930	0.352
Diabetes (b)	-0.736	0.337	-2.187	0.029
σ^2	0.213			
τ^2 (a)	0.291			
τ^2 (b)	0.115			

Note: (a)s correspond to the parameters in Gram-positive peritonitis and (b)s correspond to Non-Gram-positive peritonitis.

with diabetes patients. Patients without diabetes require more care in respect to Non-Gram-Positive peritonitis. Possible explanations could be that diabetes patients are always treated with insulin (Pittas et al., 2015; Oram et al., 2014). Insulin itself is anti-inflammatory, suppressing pro-inflammatory transcription factors such as NF-kappa β and activating protein 1, all mediators of inflammation (Hyun, 2011). This process may decrease other inflammations such as Non-Gram-Positive peritonitis (Quellhorst, 2002; King-Morris, 2012). A similar result could be found in Chen et al. (2015). Their work reveals that diabetes patients are somewhat less likely to have Non-Gram-Positive peritonitis compared with non-diabetes ($\hat{\beta}=-0.428$, p -value=0.13). Their research explains that patients with diabetes have higher risk for the terminal event, which precludes a further occurrence of peritonitis. Thus the Non-Gram-Positive peritonitis rate for diabetes may be smaller than for non-diabetes patients.

Table 4.8: Model Reduction 2 for PD Bivariate Recurrent Event Analysis

Parameter	Estimate	SE	Z	P
Vascular (a)	-0.658	0.372	-1.770	0.077
Diabetes (b)	-0.731	0.338	-2.166	0.030
σ^2	0.223			
τ^2 (a)	0.291			
τ^2 (b)	0.116			

Note: (a)s correspond to the parameters in Gram-positive peritonitis and (b)s correspond to Non-Gram-positive peritonitis.

We conduct the model reduction and remove the covariates of cannulate in the Gram-Positive event and vascular in the Non-Gram-Positive event since they have large p -values. Table 4.7 shows the results of our reduced model. The p -values indicate that the effect of vascular is statistically significant for Gram-Positive at the 10% level (p -value=0.065), but diabetes is not (p -value=0.523). As for the Non-Gram-Positive event, the effect of diabetes is statistically significant (p -value=0.029), but the cannulate is not (p -value=0.352). So insignificant covariates can be dropped from models. Table 4.8 shows the result with covariate of vascular for Gram-Positive event and diabetes for Non-Gram-Positive event. The effect of vascular disease is statistically significant for Gram-Positive event at the 10% level (p -value=0.077). The hazard for patients with vascular disease is $\exp(-0.658)=0.518$ times the hazard for patients without vascular disease for Gram-Positive event. Patients without vascular need to take extra care as the hazard is larger than for patients with vascular disease.

Table 4.8 shows that diabetes is significant at the 5% level for Non-Gram-Positive event (p -value=0.030). The hazard for patients with diabetes is $\exp(-0.731)=0.481$ times the hazard for patients without diabetes. Non-diabetes patients are somewhat more likely to have Non-Gram-Positive event than diabetes patients.

Similar results were reported by Chen et al. (2015) using two-stage estimation, stating that patients with vascular disease have lower risk of having the Gram-positive event than patients without vascular disease. In addition, patients with diabetes tend to have lower risk of having the Non-Gram-Positive event than patients without diabetes.

The estimates of dispersion parameters σ^2 is 0.223. The τ_j^2 are 0.291 and 0.116 for Gram-Positive and Non-Gram-Positive events, respectively. The dispersion parameters measure how dispersed those heterogeneities are. Specifically, the dispersion parameter of σ^2 shows the dispersion for heterogeneity between clusters and τ_j^2 shows the dispersion for event specific heterogeneity between clusters.

4.4 Simulation

In this section, we use the simulation method to evaluate the performance of our proposed model. In Section 4.4.1, we describe the simulation process and generate the data. In Section 4.4.2, we show the results of simulated data and evaluate our estimation.

4.4.1 Simulation Process

The process of simulation is described as follows:

Step 1: we assume that there are 100 subjects (or clusters). We generate the first-level frailties u_1, \dots, u_{100} following Gamma distribution with mean of 1 and variance of σ^2 , which can be expressed as

$$E(U_i) = 1 \quad \text{and} \quad \text{Var}(U_i) = \sigma^2.$$

Step 2: we assume there are two event types (or sub-clusters) for each subject as second-level frailties following Gamma distribution. Given the first-level frailties $U_* = (u_1, \dots, u_{100})$, the second-level frailties are conditionally independent. The mean and variance of the second-level frailties can be expressed as follows:

$$E(U_{ij} | U_*) = U_i \quad \text{and} \quad \text{Var}(U_{ij} | U_*) = \tau_j^2 U_i.$$

where $i = 1, \dots, 100$, and $j = 1, 2$

Step 3, we assume that there are 5 recurrent events for each event type on each subject. We then generate the response variable Y_{ij1}, \dots, Y_{ij5} for each event type on each subject following Poisson distribution. Given the first-level and second-level frailties, the components of Y can be shown as

$$Y_{ijk} \sim \text{Poisson}(u_{ij}\mu_{ijk})$$

where $\mu_{ijk} = \exp(\beta_j^T X)$. The X with three covariates is generated following normal distribution. In total, the simulation data are obtained by repeating step 1, 2 and 3 for 400 times.

The initial value of the dispersion parameter for the first-level frailty is fixed with $\sigma^2 = 0.11$. The initial values of the regression and dispersion parameters for the second-level frailty are also fixed. Specifically, for the first event type, we set $\beta_1 = (0.3, -0.2, 1.8)$ and $\tau_1^2 = 0.2$. For the second event type, we set $\beta_2 = (0.3, -0.6, 0.7)$ and $\tau_2^2 = 0.1$.

4.4.2 Simulation Performance

The average of the regression and dispersion parameter estimates over the 400 simulations are provided in Table 4.9. We can see that the regression parameters and dispersion parameter are almost the same as the true values, with very small standard errors for the former ones. This indicates that our estimates are accurate.

Table 4.9: Simulation Results

Parameter	True	Estimate	Bias	ESE	SSE
β_{a1}	0.300	0.301	0.001	0.048	0.048
β_{a2}	-0.200	-0.200	0.000	0.047	0.046
β_{a3}	1.800	1.800	0.000	0.060	0.062
β_{b1}	0.300	0.300	0.000	0.039	0.042
β_{b2}	-0.600	-0.602	-0.002	0.040	0.043
β_{b3}	0.700	0.701	0.001	0.040	0.041
σ^2	0.110	0.112	0.002		0.044
τ_1^2	0.200	0.191	-0.009		0.073
τ_2^2	0.100	0.097	-0.003		0.052

$\beta_{a1}, \beta_{a2}, \beta_{a3}$ correspond to the regression parameters for the first event ($j = 1$); $\beta_{b1}, \beta_{b2}, \beta_{b3}$ corresponds to the regression parameters for the second event ($j = 2$); τ_1^2 and τ_2^2 correspond to the dispersion parameter for the first and second event, respectively. Bias is calculated by Estimate - True; ESE denotes the average of estimated standard errors and SSE denotes the simulated standard errors.

Chapter 5

Discussion

5.1 Conclusion

It is common that in survival analysis, multivariate types recurrent events are measured for each subject. In this thesis, we propose the nested frailty Cox model for bivariate recurrent events using the Poisson modelling approach. Our model assumptions depend only on the first and second moments of the unobserved frailties. Therefore, our proposed models should be robust against mis-specification of the frailties distribution. We predict the frailties using the approach of orthodox best linear unbiased predictor. The proposed orthodox BLUP approach is computationally effective. In addition, we deal with the bivariate recurrent events and the extension to multivariate recurrent events is very straightforward.

The proposed models for analyzing survival data with bivariate recurrent

events is demonstrated by a motivating peritonitis dialysis study example. Results and discussions are given in Chapter 4. Simulation data is also generated and used to evaluate the performance of proposed model framework and estimation methods. Results indicate that our proposed models perform well.

Our result reveals that, patients with cardiovascular disease are less likely to have the Gram-Positive peritonitis compared with non-cardiovascular patients. Possible explanations could be: the mechanism for treating cardiovascular disease is by reducing the number of inflammatory chemicals in the blood. This procedure may reduce the hazard of other inflammation such as Gram-Positive peritonitis (Kaptoge et al., 2013; Chiu and Mehrotra, 2010). Our results also show that the non-diabetes patients are more likely to have Non-Gram-Positive peritonitis compared with diabetes patients. Possible explanations could be that diabetes patients are always treated with insulin (Pittas et al., 2015; Oram et al., 2014), which itself is anti-inflammatory. Medicine such as insulin may decrease inflammations such as the Non-Gram-Positive peritonitis (Quellhorst, 2002; King-Morris, 2012). Similar result could be found in Chen et al. (2015).

5.2 Further Study

There are several areas that we would like to examine in further research. First of all, our proposed method incorporates two levels of nested frailties

into Cox survival model. In the future work, we can use our approach as a starting point and incorporate multivariate frailties into the survival model. Secondly, in future work, Wald's test can be used for testing the hypothesis H_0 that regression parameter estimates equal 0.

In addition, neither the standard error nor the significance are provided for the dispersion parameters σ^2 and τ^2 . It is of a great interest to develop suitable testing approaches for their significance. In further research, bootstrap approach may be adopted to deal with this problem (Chatterjee 2005).

Bibliography

- [1] Aalen, O.O. and Husebye, E., 1991. Statistical analysis of repeated events forming renewal processes. *Statistics in medicine*, 10(8), pp.1227-1240.
- [2] Abu-Libdeh, H., Turnbull, B.W. and Clark, L.C., 1990. Analysis of multi-type recurrent events in longitudinal studies; application to a skin cancer prevention trial. *Biometrics*, pp.1017-1034.
- [3] Akritas, M.G., 2004. Nonparametric survival analysis. *Statistical Science*, pp.615-623.
- [4] Andersen, P.K., Klein, J.P., Knudsen, K.M. and y Palacios, R.T., 1997. Estimation of variance in Cox's regression model with shared gamma frailties. *Biometrics*, pp.1475-1484.
- [5] Andersen, P.K. and Gill, R.D., 1982. Cox's regression model for counting processes: a large sample study. *The annals of statistics*, pp.1100-1120.

- [6] Asim, M., 2017. Eosinophilic peritonitis in a continuous ambulatory peritoneal dialysis patient: Inflammation and irritation without infection. *Saudi Journal of Kidney Diseases and Transplantation*, 28(2), p.401.
- [7] Bewick, V., Cheek, L. and Ball, J., 2004. Statistics review 12: survival analysis. *Critical care*, 8(5), p.389.
- [8] Breslow, N.E. and Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421), pp.9-25.
- [9] Brockwell, Peter J., and Richard A. Davis. *Time series: theory and methods*. Springer Science and Business Media, 2013.
- [10] Box-Steffensmeier, J.M. and De Boef, S., 2006. Repeated events survival models: the conditional frailty model. *Statistics in medicine*, 25(20), pp.3518-3533.
- [11] Cai, J. and Schaubel, D.E., 2004. Marginal means/rates models for multiple type recurrent event data. *Lifetime data analysis*, 10(2), pp.121-138.
- [12] Clegg, L.X., Cai, J. and Sen, P.K., 1999. A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics*, 55(3), pp.805-812.
- [13] Chen, B.E. and Cook, R.J., 2004. Tests for multivariate recurrent events in the presence of a terminal event. *Biostatistics*, 5(1), pp.129-143.

- [14] Chen, C.M., Chuang, Y.W. and Shen, P.S., 2015. Two-stage estimation for multivariate recurrent event data with a dependent terminal event. *Biometrical Journal*, 57(2), pp.215-233.
- [15] Chatterjee, S. and Bose, A., 2005. Generalized bootstrap for estimating equations. *The Annals of Statistics*, 33(1), pp.414-436.
- [16] Chuang, Y.W., Shu, K.H., Yu, T.M., Cheng, C.H. and Chen, C.H., 2008. Hypokalaemia: an independent risk factor of Enterobacteriaceae peritonitis in CAPD patients. *Nephrology Dialysis Transplantation*, 24(5), pp.1603-1608.
- [17] Chiu, Y.W. and Mehrotra, R., 2010. Can we reduce the cardiovascular risk in peritoneal dialysis patients?. *Indian journal of nephrology*, 20(2), p.59.
- [18] Cox, D.R., 1972. The statistical analysis of dependencies in point processes. *Stochastic point processes*, 55, p.66.
- [19] Cox, D.R. and Oakes, D., 1984. *Analysis of survival data* (Vol. 21). CRC Press.
- [20] Cox, D.R., 2018. *Analysis of survival data*. Routledge.
- [21] Cook, R.J. and Lawless, J., 2007. *The statistical analysis of recurrent events*. Springer Science and Business Media.

- [22] Cook, R. J., Ng, E., Mukherjee, J., and Vaughn, D., (1999). Two-state mixed renewal processes for chronic disease. *Statistics in Medicine* 18, 175-188.
- [23] Crowder, M., 1986. On consistency and inconsistency of estimating equations. *Econometric Theory*, 2(3), pp.305-330.
- [24] Crowder, M., 1987. On linear and quadratic estimating functions. *Biometrika*, 74(3), pp.591-597.
- [25] David, Cox R. "Regression models and life tables (with discussion)." *Journal of the Royal Statistical Society* 34 (1972): 187-220.
- [26] Eberl, M., Topley, N. and Lin, C.Y., Cardiff University College Consultants Ltd, 2016. Method for identifying a bacterial infection. U.S. Patent Application 14/775,738.
- [27] Elashoff, R.M., Li, G. and Li, N., 2008. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*, 64(3), pp.762-771.
- [28] Haight, F.A., 1967. *Handbook of the Poisson distribution*.
- [29] He, X., Tong, X., Sun, J. and Cook, R.J., 2007. Regression analysis of multivariate panel count data. *Biostatistics*, 9(2), pp.234-248.
- [30] Hosmer, D.W., Lemeshow, S. and May, S., 2011. *Applied survival analysis*. Wiley Blackwell.

- [31] Huet, F., Akodad, M., Fauconnier, J., Lacampagne, A. and Roubille, F., 2017. Anti-inflammatory drugs as promising cardiovascular treatments. *Expert review of cardiovascular therapy*, 15(2), pp.109-125.
- [32] Hyun, E., Ramachandran, R., Hollenberg, M.D. and Vergnolle, N., 2011. Mechanisms behind the anti-inflammatory actions of insulin. *Critical Reviews in Immunology*, 31(4).
- [33] Ibrahim, J.G., Chu, H. and Chen, L.M., 2010. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16), p.2796.
- [34] Ibrahim, J.G., Chen, M.H. and Sinha, D., 2005. Bayesian survival analysis. John Wiley and Sons, Ltd.
- [35] Ishani, A., Herzog, C.A., Collins, A.J. and Foley, R.N., 2004. Cardiac medications and their association with cardiovascular events in incident dialysis patients: cause or effect?. *Kidney international*, 65(3), pp.1017-1025.
- [36] Jorgensen, B., LUNDBYE?CHRISTENSEN, S..R.E.N., SONG, X.K. and Sun, L., 1996. A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia. *Statistics in medicine*, 15(7?9), pp.823-836.
- [37] Jorgensen, B., 1997. The theory of dispersion models. Chapman and Hall. London, UK.

- [38] Jorgensen, B., 1987. Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society B*, 49, pp.127-162.
- [39] Kaplan, Edward L., and Paul Meier. "Nonparametric estimation from incomplete observations." *Journal of the American statistical association* 53.282 (1958): 457-481.
- [40] Kaptoge, S., Seshasai, S.R.K., Gao, P., Freitag, D.F., Butterworth, A.S., Borglykke, A., Di Angelantonio, E., Gudnason, V., Rumley, A., Lowe, G.D. and Jrgensen, T., 2013. Inflammatory cytokines and risk of coronary heart disease: new prospective study and updated meta-analysis. *European heart journal*, 35(9), pp.578-589.
- [41] Klein, J.P., 1992. Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, pp.795-806.
- [42] Klein, J.P. and Moeschberger, M.L., 2005. *Survival analysis: techniques for censored and truncated data*. Springer Science and Business Media.
- [43] Kiefer, N.M., 1988. Economic duration data and hazard functions. *Journal of economic literature*, 26(2), pp.646-679.
- [44] King-Morris, K. and Ikizler, T.A., 2012. *Insulin Resistance in Patients Undergoing Peritoneal Dialysis: Can We Improve It?*.
- [45] Kotsanas, D., Polkinghorne, K.R., Korman, T.M., Atkins, R.C. and Brown, F., 2007. Risk factors for peritoneal dialysis - related peritonitis:

- Can we reduce the incidence and improve patient selection?. *Nephrology*, 12(3), pp.239-245.
- [46] Lane, W.R., Looney, S.W. and Wansley, J.W., 1986. An application of the Cox proportional hazards model to bank failure. *Journal of Banking and Finance*, 10(4), pp.511-531.
- [47] Langova, Katerina. "Survival analysis for clinical studies." *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 152.2 (2008): 303-307.
- [48] Laird, N. and Olivier, D., 1981. Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, 76(374), pp.231-240.
- [49] Lawless, J.F. and Nadeau, C., 1995. Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37(2), pp.158-168.
- [50] Lee, Y. and Nelder, J.A., 1996. Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.619-678.
- [51] Li, P.K.T., Szeto, C.C., Piraino, B., de Arteaga, J., Fan, S., Figueiredo, A.E., Fish, D.N., Goffin, E., Kim, Y.L., Salzer, W. and Struijk, D.G., 2016. ISPD peritonitis recommendations: 2016 update on prevention and treatment. *Peritoneal Dialysis International*, 36(5), pp.481-508.
- [52] Lin, Haiqun, and Daniel Zelterman. "Modeling survival data: extending the Cox model." (2002): 85-86.

- [53] Lin, D.Y., Wei, L.J., Yang, I. and Ying, Z., 2000. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4), pp.711-730.
- [54] Lin, D.Y., Wei, L.J. and Ying, Z., 2001. Semiparametric transformation models for point processes. *Journal of the American Statistical Association*, 96(454), pp.620-628.
- [55] Ma, R., 1999. An orthodox BLUP approach to generalized linear mixed models (Doctoral dissertation, University of British Columbia).
- [56] MA, R., WILLMS, J.D. and BURNETT, R.T., 2002. A nested frailty survival model for recurrent events. In *Recent Advances In Statistical Methods* (pp. 186-197).
- [57] Ma, Renjun, Daniel Krewski, and Richard T. Burnett. "Random effects Cox models: a Poisson modelling approach." *Biometrika* 90.1 (2003): 157-169.
- [58] Ma, R. and Jrgensen, B., 2007. Nested generalized linear mixed models: an orthodox best linear unbiased predictor approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4), pp.625-641.
- [59] McGilchrist, C. A. "Estimation in generalized mixed models." *Journal of the Royal Statistical Society. Series B (Methodological)* (1994): 61-69.

- [60] Miller Jr, R.G., 2011. Survival analysis (Vol. 66). John Wiley and Sons.
- [61] Narendranathan, W. and Stewart, M.B., 1991. PRACTITIONERS'CORNER: Simple Methods for Testing for the Proportionality of Cause - Specific Hazards in Competing Risks Models. Oxford Bulletin of Economics and Statistics, 53(3), pp.331-340.
- [62] Ng, E. T. M. and Cook, R. J. (1997). Modeling two-state disease processes with random effects. Lifetime Data Analysis 3, 315-335.
- [63] Ng, E. T. M. and Cook, R. J. (1999). Robust inference for bivariate point processes. Canad. J. Statist. 27, 509-524.
- [64] Ning, J., Chen, Y., Cai, C., Huang, X. and Wang, M.C., 2015. On the dependence structure of bivariate recurrent event processes: inference and estimation. Biometrika, 102(2), pp.345-358.
- [65] O'Quigley, J., 2008. Proportional hazards regression (Vol. 542). New York: Springer.
- [66] Oram, R.A., Jones, A.G., Besser, R.E., Knight, B.A., Shields, B.M., Brown, R.J., Hattersley, A.T. and McDonald, T.J., 2014. The majority of patients with long-duration type 1 diabetes are insulin microsecretors and have functioning beta cells. Diabetologia, 57(1), pp.187-191.
- [67] Piraino, B., 2011. Approaches to preventing peritonitis based on organism-specific rates. Peritoneal Dialysis International, 31(6), pp.636-638.

- [68] Pittas, A.G., Westcott, G.P. and Balk, E.M., 2015. Efficacy, safety, and patient acceptability of Technosphere inhaled insulin for people with diabetes: a systematic review and meta-analysis. *The lancet Diabetes and endocrinology*, 3(11), pp.886-894.
- [69] Poon, P.Y.K., Lan, H.Y., Kwan, B.C.H., Huang, X.R., Chow, K.M., Szeto, C.C. and Li, P.K.T., 2017. Peritoneal inflammation and fibrosis in C?reactive protein transgenic mice undergoing peritoneal dialysis solution treatment. *Nephrology*, 22(2), pp.125-132.
- [70] Prentice, R.L., Williams, B.J. and Peterson, A.V., 1981. On the regression analysis of multivariate failure time data. *Biometrika*, 68(2), pp.373-379.
- [71] Quellhorst, E., 2002. Insulin therapy during peritoneal dialysis: pros and cons of various forms of administration. *Journal of the American Society of Nephrology*, 13(suppl 1), pp.S92-S96.
- [72] Ratcliffe, S.J., Guo, W. and Ten Have, T.R., 2004. Joint modeling of longitudinal and survival data via a common frailty. *Biometrics*, 60(4), pp.892-899.
- [73] Ridker, P.M., Danielson, E., Fonseca, F.A., Genest, J., Gotto Jr, A.M., Kastelein, J.J., Koenig, W., Libby, P., Lorenzatti, A.J., MacFadyen, J.G. and Nordestgaard, B.G., 2008. Rosuvastatin to prevent vascular

- events in men and women with elevated C-reactive protein. *New England Journal of Medicine*, 359(21), p.2195.
- [74] Sargent, D.J., 1998. A general framework for random effects survival analysis in the Cox proportional hazards setting. *Biometrics*, pp.1486-1497.
- [75] Schaubel, D.E. and Cai, J., 2005. Analysis of clustered recurrent event data with application to hospitalization rates among renal failure patients. *Biostatistics*, 6(3), pp.404-419.
- [76] Schaubel, D.E., Zeng, D. and Cai, J., 2006. A semiparametric additive rates model for recurrent event data. *Lifetime Data Analysis*, 12(4), pp.389-406.
- [77] Sharef Moreno, E.S., 2008. *Nonparametric Frailty Models for Clustered Survival Data*.
- [78] Sinha, D. and Dey, D.K., 1997. Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, 92(439), pp.1195-1212.
- [79] Sun, L., Zhu, L. and Sun, J., 2009. Regression analysis of multivariate recurrent event data with time-varying covariate effects. *Journal of Multivariate Analysis*, 100(10), pp.2214-2223.

- [80] Swirski, F.K. and Nahrendorf, M., 2013. Leukocyte behavior in atherosclerosis, myocardial infarction, and heart failure. *Science*, 339(6116), pp.161-166.
- [81] Tsiatis, A.A. and Davidian, M., 2004. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, pp.809-834.
- [82] Therneau, T.M. and Grambsch, P.M., 2013. Modeling survival data: extending the Cox model. Springer Science and Business Media.
- [83] Yau, K.K., 2001. Multilevel models for survival analysis with random effects. *Biometrics*, 57(1), pp.96-102.
- [84] Wei, L.J., Lin, D.Y. and Weissfeld, L., 1989. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American statistical association*, 84(408), pp.1065-1073.
- [85] Whitehead, J., 1980. Fitting Cox's regression model to survival data using GLIM. *Applied Statistics*, pp.268-275.
- [86] Wienke, A., 2010. Frailty models in survival analysis. CRC Press.
- [87] Wu, L., Liu, W., Yi, G.Y. and Huang, Y., 2012. Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 2012.

- [88] Wulfsohn, M.S. and Tsiatis, A.A., 1997. A joint model for survival and longitudinal data measured with error. *Biometrics*, pp.330-339.
- [89] Ye, Y., Kalbfleisch, J., and Schaubel, D. E., (2007). Semi-parametric analysis for correlated recurrent and terminal events. *Biometrics* 63, 78-87.
- [90] Zhu, L., Sun, J., Tong, X. and Srivastava, D.K., 2010. Regression analysis of multivariate recurrent event data with a dependent terminal event. *Lifetime data analysis*, 16(4), pp.478-490.

Vita

Candidate's full name: Jiejie Wang

Master of Science, Forestry & Environmental Management, UNB, 2017

Bachelor of Science, Forestry, Nanjing Forestry University, 2013

Publication:

- Wang, J.J., Weng, Y.H., Krasowski, M., Yan, G.H. and Fullarton, M., 2017. Genetic parameters of growth and stem forking for black spruce progeny tested in New Brunswick, Canada. *New Forests*, pp.1-13.

Conference Presentation:

- Wang, J.J., Weng, Y.H., Krasowski, M. (2015). Genetic parameters of growth and stem forking for black spruce in first-generation progeny tests in New Brunswick. 34th Canadian Forest Genetics Association