Flexible Methods for the Analysis of Clustered Event Data in Observational Studies

by

Lili Wang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2020

Doctoral Committee:

Professor Douglas E. Schaubel, Co-Chair, University of Pennsylvania
Professor Peter X.K. Song, Co-Chair
Research Professor Mousumi Banerjee
Research Associate Professor Kevin He
Professor Richard A. Hirth

Lili Wang

lilywang@umich.edu

ORCID iD: 0000-0003-4276-3930

ACKNOWLEDGEMENTS

I came to the University of Michigan as a graduate student in Molecular Biology. One year later, I started taking Biostat 601, taught by Prof. Yi Li, and Bistat 602 by Prof. Bin Nan, both from the Department of Biostatistics, a major that I did not know when I was in China. I wanted to thank my research advisor at that time, Prof. James Bardwell, who kindly allowed me to take courses with freedom and generously supported my later decision to change to the Biostatistics MS program. I remember he said to me, "If there is a decision haunting in your mind for a year, then go for it." I should admit that the two biostatistics courses I took then were terrific and made up my mind to pursue my career in that field. Meanwhile, I was so fortunate that I started getting a consistent helping hand from Nicole Fenech, our program coordinator of the Biostatistics Department.

I started working hard on my coursework to lay a firm foundation in statistics and, very importantly, get some financial support. I enjoyed a lot in the courses like Biostat 650 taught by Prof. Brisa Sánchez, Biostat 801 by Prof. Bin Nan, Biostat 802 by Prof. Peter Song, Biostat 615 by Prof. Jian Kang, Biostat 653 and Biostat 880 by Prof. Lu Wang, Biostat 651 by Associate Prof. Shawn Lee, special-topic Biostat 830 by Assistant Prof. Peisong Han, and Biostat 875 from Prof. Douglas Schaubel. Note that I put together many courses I took for both my master's and Ph.D. degrees. To point out, Peisong continuously helped me on my second project even years after we finished the course. Jian also kindly gave me a lot of helpful suggestions on the COVID-19 project within the Bayesian framework.

My academic advisor then, Prof. Douglas Schaubel, who later became my thesis advisor, kindly offered me a summer research position to work on a survival project. Doug patiently trained me and helped me understand the statistical part of the project and its clinical background. I started to learn from him how to write and run the programs for survival analysis. I enjoyed working with him, which made me decide to invite him to be my thesis advisor when the Ph.D. program admitted me. For the past five years working with him, Doug never stopped encouraging me to think independently and creatively. He patiently and thoughtfully discussed the ideas I came up with, no matter how silly they were. Later I realized that the intent was to help me reform my “naive” ideas into well-designed research projects, which largely stimulated my research passion. In sum, he helps me not only with my research projects but also to build my confidence and independence. In Chinese, we say, “Give a man a fish, and you feed him for a day; teach a man to fish, and you feed him for a lifetime.” Doug is the person who taught me to fish and helped me catch the fish. Thank you so much, Doug!

After I finished the summer research, Doug referred me to a research team of professionals trained in different backgrounds at the Kidney, Epidemiology, and Cost Center (KECC). I want to thank the project investigator Prof. Richard Hirth for his trust and offering me the opportunity to work on the research project, which since then, I have been serving for five years. It is also so kind of him to agree to serve as a cognate member of my committee. In KECC, I got to know a giant in survival analysis, Prof. Jack Kalbfleisch. Jack never makes me feel that there is any distance between us. I especially want to thank him for the consistent guidance on my research work. He is always willing to help students who reach out to him in the department. He answered many of my questions, scheduled meetings with me, and lent me books. Besides, I have been working closely with Research Associate Prof. Kevin He, and his close bond with students much touched me. He spends much time studying and working with students and is always willing to help. Moreover, I would like to especially thank Prof. Yi Li for his continuing assistance with a methodology paper we are writing.

When approaching the end of my third year, I wanted to push my research boundary further. I was impressed by the excellent research work on longitudinal data analysis and data integration in Prof. Peter Song's group. With a strong motivation to work on these methodology projects, I asked him to advise me after Doug left for his new position at the University of Pennsylvania. Peter generously accepted my invitation and kindly took responsibility. I appreciate it a lot that I can work with one of the most productive and talented teams of students and learn from them. Peter closely follows a weekly meeting to ensure my research progress by giving me timely advice. I also got the chance to work on the exciting COVID-19 project and published my first discussion paper. All in all, the one-year research is so fruitful that it makes me a more knowledgeable researcher and better-prepared problem solver. Thank you a lot, Peter!

Last year, I heard the moving story from Research Prof. Mousumi Banerjee about her experience of being a biostatistician and a musician simultaneously. As a neighbor to her office, I got the chance to chat with her several times on statistics, COVID-19 and Indian folk music. I want to thank her for being such a valuable committee member with her unceasing encouragement on my COVID-19 project and a lot of great questions and suggestions, as well as her CDs and melodious voice.

Prof. Michael Boehnke chatted with me four years ago, which made me decide to accept the offer from our department as a Ph.D. student. This decision turned out to be one of the best I have made for the past several years.

Last but not least, I would love to thank my family. I want to thank my dad, who is such a persistent leaner that never stops learning. He told me that the real graduation certificate would be the person's death certificate. Moreover, my dad encouraged me to think critically. His open-mindedness supports my decision to pursue my career in Biostatistics, although he disliked the decision a lot at the beginning. My mom is a tolerant person and never pushes me. She makes me optimistically believe that there are always alternative options in life. Thank you both for being such wonderful parents and consistently trusting me. My husband is hardworking and wrote his motto as “Think hard and work smart.” I enjoy so much discussing with him about our research and debating on soccer. Thank you for accompanying me during those struggling times and being a patient listener. I love you all.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Clustered event data are frequently encountered in observational studies. In this dissertation, I am focusing on correlated event outcomes clustered by subjects (multivariate events), facilities, and both hierarchically.

The main approaches to analyzing correlated event data include frailty models with random effects and marginal models with robust variance estimation. Difficulties for the existing methods include a) computational demands and speed in the presence of numerous clusters (e.g., recurrent events); b) lacking rigorous diagnostic tools to prespecify the distribution of the random effects; c) analyzing a multi-state model that follows a semi-Markov renewal process. The growing need for flexible, computationally fast, and accurate estimating approaches to analyzing clustered event data motivates my methodological exploration in the following chapters.

In Chapter II, I propose a log-normal correlated frailty model to analyze recurrent event incidence rates and duration jointly. The regression parameters are estimated through a penalized partial likelihood, and the variance-covariance matrix of the frailty is estimated via a recursive estimating formula. The proposed methods are more flexible and faster than existing approaches and have the potential to be extended to other frequently encountered data structures (e.g., joint modeling with longitudinal outcomes).

In Chapter III, I propose a class of semiparametric frailty models that leave the distribution of frailties unspecified. Parameter estimation proceeds through estimating equations derived from first- and second-moment conditions. Estimation techniques

have been developed for three different models, including a shared frailty model for a single event; a correlated frailty model for multiple events; and a hierarchically structured nested failure time model. Extensive simulation studies demonstrate that the proposed approach can accurately estimate the regression parameters, baseline event rates, and variance components. Moreover, the computation time is fast, permitting application to very large data sets.

In Chapter IV, I develop a class of multi-state rate models to study the association of exposure to lead, a major endocrine disruptive agent, with behavioral changes captured by accelerometer measurements from wearable device ActiGraph GT3X. Categorized from personal activity counts over time by validated cutoffs, activity states are defined and analyzed through their in-state transitions using the proposed multi-state rate models in which the baseline rates are estimated nonparametrically. The proposed models combine the advantage of regular event rate models with the concept of competing risks, allowing to incorporate a daily renewal property and share baselines in the activity transition rates across different days. The regression parameters are specified in the event rate functions, leading to a semiparametric modeling framework. Statistical inference is based on a robust sandwich variance estimator that accounts for correlations between different event types and their recurrences. I found that the evaluated exposure to lead is associated with an increased transition from low activity to vigorous activity.

Chapter V is a special project of modeling the COVID-19 surveillance data in China, in which I develop two extended susceptible-infected-recovered (SIR) state-space models under a Bayesian state-space model framework. I propose to include a time-varying transmission rate or a time-dependent quarantine process in the classical SIR model to assess the effectiveness of macro-control measures issued by the government to mitigate the pandemic. The proposed compartment models enable to predict both short-term and long-term prevalence of the COVID-19 infection with

quantification of prediction uncertainty. I provide and maintain an open-source R package on GitHub (lilywang1988/eSIR) for the developed analytics.

CHAPTER I

Introduction

Survival outcomes are frequently encountered in health studies and clinical trials. Event data are very commonly clustered in observational studies, i.e., event times from the same cluster are correlated for some observed or unobserved reasons. Examples include the natural genetic similarity of people from the same family; patients treated in same hospital experiencing similar responses due to having been subject to similar treatment practices, and, a series of multivariate events from the same subject. Recurrent events are homogeneous, ordered, multivariate failure time data, which are naturally clustered within subjects. Many methods exist for modeling recurrent event data, commonly with regard to either total times (time since the origin) or gap (inter-event) times.

Andersen and Gill (1982) introduced the Cox (*Cox*, 1972, 1975) type model for recurrent event intensities based on the event increments given their past histories (filtration) and leaving the transition of the changing at-risk set unspecified. Their intensity model for multivariate events follows $E(dN^*(t)|\mathcal{F}_{t^-}) = \lambda_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{Z}(t))dt$, where $dN^*(t)$ denotes the instantaneous increment $N^*(t) - N^*(t^-)$ within a transient time interval $[t - dt, t]$ with $dt \to 0$. Suppose that C denotes the censoring time and $Y(t) = I(t \leq C)$ denotes the at-risk process, which implies that after C the subject is no longer at risk. As follows, the observed event process is $N(t) = \int_0^t Y(s)dN^*(s)$

or $dN(t) := Y(t)dN^*(t)$ with its increment at time t to be $dN(t) = N(t) - N(t^-)$. Let $\mathcal{F}_t$ be the filtration, which is usually a σ-field generated by the event and at-risk processes as well as the covariate, i.e. $\mathcal{F}_t := \{N(s), Y(s^+), \boldsymbol{Z}(s^+) : 0 \leq s \leq t\}$. The Cox-type models also implicitly assume that $E(dN^*(t)|\mathcal{F}_{t^-}) = E(dN^*(t)|\boldsymbol{Z}(t))$, or given $\boldsymbol{Z}(t)$, the intensity formula is independent of other historical events and covariates from $\mathcal{F}_{t^-}$. If $\boldsymbol{Z}(t)$ is time-invariant with $\boldsymbol{Z}(t) = \boldsymbol{Z}$, the Cox-type counting process can be reduced to a non-stationary Poisson process, i.e., the event increments from exclusive intervals are independent given $\boldsymbol{Z}$ (*Lin et al.*, 2000). Based on the powerful martingale theory, *Andersen and Gill* (1982) developed a series of large sample properties for the multivariate event *intensity* estimation via the partial likelihoods (*Cox*, 1975). Note that the event intensity here is in correspondence to the event *hazard* under the framework of modeling univariate event data (e.g., deaths). Subsequently, *Pepe and Cai* (1993), *Lawless and Nadeau* (1995) and *Lin et al.* (2000) suggested a proportional *rate* model $E(dN^*(t)|\boldsymbol{Z}(t)) = \lambda_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{Z}(t))$, which is equivalent to marginalizing the transitory intensities over its event history. Instead of using martingale theory, empirical process theory was employed to establish large sample properties.

There are a series of inspiring contributions on intensity models. *Prentice et al.* (1981) exploited a class of conditional models which allow the baseline intensities and regression parameters to vary between ordered events, $E(dN^*(t) \mid \mathcal{F}_{t^-}) = E(dN^*(t)|N^*(t^-) = k-1, \boldsymbol{Z}(t)) = \lambda_{0k}(t)\exp(\boldsymbol{\beta}_k'\boldsymbol{Z}(t))$, which can be estimated with a stratified Cox model. *Prentice et al.* (1981) also proposed a model for gap times as a special case of semi-Markov process. *Wei et al.* (1989) avoided dealing with the dependence between events by fitting separate marginal models $E(dN_k^*(t)|\mathcal{F}_k(t^-)) = \lambda_{0k}(t)\exp(\boldsymbol{\beta}_k'\boldsymbol{Z}(t))$, solely conditional on the kth event history $\mathcal{F}_k(t^-)$. Note that, in this marginal model, subjects are at risk for event k even before event $k-1$ occurs. A drawback of using such total event times in the presence of recurrent event data is the "carry-over effect", i.e., the prior event times can influence the future event times since they are cumulative.

To overcome such carry-over effect, practitioners sometimes prefer using gap times. There are mainly two statistical problems that hinder the use of gap times for multivariate event analysis: 1) the marginal distribution of gap times other than the first one are always not identifiable because their observations are dependent on the fact that they are uncensored, or $T_k^* \leq C$ (here again superscript T^* indicates true event times subject to right censoring C and its subscript k denotes the event order), and 2) induced dependent censoring (*Wang*, 1999; *Huang*, 2002; *Schaubel and Cai*, 2004b). To be specific, with the gaps times denoted by $\widetilde{T}_k^* := T_k^* - T_{(k-1)}^*$, the observation of $\widetilde{T}_k^*$ is always conditional on the fact that $T_{(k-1)}^*$ has been observed. Moreover, we have dependent censoring for gap times after the first event, i.e., $\widetilde{T}_k^* \not\perp C - \sum_{j=1}^{(k-1)} \widetilde{T}_j^*$. Both problems can be solved when the gap times are independent of each other, which is mostly an untenable assumption.

A variety of approaches for gap-time data analysis have been established. They mainly follow the subsequent several strategies, including 1) using conditional distribution of the current gap time given prior events, 2) introducing random effects to account for unobserved associations between events via shared or correlated random effects, 3) modeling survival functions of gap times using Copula with some association parameters, 4) marginal models with robust inference to account for dependent gap times. In addition, one can directly estimate the gap time distributions marginally through some nonparametric methods which have been extensively studied in a rich literature (*Visser*, 1996; *Wang and Wells*, 1998; *Wang*, 1999; *Wang and Chang*, 1999; *Lin et al.*, 1999; *Huang*, 2000, 2002; *Schaubel and Cai*, 2004a; *Lee et al.*, 2016). However, the nonparametric approaches do not provide regression effect estimation. *Schaubel and Cai* (2004b) proposed a semiparametric relative-risk model and *Huang* (2002) proposed AFT-type gap time models for regression parameter estimations. Among all the options, only frailty models and Copula methods are trying to estimate not only the effects from covariates, but also the associations between events. Moreover, many estimation methods for frailty models, e.g., expectations-maximization (EM) (*Dempster et al.*, 1977) and numerical approximations (*Lange*, 2010), predict the subject-specific random effects, which can be quite useful when predicting subject-specific hazards or survival functions.

Almost all the recurrent events methods mentioned above treated the duration of the event status as a point process. Taking recurrent hospitalizations as an example, the inpatient periods are usually counted into the waiting time before the next admission, which may cause bias in the estimation of the readmission process; or the inpatient episodes are ignored, which could decrease the estimating efficiency by neglecting related information in the data. Moreover, both event processes might be interesting to researchers. Bivariate events and the extension to multivariate events using log-normal frailty models have been studied by *Xue and Brookmeyer* (1996), but the slow EM algorithm limits their use in practice. To analyze alternating gap times, *Huang and Wang* (2005) specified a non-parametric framework; *Yan and Fine* (2008) implemented a combination of marginal mean models and temporal regression models; and *Lee et al.* (2018) developed AFT based estimating approaches. However, none of these three methods explicitly assess the associations within and between the two event processes.

In Chapter II, I propose an alternative approach based on a penalized survival model to analyze alternating recurrent events. The penalty term is naturally inherent from the correlated frailty model, where the association between the two alternating states can be explained by the covariates and a pair of correlated random effects. Note that most existing frailty models assume independent random effects and control the association direction by regulating their multipliers. See the hierarchical models given by *Ripatti and Palmgren* (2000) and *Dharmarajan et al.* (2018) for details. However, the direction of the correlation between two event types is often unknown and often difficult to anticipate. For instance, a longer care in the hospital may lead to a longer waiting time before readmission, or associate with a sicker status and hence predict a sooner readmission. Extending *Ripatti and Palmgren* (2000) and *Therneau et al.* (2003), I derive a Laplace approximation based estimation method permitting complete variance-covariance estimation for the correlated random effects. The regression parameters are estimated from a penalized partial likelihood. In addition, based on the approximate marginal likelihood, I develop a likelihood ratio test (LRT) to evaluate the existence of dependence between two event intensities. Our method provides accurate parameter estimation and relatively fast computation time. The proposed method can also be extended to accommodate more complicated models with an arbitrary number of event types and jointly with longitudinal observations.

Frailty models have a general limitation that the distribution of random effects needs to be prespecified, and its mis-specification may result in a biased estimation (*Wienke*, 2010). Moreover, the distributional assumption on the frailties cannot be verified in practice. To circumvent this issue, one can treat the frailties as nuisance parameters (*Wang et al.*, 2001), at the expense of some loss in estimation efficiency relative to parametric frailty methods (*Ye et al.*, 2007). Alternatively, according to *Laird* (1978), *Heckman and Singer* (1982) and *Heckman and Singer* (1984), the distributions of random effects can be estimated via nonparametric maximum likelihood estimators (NPMLE), which has been extensively discussed in Chapter IV of *Xu* (2011). In Chapter III, I develop a novel estimating equation framework for a family of semiparametric models of recurrent events, extending the work of *Wang et al.* (2001). The proposed estimating approach allows consistent estimation of the regression parameters, the nonparametric baseline and the variance of the random effects. Moreover, the proposed framework can accommodate multiple types of events following correlated frailty models, and hierarchically structured event data with nested frailty models.

Both Chapters II-III utilize frailty models. In chapter IV, I propose a type of marginal multistate rate models to analyze the time-varying transitions of physical activity states among the children in a cohort from Mexico City. We transform the tri-axial accelerometer data into time-varying categorical states and analyze the nature of daily changing pattern of the transition rates between states. The proposed multistate rate models are analyzing "competing rates" jointly, in the spirit of both event rate models and the competing risk methods. Our proposed multistate rate models demonstrate that Pb exposure is positively associated with activity transitions to vigorous states among the boys, while transitions to moderate activity states among the girls. The proposed models enjoy the adaptability to a finer stratification, and the flexibility to shared covariate effects. Moreover, the proposed models can be easily implemented in a well-established R package Survival (*Therneau*, 2015).

When I was preparing my dissertation, the novel coronavirus (COVID-19) pandemic attacked hundreds of countries around the world and caused hundreds of thousands of deaths. I added this chapter to foster the development of convenient computational tools for public-health practitioners to make public decisions conveniently based on reasonable forecasting models. The proposed epidemiological models are built upon a hierarchical structure; with two observed time series of daily proportions of the infected and removed cases (dead and recovered). The two observational layers are generated from the underlying infection dynamics governed by a Markov Susceptible-Infectious-Removed (SIR) infectious disease process under a Bayesian framework. The latent SIR model based on the three ordinary differential equations are solved via the fourth-order Runge-Kutta approximation. The regular SIR model has been extended by including either a time-dependent modifier of the transmission rate, or a time-varying quarantine compartment whose prevalence follows a Dirac delta function. To deal with the accuracy in face of the strong discontinuity of the latter extension, I propose a two-step computation procedure. The proposed compartment models are established in an open-source R package on GitHub (lilywang1988/eSIR).

CHAPTER II

Penalized Survival Models for the Analysis of Alternating Recurrent Event Data

2.1 Introduction

Recurrent event data are commonly encountered in clinical experiments and observational studies. Examples for recurrent events include repeated hospitalizations, recurrent opportunistic infections for HIV-infected patients, and recurrent tumors for cancer patients. Various methods have been developed to analyze multivariate failure times by formulating models based on either intensity functions or rate functions (*Prentice et al.*, 1981; *Andersen and Gill*, 1982; *Pepe and Cai*, 1993; *Lawless and Nadeau*, 1995; *Lin et al.*, 2000). Despite the utility of these approaches, an important limitation is that each of these methods treats the recurrent event sequence as a point process and hence assumes (at least implicitly) that event durations are negligible. In cases where event duration is variable and not negligible, information and accuracy are sacrificed if the event duration is not considered. Taking recurrent hospitalizations as an example, the inpatient periods were counted into the awaiting time before the next admission, which may cause bias in the estimation of the readmission process; or the inpatient episodes are ignored, which could decrease the efficiency by neglecting some associated information in the data. Moreover, both event processes might be interesting to researchers. Other examples include recurrent infections among patients who have HIV or who have experienced hematopoietic cell transplantations, whereby the durations of infections could be of various lengths.

A natural way to accommodate recurrent events with non-negligible duration is to cast the data structure as an alternating gap time sequence. Very few methods have been proposed along these lines. A non-parametric approach was developed by *Huang and Wang* (2005) to estimate the joint distribution of two alternating events and the marginal distribution of the first recurrent event; however, this method does not provide inference on covariate effects. Recently, *Lee et al.* (2018) developed an estimating equation approach for accelerated failure time (AFT) model for an alternating recurrent event data. In their proposed estimating procedure, the distribution of the possibly-correlated random variables for two processes was left unspecified. Like *Huang and Wang* (2005), *Lee et al.* (2018) does not provide information on the correlations within or between the two recurrent event sequences.

Alternatively, correlated frailty models can be developed to accommodate the alternating recurrent event setting by adapting bivariate frailty models for clustered multivariate failure time data (*Yashin et al.*, 1995; *Xue and Brookmeyer*, 1996). The marginal likelihood (integrating out the unobserved frailties) can be obtained through the expectation-maximization (EM) algorithm, Gaussian-Hermite quadrature or a Laplace approximation (*Vaida and Xu*, 2000; *Ripatti et al.*, 2002; *Huang and Liu*, 2007; *Liu and Huang*, 2008). *Ripatti and Palmgren* (2000) developed a Laplace approximation based approach to estimate a penalized partial-likelihood (PPL) for a multivariate frailty model, which was shown to converge much faster than EM (*Therneau et al.*, 2003).

In this report, we propose a novel PPL estimating approach for alternating recurrent event data using correlated log-normal frailties. This is equivalent to modeling two recurrent event processes jointly, incorporating a bivariate random intercept (with correlated elements) to represent between- and within-process correlations. In contrast to *Ripatti and Palmgren* (2000), our estimating equation for the variance components is obtained by differentiating the approximate marginal likelihood with respect to its inverse variance matrix, other than to its scalar elements. Possibly due to the difficulty to estimate the each parameter of a non-diagonal variance matrix separately, *Ripatti and Palmgren* (2000) only provided estimators and simulation results for independent frailties, under an assumption that the two sub-clusters (right and left hips) were positively correlated. A similar strategy was employed by *Dharmarajan et al.* (2018) to model clustered competing risk data, assuming that the competing risks are negatively correlated. Although suitable for many practical settings, the requirement that the sign of the correlation be pre-specified limits the implementation of such approaches to alternating recurrent event data. In particular, it may be very difficult to accurately pre-specify the sign of a correlation that, depending on the application at hand, could be either positive or negative.

Our objective in this report is to develop methods for analyzing alternating recurrent event data that are flexible, informative, computationally efficient and implementable in very large data sets. In contrast to previous works, our proposed method estimates the frailty covariance matrix as a whole, with no need to restrict the correlation parameter in any fashion. In addition, a likelihood ratio test (LRT) is proposed to assess whether or not the two alternating recurrent event sequences are mutually independent.

The remainder of this report is organized as follows. We introduce our model in the setting of alternating recurrent events in Section 2.2. In Section 2.3, our method is described by deriving an approximate marginal likelihood (Subsection 2.3.1) based on the model in Section 2.2, developing a new PPL estimation approach (Subsection 2.3.2), and proposing a LRT based on a marginal PPL (Subsection 2.3.3). Simulations to evaluate the proposed methods on finite sample sizes are summarized in

Section 2.4. We apply the proposed method to an analysis of end-stage renal disease patients from the Dialysis Outcomes and Practice Patterns Study (DOPPS) in Section 2.5. A more comprehensive introduction of Laplace approximation and the bias correction for our proposed model can be found in Section 2.6. An exploratory extension of the current estimating approach to general clustered survival outcomes with an arbitrary K event types is introduced in Section 2.7 with preliminary simulation results. Concluding remarks are provided in Section 2.8.

2.2 Model Specification

A total of n independent subjects are followed over time and experience two alternating states indexed by k, i.e., $k = 1$ indicates the first event and $k = 2$ indicates the second event. For subject i, let random vectors $\boldsymbol{T}_i^* = \{(T_{ij1}^*, T_{ij2}^*), j = 1, \cdots\}$ indicate the total event times (i.e., measured from the start of follow-up) for both event types; j indicates the order of recurrent event pairs, i.e., $0 < T_{i11}^* < T_{i12}^* < T_{i21}^* < T_{i22}^* < \cdots$. The gap times between recurrent events for subject i are denoted by $\widetilde{\boldsymbol{T}}_i^* = \{(\widetilde{T}_{ij1}^*, \widetilde{T}_{ij2}^*), j = 1, \cdots\}$, where $\widetilde{T}_{ij1}^* = T_{ij1}^* - T_{i(j-1)2}^*$ and $\widetilde{T}_{ij2}^* = T_{ij2}^* - T_{ij1}^*$.

The covariate vector is denoted by $\boldsymbol{Z}_i(t)$, with any time-dependent elements restricted to be external covariates (*Kalbfleisch and Prentice*, 2002). Moreover, we set the covariates for gap time $\widetilde{T}_{ijk}$ to $\mathbf{Z}_{ijk}$, such that all elements (even if time-dependent) are set to their values at their respective gap time origins; i.e., $\mathbf{Z}_{ij1} = \mathbf{Z}_{ij}(T_{i,j-1,2})$ and $\mathbf{Z}_{ij2} = \mathbf{Z}_{ij}(T_{ij1})$. Note that the covariates comprising the two recurrent event types, $\mathbf{Z}_{ij1}$ and $\mathbf{Z}_{ij2}$, can be different. If no time-varying external covariates are considered, we only use their baseline measurements.

The alternating recurrent event process is right-censored by C_i. Let m_i be the number of observed complete event pairs from subject i. It is possible to have more than m_i event pairs, but due to the censoring, there are only m_i complete pairs being observed; it is still possible to observe event time $T_{i(m_i+1)1}^*$, but $T_{i(m_i+1)2}^*$ is always

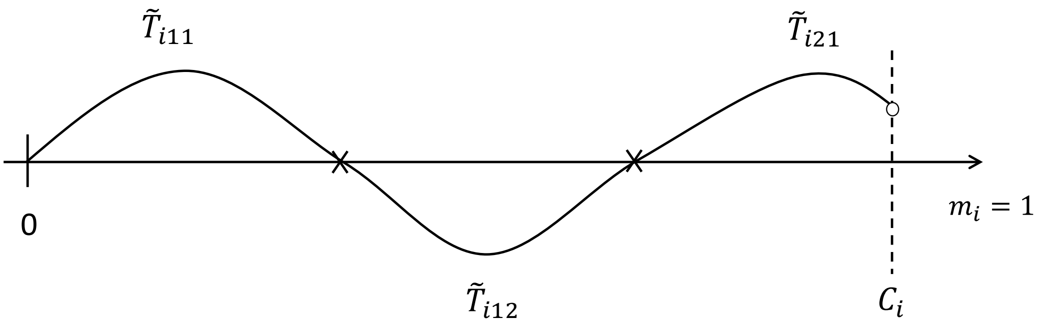


Figure 2.1: Alternating recurrent events under right censoring. $\widetilde{T}_{i11}$ is the observed awaiting time to the first hospitalization of subject i. $\widetilde{T}_{i12}$ is the length of stay before discharge. $\widetilde{T}_{i21}$, the observed time to the second hospitalization, is censored. Therefore, we only observe 1 complete event pair ($m_i = 1$).

censored. Correspondingly, we denote the observed covariate history by $\boldsymbol{Z}_i = \{\boldsymbol{Z}_{ijk}, j = 1, \cdots, m_i + 1, k = 1, 2\}$, and assume that C_i is independent of $\boldsymbol{T}_i^*$ given $\boldsymbol{Z}_i$.

As depicted in Figure 2.1, the two alternating recurrent events can be a series of alternating admission and discharge from a hospital. Let $\widetilde{T}_{ij1}$ denote the gap time between the previous discharge (or time 0 for $j = 1$) and admission, and let $\widetilde{T}_{ij2}$ denote the length of stay, i.e., time between hospital admission and discharge.

The event indicators and total observation times are defined as $\delta_{ijk} = I(T_{ijk}^* < C_i)$ and $T_{ijk} = T_{ijk}^* \wedge C_i$, where $I(\cdot)$ is an indicator function and $a \wedge b = min(a, b)$. Note that it is always true that for $j = 1, \ldots, m_i$, we have $\delta_{ijk} = 1$ and $T_{ijk} = T_{ijk}^*$, while $\delta_{i(m_i+1)2} = 0$ and $T_{i(m_i+1)2} = C_i$. Event gap times $\widetilde{T}_{ijk}^*$ are subject to censoring $\widetilde{C}_{ijk}$, where $\widetilde{C}_{ij1} = C_i - T_{i(j-1)2}$ and $\widetilde{C}_{ij2} = C_i - T_{ij1}$. Note that $\widetilde{T}_{i0k}^* = 0$ and $\widetilde{T}_{i0k} = 0$ for $k = 1, 2$. Consequently, the observed gap times are $\widetilde{T}_{ijk} = \widetilde{T}_{ijk}^* \wedge \widetilde{C}_{ijk}$. For each event gap, we introduce an at-risk indicator $Y_{ijk}(t) = I(t \leq \widetilde{T}_{ijk})$.

The assumed hazard model for $\widetilde{T}_{ijk}$ is given by

$$\lambda_{ijk}(t \mid \boldsymbol{Z}_i, \boldsymbol{\gamma}_i) = \lambda_{0k}(t) \exp(\boldsymbol{\beta}_k' \boldsymbol{Z}_{ijk} + \gamma_{ik}), \tag{2.1}$$

where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2})'$ are independent draws from a mean-zero bivariate normal distribution, BVN$(\mathbf{0}_2, \boldsymbol{D}_{2\times 2})$. The shared baseline hazards in (2.1) implies that the alternating event process follows a renewal property so that the baselines are shared within each event type. We assume that the $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots$ are mutually independent, and that each γ_i is independent of $\boldsymbol{Z}_i$. Moreover, for $(j,k) \neq (p,q)$, $\widetilde{T}^*_{ijk}$ and $\widetilde{T}^*_{ipq}$ are independent given $\{\boldsymbol{\gamma}_i, \boldsymbol{Z}_{ijk}, \boldsymbol{Z}_{ipq}\}$. Consequently, the $\widetilde{T}^*_{ijk}$ are subject to censoring via $\widetilde{C}_{ijk}$ which is conditionally independent given $\boldsymbol{\gamma}_i$ and $\boldsymbol{Z}_i$.

We stack $\boldsymbol{\gamma}_i$ from the n subjects into a vector $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', \cdots, \boldsymbol{\gamma}_n')'$, which then follows a mean-zero multivariate normal distribution MVN$(\mathbf{0}_{2n}, \boldsymbol{\Sigma}_{2n\times 2n})$. Note that $\boldsymbol{\Sigma} = \boldsymbol{D} \otimes \boldsymbol{I}_{n\times n}$ is a block-diagonal matrix, where $\otimes$ is a Kronecker product and $\boldsymbol{I}_{n\times n}$ is an n by n identity matrix. We also include frailty design vectors $\boldsymbol{R}_{ik} = (0_{(1)}, \ldots, 1_{(2i-2+k)}, \ldots, 0_{(2n)})$ to indicate γ_{ik} in the $(2i-2+k)^{th}$ entry of $\boldsymbol{\gamma}$ is present. We only account for random intercepts here, though it is possible for the models to be extended such that various covariates have random effects. The proposed event-specific intensity in (2.1) will become

$$\lambda_{ijk}(t \mid \boldsymbol{Z}_i, \boldsymbol{R}_{ik}, \boldsymbol{\gamma}) = \lambda_{0k}(t)\exp(\boldsymbol{\beta}_k'\boldsymbol{Z}_{ijk} + \boldsymbol{\gamma}'\boldsymbol{R}_{ik}). \tag{2.2}$$

The likelihood for subject i conditional on $\boldsymbol{\gamma}$ and $\boldsymbol{Z}_i$ is given by

$$\begin{aligned} &L_i(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{01}(\cdot), \lambda_{02}(\cdot) \mid \boldsymbol{\gamma}, \boldsymbol{Z}_i) \\ &= \prod_{j=1}^{m_i+1}\prod_{k=1}^{2}\Big[\lambda_{0k}(\widetilde{T}_{ijk})\exp(\eta_{ijk})\Big]^{\delta_{ijk}} \exp\Big\{-\Lambda_{0k}(\widetilde{T}_{ijk})\exp(\eta_{ijk})\Big\}, \end{aligned} \tag{2.3}$$

where the cumulative intensity of event process k is $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(s)ds$, and $\eta_{ijk} = \boldsymbol{\beta}_k'\boldsymbol{Z}_{ijk} + \boldsymbol{\gamma}'\boldsymbol{R}_{ik}$. It follows from (2.3) that the marginal likelihood is

$$\begin{aligned} L_m &= L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{01}(\cdot), \lambda_{02}(\cdot)) \\ &= \int_{\boldsymbol{\gamma}}\prod_{i=1}^{n} L_i(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{01}(\cdot), \lambda_{02}(\cdot) \mid \boldsymbol{\gamma}, \boldsymbol{Z}_i)\exp(-\tfrac{1}{2}\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma})\big|\boldsymbol{\Sigma}\big|^{-\frac{1}{2}}\mathrm{d}\boldsymbol{\gamma}. \end{aligned} \tag{2.4}$$

Note that the marginal likelihood in (2.4) is not in a closed form. A PPL-based estimation procedure is developed for the proposed model.

2.3 Parameter Estimation

2.3.1 Approximate Likelihood

We derive an approximate marginal likelihood for the log-likelihood from (2.3). The joint likelihood function for the observations from $\boldsymbol{n}$ subjects and their frailties can be represented as

$$L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{01}(\cdot), \lambda_{02}(\cdot), \boldsymbol{\gamma}) = \left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} \exp(-\boldsymbol{K}(\boldsymbol{\gamma})), \tag{2.5}$$

where we define $\boldsymbol{K}(\boldsymbol{\gamma})$ is

$$\boldsymbol{K}(\boldsymbol{\gamma}) = \tfrac{1}{2}\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}+ \\ \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m_i+1} \sum_{k=1}^{2} \Lambda_{0k}(\widetilde{T}_{ijk}) \exp(\eta_{ijk}) - \delta_{ijk} \left\{\eta_{ijk} + \log(\lambda_{0k}(\widetilde{T}_{ijk}))\right\}.$$

Through a Taylor expansion, $\boldsymbol{K}(\boldsymbol{\gamma})$ is approximated by

$$\boldsymbol{K}(\boldsymbol{\gamma}) \approx \widehat{\boldsymbol{K}}(\boldsymbol{\gamma}) = \boldsymbol{K}(\widetilde{\boldsymbol{\gamma}}) + \frac{1}{2}(\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}})'\boldsymbol{K}_2(\widetilde{\boldsymbol{\gamma}})(\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}}),$$

where $\widetilde{\boldsymbol{\gamma}}$ is the solution of $\boldsymbol{K}_1(\boldsymbol{\gamma}) = 0$, with

$$\boldsymbol{K}_1(\boldsymbol{\gamma}) = \frac{\partial \boldsymbol{K}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma} + \sum_{i=1}^{n} \sum_{j=1}^{m_i+1} \sum_{k=1}^{2} \left\{\Lambda_{0k}(t_{ijk}) \exp(\eta_{ijk}) - \delta_{ijk}\right\} \boldsymbol{R}_{ik}. \tag{2.6}$$

The corresponding second derivative is given by

$$\boldsymbol{K}_2(\boldsymbol{\gamma}) = \frac{\partial^2 \boldsymbol{K}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = \boldsymbol{\Sigma}^{-1} + \sum_{i=1}^{n} \sum_{j=1}^{m_i+1} \sum_{k=1}^{2} \Lambda_{0k}(t_{ijk}) \exp(\eta_{ijk}) \boldsymbol{R}_{ik}^{\otimes 2}, \tag{2.7}$$

where we define $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}'$. Note that $\boldsymbol{K}_2(\boldsymbol{\gamma})$ in (2.7) is a block-diagonal matrix.

Through a Laplace approximation, we plug $\widehat{\boldsymbol{K}}(\boldsymbol{\gamma})$ into (2.5) and integrate it to obtain an approximate marginal log-likelihood,

$$l_m = \log(L_m) \approx -\frac{1}{2}\log\left|\boldsymbol{\Sigma}\right| - \boldsymbol{K}(\widetilde{\boldsymbol{\gamma}}) - \frac{1}{2}\log|\boldsymbol{K}_2(\widetilde{\boldsymbol{\gamma}})|. \tag{2.8}$$

The function $\boldsymbol{K}(\boldsymbol{\gamma})$ can be decomposed as follows,

$$\boldsymbol{K}(\boldsymbol{\gamma}) = -\text{PPLL} - h(\lambda_{01}(\cdot), \lambda_{02}(\cdot), \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma}), \tag{2.9}$$

where PPLL represents the penalized partial log-likelihood (Subsection 2.3.2) and $h(\cdot)$ is defined in Appendix 1. In practice, inference on the regression parameters is often simplified by solely focusing on the PPLL term. *Ripatti and Palmgren* (2000) adopted a similar simplification, and demonstrated in their simulation studies that the information loss due to neglecting $h(\cdot)$ was negligible.

2.3.2 Penalized Partial Likelihood Estimation

To estimate the regression coefficients and variance components, we need two iterating steps, the inner loops and the outer loops. In the inner loop, a Newton-Raphson algorithm is conducted based on PPLL, treating both $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}_k$ as parameters, and treating $\widehat{\boldsymbol{D}}$ as known from the previous outer loop. The outer loop is grounded in an approximate marginal likelihood, fixing $\widehat{\boldsymbol{\theta}}$ from the recent inner loop, where $\boldsymbol{\theta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\gamma}')'$. We outline the proposed algorithm below.

2.3.2.1 Inner Loop

Given the variance matrix $\widehat{\boldsymbol{\Sigma}}$ (or $\widehat{\boldsymbol{D}}$), the PPLL is expressd as

$$\begin{aligned}\text{PPLL} = \; & -\tfrac{1}{2}\boldsymbol{\gamma}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\gamma} \\ & + \sum_{i=1}^{n}\sum_{j=1}^{m_i+1}\sum_{k=1}^{2}\delta_{ijk}\left\{\eta_{ijk} - \log(\sum_{l=1}^{n}\sum_{p=1}^{m_l+1} Y_{lpk}(t_{ijk})\exp(\eta_{lpk}))\right\}. \end{aligned} \tag{2.10}$$

For notation simplicity, let

$$\begin{aligned} \boldsymbol{S}_{Z_k}^{(d)}(t) &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i+1} Y_{ijk}(t)\exp(\eta_{ijk})\boldsymbol{Z}_{ijk}^{\otimes d}, \\ \boldsymbol{S}_{R_k}^{(d)}(t) &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i+1} Y_{ijk}(t)\exp(\eta_{ijk})\boldsymbol{R}_{ik}^{\otimes d}, \\ \boldsymbol{S}_{ZR_k}(t) &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i+1} Y_{ijk}(t)\exp(\eta_{ijk})\boldsymbol{Z}_{ijk}\boldsymbol{R}'_{ik}, \end{aligned} \tag{2.11}$$

where $d \in \{0, 1, 2\}$.

Let $\overline{\boldsymbol{Z}}_k(t) = \boldsymbol{S}_{Z_k}^{(1)}(t)/S_{Z_k}^{(0)}(t)$, $\overline{\boldsymbol{R}}_k(t) = \boldsymbol{S}_{R_k}^{(1)}(t)/S_{R_k}^{(0)}(t)$, $\boldsymbol{V}_{Z_k}(t) = \boldsymbol{S}_{Z_k}^{(2)}(t)/S_{Z_k}^{(0)}(t)$, $\boldsymbol{V}_{R_k}(t) = \boldsymbol{S}_{R_k}^{(2)}(t)/S_{R_k}^{(0)}(t)$ and $\boldsymbol{V}_{ZR_k}(t) = \boldsymbol{S}_{ZR_k}(t)/S_{Z_k}^{(0)}(t) = \boldsymbol{S}_{ZR_k}(t)/S_{R_k}^{(0)}(t)$. Taking first and second partial derivatives of PPLL with respect to $\boldsymbol{\beta}_k$ and $\boldsymbol{\gamma}$, we obtain corresponding score functions

$$\frac{\partial PPLL}{\partial \boldsymbol{\beta}_k} = \sum_{i=1}^{n}\sum_{j=1}^{m_i+1}\delta_{ijk}\left\{\boldsymbol{Z}_{ijk} - \overline{\boldsymbol{Z}}_k(\widetilde{T}_{ijk})\right\} \tag{2.12}$$

$$\frac{\partial PPLL}{\partial \boldsymbol{\gamma}} = \sum_{i=1}^{n}\sum_{j=1}^{m_i+1}\sum_{k=1}^{2}\delta_{ijk}\left\{\boldsymbol{R}_{ik} - \overline{\boldsymbol{R}}_k(\widetilde{T}_{ijk})\right\} - \widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\gamma}, \tag{2.13}$$

and information matrices

$$
\begin{aligned}
-\frac{\partial^2 PPLL}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_k{}'} &= \sum_{i=1}^{n} \sum_{j=1}^{m_i+1} \delta_{ijk} \left\{ \boldsymbol{V}_{Z_k}(\widetilde{T}_{ijk}) - \overline{\boldsymbol{Z}}_k(\widetilde{T}_{ijk})^{\otimes 2} \right\} && (2.14) \\
-\frac{\partial^2 PPLL}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} &= \sum_{i=1}^{n} \sum_{j=1}^{m_i+1} \sum_{k=1}^{2} \delta_{ijk} \left\{ \boldsymbol{V}_{R_k}(\widetilde{T}_{ijk}) - \overline{\boldsymbol{R}}_k(\widetilde{T}_{ijk})^{\otimes 2} \right\} + \widehat{\boldsymbol{\Sigma}}^{-1} && (2.15) \\
-\frac{\partial^2 PPLL}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\gamma}'} &= \sum_{i=1}^{n} \sum_{j=1}^{m_i+1} \delta_{ijk} \left\{ \boldsymbol{V}_{ZR_k}(\widetilde{T}_{ijk}) - \overline{\boldsymbol{Z}}_k(\widetilde{T}_{ijk}) \overline{\boldsymbol{R}}_k(\widetilde{T}_{ijk})' \right\}. && (2.16)
\end{aligned}
$$

If we substitute $\Lambda_{0k}(t)$ in (2.6) with their corresponding Breslow estimators, $\boldsymbol{K}_1(\boldsymbol{\gamma})$ will be equivalent to $\partial PPLL/\partial \boldsymbol{\gamma}$. Unlike $\boldsymbol{K}_2(\boldsymbol{\gamma})$ in (2.7), $\boldsymbol{K}_{PPL2}(\boldsymbol{\gamma}) := [\partial^2 \mathrm{PPLL}/\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}']$ is not block diagonal. It is trivial to prove that $\partial^2 \mathrm{PPLL}/\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2{}'$ equals 0.

Let PLL be the partial log-likelihood without a penalty term. The Hessian matrix $\boldsymbol{H}(\boldsymbol{\theta})$ is given by

$$
\boldsymbol{H}(\boldsymbol{\theta}) = \boldsymbol{I}(\boldsymbol{\theta}) + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \widehat{\boldsymbol{\Sigma}}^{-1} \end{bmatrix}, \tag{2.17}
$$

where we have

$$
\boldsymbol{I}(\boldsymbol{\theta}) = -\frac{\partial^2 \mathrm{PLL}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\begin{bmatrix} \frac{\partial^2 \mathrm{PLL}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1{}'} & \mathbf{0} & \frac{\partial^2 \mathrm{PLL}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\gamma}'} \\ \mathbf{0}' & \frac{\partial^2 \mathrm{PLL}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2{}'} & \frac{\partial^2 \mathrm{PLL}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\gamma}'} \\ \frac{\partial^2 \mathrm{PLL}}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}_1{}'} & \frac{\partial^2 \mathrm{PLL}}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}_2{}'} & \frac{\partial^2 \mathrm{PLL}}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \end{bmatrix}.
$$

There are two options for the asymptotic covariance estimate of $\widehat{\boldsymbol{\theta}}$, $\boldsymbol{H}(\widehat{\boldsymbol{\theta}})^{-1} \boldsymbol{I}(\widehat{\boldsymbol{\theta}}) \boldsymbol{H}(\widehat{\boldsymbol{\theta}})^{-1}$ (*Gray*, 1992) and $\boldsymbol{H}(\widehat{\boldsymbol{\theta}})^{-1}$ (*Verweij and Van Houwelingen*, 1994). In the inner loop, the variances for $\widehat{\boldsymbol{\theta}}$ are underestimated since $\widehat{\boldsymbol{\Sigma}}$ is fixed. Since $\boldsymbol{H}(\boldsymbol{\theta})^{-1}$ has been demonstrated in related contexts to be more conservative in Wald tests (*Therneau et al.*, 2003), $\boldsymbol{H}(\boldsymbol{\theta})^{-1}$ is employed here with the intention of increasing coverage probability. Note that when the sample size is large, one can sparsen the Hessian matrix by reducing the off-2×2-block-diagonal part of $\boldsymbol{K}_{PPL2}(\boldsymbol{\gamma})$ to be 0 before calculating the inverse.

2.3.2.2 Outer Loop

Fixing $\widehat{\boldsymbol{\theta}}$ from the previous inner loop, $\boldsymbol{D}$ can be estimated in the outer loop through an approximate marginal profile log-likelihood by dropping irrelevant terms from (2.8) (e.g., not including $\boldsymbol{D}$), such that

$$\begin{aligned} l_m &\approx -\tfrac{1}{2}\log|\boldsymbol{\Sigma}| - \tfrac{1}{2}\log|\boldsymbol{K}_2(\widehat{\boldsymbol{\gamma}})| - \boldsymbol{K}(\widehat{\boldsymbol{\gamma}}) \\ &\propto -\tfrac{1}{2}\log|\boldsymbol{\Sigma}| - \tfrac{1}{2}\log|\boldsymbol{K}_2(\widehat{\boldsymbol{\gamma}})| - \tfrac{1}{2}\widehat{\boldsymbol{\gamma}}'\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\gamma}}. \end{aligned} \tag{2.18}$$

We now derive an estimator for the entire variance-covariance matrix of the frailties. Given the baselines fixed, $\boldsymbol{K}_2(\widehat{\boldsymbol{\gamma}})$ is a block-diagonal matrix with each block defined as $\boldsymbol{K}_2(\widehat{\boldsymbol{\gamma}})_{ii}$, thus the marginal likelihood from (2.18) can be re-arranged as a function of $\boldsymbol{D}^{-1}$,

$$\tfrac{n}{2}\log|\boldsymbol{D}^{-1}| - \tfrac{1}{2}\sum_{i=1}^{n}\log|\boldsymbol{K}_2(\widehat{\boldsymbol{\gamma}})_{ii}| - \tfrac{1}{2}\sum_{i=1}^{n}\widehat{\boldsymbol{\gamma}}_i'\boldsymbol{D}^{-1}\widehat{\boldsymbol{\gamma}}_i, \tag{2.19}$$

where we have

$$\begin{aligned} \boldsymbol{K}_2(\widehat{\boldsymbol{\gamma}})_{ii} &= \boldsymbol{D}^{-1} + \boldsymbol{M}_i(\widehat{\boldsymbol{\theta}}), \\ \boldsymbol{M}_i(\widehat{\boldsymbol{\theta}}) &= \mathrm{diag}\left[\sum_{\mathrm{j=1}}^{\mathrm{m_i+1}} \Lambda_{01}(\widetilde{\mathrm{T}}_{\mathrm{ij1}})\exp(\widehat{\eta}_{\mathrm{ij1}}), \sum_{\mathrm{j=1}}^{\mathrm{m_i+1}} \Lambda_{02}(\widetilde{\mathrm{T}}_{\mathrm{ij2}})\exp(\widehat{\eta}_{\mathrm{ij2}})\right], \end{aligned}$$

and when the diag(·) function maps a vector to a diagonal matrix. Taking the first derivative of approximate marginal likelihood with respect to $\boldsymbol{D}^{-1}$, we obtain the estimating equation

$$\tfrac{n}{2}\boldsymbol{D} - \tfrac{1}{2}\sum_{i=1}^{n}\boldsymbol{K}_2(\widehat{\boldsymbol{\gamma}})_{ii}^{-1} - \tfrac{1}{2}\sum_{i=1}^{n}\widehat{\boldsymbol{\gamma}}_i\widehat{\boldsymbol{\gamma}}_i' = 0, \tag{2.20}$$

and its solution

$$\boldsymbol{D} = \frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{K}_2(\widehat{\boldsymbol{\gamma}})_{ii}^{-1} + \widehat{\boldsymbol{\gamma}}_i\widehat{\boldsymbol{\gamma}}_i'\right]. \tag{2.21}$$

Let $\widehat{\boldsymbol{D}}_t$ denote the variance-covariance matrix estimate from the t^{th} outer loop. The $(t+1)^{th}$ estimator can thus be expressed as

$$\widehat{\boldsymbol{D}}_{t+1} = \frac{1}{n}\sum_{i=1}^{n}\left[(\widehat{\boldsymbol{D}}_t^{-1} + \boldsymbol{M}_i(\widehat{\boldsymbol{\theta}}))^{-1} + \widehat{\boldsymbol{\gamma}}_i\widehat{\boldsymbol{\gamma}}_i'\right]. \tag{2.22}$$

The variance-covariance estimator (2.22) is analogous to the recursive estimating formula for logistic regression models derived through a Laplace approximation (*Demidenko* (2004); Ch 7.7.2). The convergence of this type recursive estimator is verified by the Fixed Point Theorem (*Zamfirescu*, 1972). We suggest initializing $\boldsymbol{D}$ with a diagonal matrix; e.g., identity matrices were employed in our simulations. Standard errors for the variance components, if of interest, could be obtained by bootstrapping.

We found that directly replacing $\Lambda_{0k}(t)$ with its corresponding Breslow estimator tends to result in overestimation of the diagonal entries for $\boldsymbol{D}$. Moreover, the baseline calculation could be intensive, especially when the sample size is large. Both $\boldsymbol{K}_2(\boldsymbol{\gamma})$ and $\boldsymbol{K}_{PPL2}(\boldsymbol{\gamma})$ are summations involving a $\boldsymbol{\Sigma}^{-1}$, and their numerical difference can be captured by the second derivative of $h(\cdot)$ with respect to $\boldsymbol{\gamma}$. *Ripatti and Palmgren* (2000) suggested to replace the $\boldsymbol{K}_2(\boldsymbol{\gamma})$ with $\boldsymbol{K}_{PPL2}(\boldsymbol{\gamma})$ for a more accurate estimation on the variance components, and to avoid computing the baselines for each updating step. In a similar vein, we substitute the 2×2 matrices located on the block-diagonal of $[\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})]^{-1}$, which is denoted as $[\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})^{-1}]_{blk_i}$, for $\boldsymbol{K}_2(\widehat{\boldsymbol{\gamma}})_{ii}^{-1}$ in (2.21). Subsequently,we have a new estimator

$$\widehat{\boldsymbol{D}}^{\#} = \frac{1}{n}\sum_{i=1}^{n}\left\{\left[\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})^{-1}\right]_{blk_i} + \widehat{\boldsymbol{\gamma}}_i\widehat{\boldsymbol{\gamma}}_i'\right\}, \tag{2.23}$$

which is shown to be positive-definite in Appendix A.1.

2.3.3 Likelihood Ratio Test

In order to test whether the frailties from the two alternating processes are independent to each other, we also propose a likelihood ratio test (LRT) based on an approximate marginal penalized partial log-likelihood (MPPL).

$$\mathrm{MPPL} = -\tfrac{n}{2}\log|\boldsymbol{D}| + \tfrac{1}{2}\sum_{i=1}^{n} \mid \left[\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})^{-1}\right]_{blk_i} \mid + \mathrm{PPLL}(\widehat{\boldsymbol{\gamma}}) \tag{2.24}$$

The procedure includes contrasting the estimated MPPL in (2.24) under the null and alternative hypotheses respectively. Under the null, we restrict $\boldsymbol{D}$ to be diagonal; while under the alternative, we do not. Notice that, under the null, the two recurrent processes can either be fitted separately, or be fitted together while restricting the off-diagonal entries of $\boldsymbol{D}^{\#}$ to be 0 in each outer step. Let $\widehat{\boldsymbol{D}}_0^{\#}$ be the variance estimator under the independence assumption,

$$\widehat{\boldsymbol{D}}_0^{\#} = \mathrm{extdiag}\left[\frac{1}{n}\sum_{i=1}^{N}\left\{\left[\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}_0})^{-1}\right]_{blk_i} + \widehat{\boldsymbol{\gamma}_0}_i\widehat{\boldsymbol{\gamma}_0}_i'\right\}\right], \tag{2.25}$$

where extdiag($\cdot$) is a function to extract diagonal part of the matrix, distinct from diag($\cdot$) defined previously. Correspondingly, if the parameters subscripted with 1 are from the unrestricted alternative, and those subscripted with 0 are from the null, the test statistic is then given by

$$\mathrm{LRT} = 2\left[\mathrm{MPPL}\left(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\Sigma}}_1^{\#}\right) - \mathrm{MPPL}\left(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\Sigma}}_0^{\#}\right)\right]. \tag{2.26}$$

One would reject the null hypothesis of independence if LRT> $\chi^2_{(1)\alpha}$, where α is the prespecified type 1 error and $\chi^2_{(1)}$ is chi-square with degree of freedom 1.

2.3.4 Comparison

To the best of our knowledge, there is no literature that derived a direct estimator for the entire variance-covariance matrix within a hierarchical survival model. Recall that in hierarchical frailty model discussed by *Ripatti and Palmgren* (2000), the authors incorporated three independent frailties to ensure a positive correlation between two events in their model. *Dharmarajan et al.* (2018) built a correlated frailty model of competing risks with a negative correlation by flipping a sign for one of its random effects. They both predefined the correlation directions between two events, which can be summarized as

$$\begin{aligned} \lambda_{ij1}(t \mid \boldsymbol{Z}_1, \boldsymbol{\gamma}_i^*) &= \lambda_{01}(t)\exp(\boldsymbol{\beta}_1'\boldsymbol{Z}_{ij1} + \gamma_{i1}^* + \gamma_{i0}^*) \\ \lambda_{ij2}(t \mid \boldsymbol{Z}_2, \boldsymbol{\gamma}_i^*) &= \lambda_{02}(t)\exp(\boldsymbol{\beta}_2'\boldsymbol{Z}_{ij2} + \gamma_{i2}^* \pm \gamma_{i0}^*), \end{aligned} \tag{2.27}$$

where $\boldsymbol{\gamma}_i^* = [\gamma_{i0}^*, \gamma_{i1}^*, \gamma_{i2}^*]'$ are independent and identically distributed draws from a normal distribution $N(0, \boldsymbol{D}^*)$ and $\boldsymbol{D}^* = \mathrm{diag}[\phi_0, \phi_1, \phi_2]$, and the plus-minus sign ± controls the the correlation sign between two events. Note that we use a superscript $*$ to distinguish the analogs in (*Ripatti and Palmgren*, 2000) from ours. By introducing in a 3 by n design matrices $\boldsymbol{R}_{ik}^*$, we can rewrite (2.27) into its analog of (2.2), substituted with $\boldsymbol{R}_{ik}^*$ for $\boldsymbol{R}_{ik}$, and $\boldsymbol{\gamma}^*$ for $\boldsymbol{\gamma}$.

In the outer loop, the estimating equations for ϕ_d with $d = 0, 1, 2$ are

$$-\frac{1}{2}\left\{tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_d}\right) + tr\left(\boldsymbol{K}_2^*(\widehat{\boldsymbol{\gamma}^*})^{-1}\frac{\partial\boldsymbol{\Sigma}^{*-1}}{\partial\phi_d}\right) - \widehat{\boldsymbol{\gamma}^*}'\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_d}\boldsymbol{\Sigma}^{*-1}\widehat{\boldsymbol{\gamma}^*}\right\} = 0. \tag{2.28}$$

Ripatti and Palmgren (2000) also suggested using the second derivative of the penalized partial log likelihood $K_{PPL2}^*(\boldsymbol{\gamma}) = \partial^2 PPLL/\partial\boldsymbol{\gamma}^*\partial\boldsymbol{\gamma}^{*\prime}$ in place of $K_2^*(\boldsymbol{\gamma}) = \partial^2\boldsymbol{K}(\boldsymbol{\gamma}^*)/\partial\boldsymbol{\gamma}^*\partial\boldsymbol{\gamma}^{*\prime}$ to achieve a better estimating performance. Note that these estimating equations (2.28) are corresponding to scalar entries of $\boldsymbol{D}$ matrix other than its whole in our method (2.20). In practice, due to the difficulty to obtain $\partial\boldsymbol{\Sigma}^{*-1}/\partial\phi_d$ and that its Newton-Raphson estimating procedure in the outer loop cannot ensure it to be positive-definite (though re-parametrization of the variance-covariance matrix can be another option), practitioners assume the independent frailty model with the sign of the covariance of $\boldsymbol{D}$ pre-specified as in (2.27) for convenience. Under such assumption as in models (2.27), the estimators for variance-covariance components are

$$\widehat{\phi}_d = \frac{[\widehat{\boldsymbol{\gamma}}^{*d}]'[\widehat{\boldsymbol{\gamma}}^{*d}]+tr\left[\{K^*_{PPL2}(\widehat{\boldsymbol{\gamma}}^*)^{-1}\}_d\right]}{n} \quad d \in \{0,1,2\}, \tag{2.29}$$

where we define $\widehat{\boldsymbol{\gamma}}^{*d} = [\widehat{\gamma}^*_{1d}, \ldots, \widehat{\gamma}^*_{nd}]'$, and $\{K^*_{PPL2}(\widehat{\boldsymbol{\gamma}}^*)^{-1}\}_d$ is the d^{th} sub-matrix of $K^*_{PPL2}(\widehat{\boldsymbol{\gamma}}^*)^{-1}$ corresponding to $\widehat{\boldsymbol{\gamma}}^{*d}$. In other words, $tr[\{K^*_{PPL2}(\widehat{\boldsymbol{\gamma}}^*)^{-1}\}_d]$ is summing up every d^{th} elements on the diagonal of $K^*_{PPL}(\widehat{\boldsymbol{\gamma}}^*)^{-1}$. Note that $K^*_2(\boldsymbol{\gamma}^*)$ is diagonal while $K^*_{PPL}(\boldsymbol{\gamma}^*)$ is not, thus taking the trace $tr[\{K^*_{PPL2}(\widehat{\boldsymbol{\gamma}}^*)^{-1}\}_d]$ is comparable to taking the sum of the diagonal blocks in our method (2.23). Consequently, the entire variance matrix estimator $\widehat{\boldsymbol{D}}^*$ can be assembled as

$$\widehat{\boldsymbol{D}}^* = \begin{bmatrix} \widehat{\phi}_1 + \widehat{\phi}_0 & \pm\widehat{\phi}_0 \\ \pm\widehat{\phi}_0 & \widehat{\phi}_2 + \widehat{\phi}_0 \end{bmatrix}. \tag{2.30}$$

Note that the difference between the two parameterizations, (2.27) and (2.1), mainly affects the outer loop procedure. Since the regression parameters have their score functions generally untouched in the inner loop, we would expect that there should not exist much difference between our proposed method and *Ripatti and Palmgren* (2000) in terms of regression parameter estimation. In Section 2.4 we also confirmed that the changes in regression parameter estimates were negligible.

In addition, our method largely speeds up the calculation and reduces the memory cost for several reasons. Firstly in the inner loop, our method estimates $2n + p$ of parameters (p is the number of fixed effects in both events), while the model in (2.27) needs $3n + p$. The information matrix is also reduced by 5/9. Once we implement

PPL for more types of events (e.g. three event processes), this dimension reduction will be even larger. Secondly, sparsening the matrix for large sized data can also dramatically reduce the computation burden. Moreover, we let the matrices from the inner loop be computed in through Rcpp, and improve the computing algorithms for score functions and information matrices to expedite the computation procedure significantly.

2.4 Simulation Studies

Simulations under different settings were carried out to evaluate the proposed method in reasonable sample sizes. To begin, the $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2})'$ were drawn independently from a mean-zero bivariate normal distribution with various specifications for $\boldsymbol{D}$. Given $\boldsymbol{\gamma}$, event times were generated in alternating turns, added up and recorded until the censoring time $C_i = 10$. The covariates were drawn from independent standard normal distributions. The intensity function with respect to the j^{th} occurrence of the event type k is given by $\lambda_{0k} \exp(\boldsymbol{\beta}_k' \boldsymbol{Z}_{ijk} + \gamma_{ik})$. We denote the $(a, b)^{th}$ entry of the variance matrix $\boldsymbol{D}$ by $\boldsymbol{D}[\mathrm{a}, \mathrm{b}]$. When $\boldsymbol{D}[1, 2] = 0$, the two alternating sequences are independent. We generated 500 samples and set the convergence tolerance to be 10^{-6} for each replicate.

Table 2.1 provides results for samples with different sizes and baseline intensities, $\lambda_{0k} = 1.5$ $(k = 1, 2)$. The median number of uncensored complete recurrent event pairs was ≈4. We varied sample sizes from $n = 100$ to $n = 1,000$. Estimated regression parameters were approximately unbiased, with asymptotic standard error (ASE) generally close to the empirical standard deviation (ESD). The results are similar across different sample sizes, implying that the estimation is not affected much by sample size if the cluster sizes or recurrent event numbers are fixed. By comparing their results, we noticed that the negatively correlated setting would experience less bias in their $\boldsymbol{D}$ matrix estimation than the positively correlated setting. For data with more recurrent events (greater m_i), the bias in estimating the $\boldsymbol{D}$ matrix decreased dramatically (Appendix A.2). Note that, the ASEs for regression parameters were calculated from the inverse Hessian matrix, but the ASEs for the entries of $\boldsymbol{D}$ were obtained via bootstrapping, for which we tried different bootstrapping sample sizes with $\boldsymbol{B} = 50$, $\boldsymbol{B} = 100$ and $\boldsymbol{B} = 200$. We notice that using $\boldsymbol{B} = 50$ provides sufficiently accurate estimates for the standard errors while takes relatively less computation time. For large n, we turned to m-out-of-n bootstrapping: resampling m observations out of n subjects with replacement (*Bickel et al.*, 1997; *Bickel and Sakov*, 2008). By comparing different $\sqrt{n}ESE$ of all the estimated parameters with $n = 50, 100, 200, \ldots, 1,000$, we noticed that when $n = 200$, it started to provide stable size adjusted standard errors ($\sqrt{n}ESE$). Thus we set $m = 200$ for datasets with $n > 200$, and the estimated ASEs from m-bootstraps were adjusted by the size difference by multiplying $\sqrt{m/n}$. The resulting coverage probabilities (CP) were fairly acceptable for both the regression parameters and the variance components.

Table 2.1: Simulation results: Estimating regression coefficients and variance components based on 500 replicates, with $\lambda_{0k} = 1.5$ for $k = 1, 2$ and a median m_i of ≈ 4. Sample sizes vary from $n = 100$ to $n = 1,000$, and "cr" denotes censoring rate, or the proportion of subjects with $m_i = 0$.

	True	$\boldsymbol{D}[1,2] > 0$				True	$\boldsymbol{D}[1,2] < 0$			
	Value	Bias	ESD	ASE	CP	Value	Bias	ESD	ASE	CP
$n = 100$	cr: 3.81%					cr: 3.64%				
β_1	1	-0.005	0.062	0.056	0.934	1	-0.006	0.064	0.059	0.920
β_2	-1	0.003	0.062	0.059	0.940	-1	0.001	0.064	0.062	0.934
$\boldsymbol{D}[1,1]$	0.25	-0.010	0.067	0.066	0.880	0.25	-0.007	0.075	0.073	0.922
$\boldsymbol{D}[2,2]$	0.25	-0.012	0.068	0.068	0.902	0.25	-0.003	0.081	0.078	0.912
$\boldsymbol{D}[1,2]$	0.125	-0.014	0.047	0.047	0.928	-0.125	-0.008	0.053	0.052	0.934
$n = 500$	cr: 3.84%					cr: 3.70%				
β_1	1	-0.004	0.025	0.025	0.950	1	-0.004	0.028	0.026	0.922
β_2	-1	0.003	0.028	0.026	0.930	-1	0.005	0.029	0.028	0.934
$\boldsymbol{D}[1,1]$	0.25	-0.012	0.028	0.030	0.932	0.25	-0.006	0.032	0.033	0.942
$\boldsymbol{D}[2,2]$	0.25	-0.012	0.032	0.031	0.900	0.25	-0.006	0.034	0.035	0.938
$\boldsymbol{D}[1,2]$	0.125	-0.013	0.022	0.021	0.868	-0.125	-0.005	0.023	0.024	0.954
$n = 1,000$	cr: 3.86%					cr: 3.68%				
β_1	1	-0.003	0.018	0.018	0.952	1	-0.004	0.019	0.019	0.924
β_2	-1	0.003	0.020	0.019	0.934	-1	0.004	0.021	0.020	0.934
$\boldsymbol{D}[1,1]$	0.25	-0.010	0.020	0.021	0.914	0.25	-0.006	0.022	0.023	0.944
$\boldsymbol{D}[2,2]$	0.25	-0.011	0.022	0.022	0.894	0.25	-0.004	0.025	0.025	0.930
$\boldsymbol{D}[1,2]$	0.125	-0.013	0.015	0.015	0.836	-0.125	-0.005	0.017	0.017	0.938

The dimension of the Hessian matrix is $(2n+p)\times(2n+p)$, where n is the sample size, and p is the total regression parameters we are estimating from two event processes. Increasing the sample size would largely inflate the dimension of the matrix and consequently the computational burden. It was found that frailty part of the Hessian matrix (2.17) is quite sparse, thus we set its off-block-diagonal part to be 0 in order to improve the computation speed and reduce the memory usage without causing much information loss. The trivial information loss for datasets with finite sample sizes is substantiated by comparing the estimating results in Table 2.2 with those from Table 2.1.

Table 2.2: Simulation results: Estimating regression coefficients and variance components based on 500 replicates with their *Hessian matrices sparsened* in the random effect part, $\lambda_{0k} = 1.5$ for $k = 1, 2$ and a median $\boldsymbol{m}_i$ of ≈ 4. Sample sizes vary from $\boldsymbol{n} = 100$ to $\boldsymbol{n} = 1,000$, and "cr" denotes censoring rate, or the proportion of subjects with $\boldsymbol{m}_i = 0$. We also included a case with $\boldsymbol{n} = 6000$ and about $75 - 80\%$ censoring rate, which is mimicking the application dataset DOPPS.

	True	$\boldsymbol{D}[1,2] > 0$				True	$\boldsymbol{D}[1,2] < 0$			
	Value	Bias	ESD	ASE	CP	Value	Bias	ESD	ASE	CP
$\boldsymbol{n} = 100$	cr: 3.81%					cr: 3.64%				
β_1	1	-0.006	0.062	0.056	0.932	1	-0.006	0.064	0.059	0.922
β_2	-1	0.004	0.062	0.059	0.940	-1	-0.000	0.064	0.062	0.932
$\boldsymbol{D}[1,1]$	0.25	-0.014	0.066	0.065	0.880	0.25	-0.010	0.074	0.072	0.916
$\boldsymbol{D}[2,2]$	0.25	-0.016	0.067	0.067	0.892	0.25	-0.007	0.079	0.077	0.900
$\boldsymbol{D}[1,2]$	0.125	-0.015	0.047	0.047	0.924	-0.125	-0.005	0.052	0.052	0.930
$\boldsymbol{n} = 500$	cr: 3.84%					cr: 3.70%				
β_1	1	-0.004	0.025	0.025	0.950	1	-0.004	0.028	0.026	0.920
β_2	-1	0.003	0.028	0.026	0.928	-1	0.005	0.029	0.028	0.934
$\boldsymbol{D}[1,1]$	0.25	-0.012	0.028	0.029	0.924	0.25	-0.007	0.032	0.033	0.936
$\boldsymbol{D}[2,2]$	0.25	-0.012	0.031	0.031	0.894	0.25	-0.007	0.034	0.034	0.936
$\boldsymbol{D}[1,2]$	0.125	-0.014	0.022	0.021	0.862	-0.125	-0.004	0.023	0.023	0.950
$\boldsymbol{n} = 1,000$	cr: 3.86%					cr: 3.68%				
β_1	1	-0.003	0.018	0.018	0.952	1	-0.004	0.019	0.019	0.922
β_2	-1	0.003	0.020	0.019	0.934	-1	0.004	0.021	0.019	0.934
$\boldsymbol{D}[1,1]$	0.25	-0.010	0.020	0.021	0.908	0.25	-0.007	0.022	0.023	0.934
$\boldsymbol{D}[2,2]$	0.25	-0.011	0.022	0.022	0.892	0.25	-0.005	0.025	0.025	0.928
$\boldsymbol{D}[1,2]$	0.125	-0.013	0.015	0.015	0.832	-0.125	-0.005	0.017	0.017	0.938
$\boldsymbol{n} = 6000$	cr: 78.1%		nr: 8722			cr: 79.4%		nr: 8460		
β_1	1	-0.011	0.020	0.018	0.880	1	-0.013	0.020	0.019	0.868
β_2	-1	0.007	0.027	0.025	0.918	-1	0.009	0.028	0.026	0.920
$\boldsymbol{D}[1,1]$	0.25	-0.019	0.030	-	-	0.25	-0.024	0.033	-	-
$\boldsymbol{D}[2,2]$	0.25	-0.014	0.035	-	-	0.25	-0.017	0.038	-	-
$\boldsymbol{D}[1,2]$	0.125	-0.009	0.021	-	-	-0.125	0.003	0.024	-	-

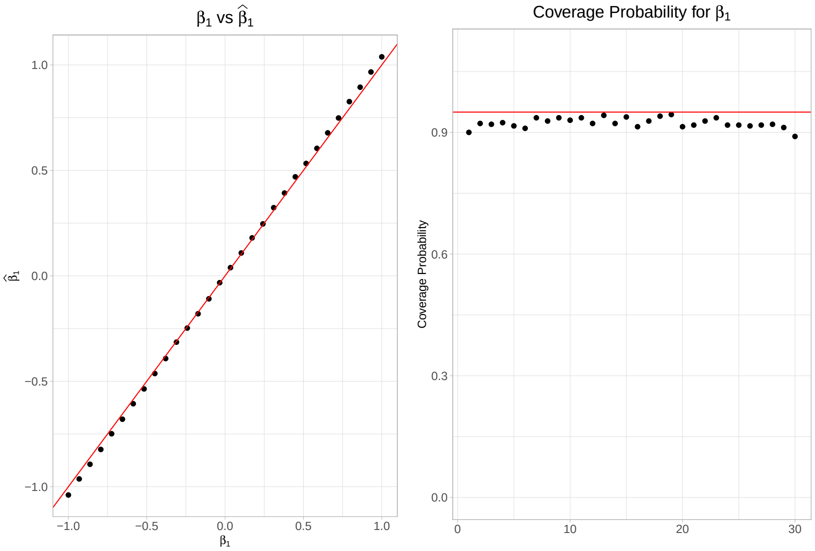


Figure 2.2: For event type 1, we have 30 parameters ranging from -1 to +1. The sub-figure on the left has shown that entries of $\boldsymbol{\beta}_1$ were plotted against the average value of their estimates, which is quite close to the red reference line $y = x$; On the right, we show the CPs for all the regression parameters form event 1, and the red line denotes the nominal value 0.95.

We also tested our method on more complicated model with a large number of regression parameters. We generated datasets with 200 subjects ($n = 200$), and there were 30 regression parameters for each event type, thus in total, there were 60 regression parameters ($p = 60$). The regression parameters for each event type is an arithmetic sequence with constant increments from -1 to $+1$. We denote $\boldsymbol{\beta}_k[c]$ as the c^{th} fixed effect for event type k, $k = 1, 2$. We set the variance matrix to be $\boldsymbol{D}[1,1] = \boldsymbol{D}[2,2] = 0.25$ and $\boldsymbol{D}[1,2] = -0.125$, and let $\lambda_{0k} = 1.5$. The median complete event number is $\widetilde{m}_i = 1$ with 33% of the subjects censored (with 0 complete event pairs). The regression parameter estimates are generally quite accurate when plotted against their corresponding true values in Figures 2.2 and 2.3. The CPs are slightly lower than their nominal value 95%, especially when the regression effects are large, e.g. the lowest one is CP = 0.878 from $\boldsymbol{\beta}_2[1] = -1$. The variance matrix estimation is quite close to its estimates in Tables 2.1 and 2.2.

Moreover, in order to test whether our proposed method works well for the DOPPS

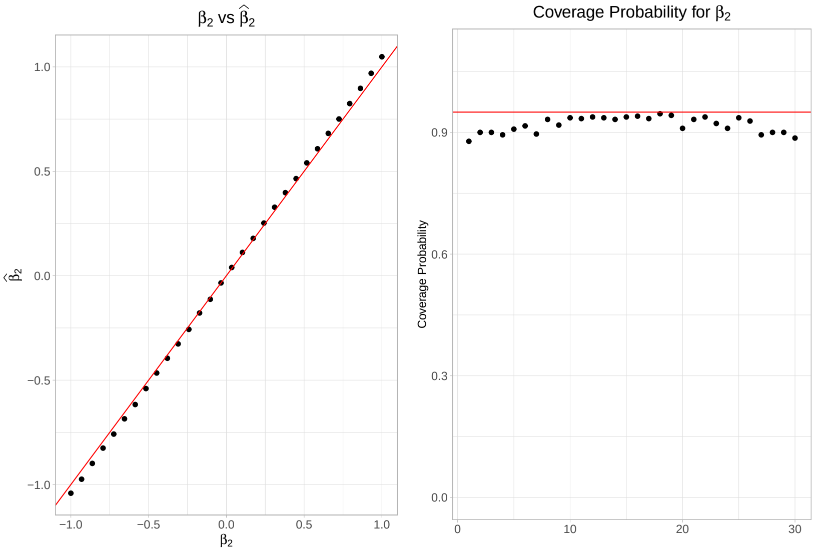


Figure 2.3: For event type 2, we have 30 parameters ranging from -1 to +1. The sub-figure on the left has shown that entries of $\boldsymbol{\beta}_1$ were plotted against the average value their estimates, which is quite close to the red reference line $y = x$; On the right, we show the CPs for all the regression parameters form event 2, and the red line denotes the nominal value 0.95.

data, we generated datasets similar to the DOPPS data: high censoring rate and large dimension of covariates. We mimic the DOPPS data by simulating $n = 6,000$ patients, with about $75-80\%$ censoring proportion ($m_i = 0$), 30 regression parameters for each event type ($p = 60$), and around $8,000$ event pairs; this is in comparison with the DOPPS dataset with 78.7% censored, and 29 regression parameters for each event ($p = 58$), and $9,065$ event pairs. We let 95% of the samples with censoring time $\boldsymbol{C} = 0.4$, and the other 5% with censoring time $\boldsymbol{C} = 10$, the baselines intensities are $\lambda_{0k} = 1.5$. The frailty is following a bivariate normal distribution $\boldsymbol{\gamma} \sim BVN(\mathbf{0}_2, \boldsymbol{D})$ where $\boldsymbol{D}[1,1] = \boldsymbol{D}[2,2] = 0.25$ and $\boldsymbol{D}[1,2] = \pm 0.125$. The gap times were generated from exponential distribution with intensity $\lambda_{0k} \exp(\boldsymbol{\beta}' \boldsymbol{Z}_{ijk})\gamma_{ik}$, where the covariates were simulated from a standard normal distribution. For each event type, there are 30 regression as an arithmetic sequence with constant increments ranging from -1 to $+1$. Eventually, there are around 9000 rows of records (pairs of admission and discharge). All the experiments were repeated 500 times.

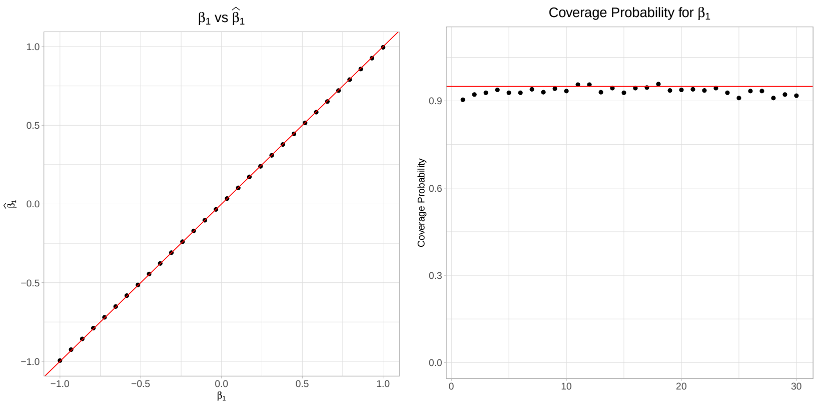


Figure 2.4: For event type 1, we have 30 parameters ranging from -1 to +1. The sub-figure on the left has shown that entries of $\boldsymbol{\beta}_1$ were plotted against the average value of their estimates, which is quite close to the red reference line $y = x$; On the right, we show the CPs for all the regression parameters form event 1, and the red line denotes the nominal value 0.95.

2.4.0.1 $\boldsymbol{D}[1, 2] = 0.125$

On average the censoring rate was 74.1% (range: 72.3-75.5%) and the total number of records (event pairs) was 8156 (range: 8002-8324). The average estimated random effect variance-covariance matrix is quite accurate with bias of the estimates for $\boldsymbol{D}[1, 1]$ to be -0.007 (ESE: 0.035), $\boldsymbol{D}[2, 2]$ to be -0.008 (ESE: 0.046), and $\boldsymbol{D}[1, 2]$ to be -0.005 (ESE: 0.025). The estimation of the regression parameters can be found in the Figures 2.4 and 2.5

2.4.0.2 $\boldsymbol{D}[1, 2] = -0.125$

On average the censoring rate was 74.5% (range: 72.7-76.2%) and the total number of records (event pairs) was 8092 (range: 7928-8246). The average estimated random effect variance-covariance matrix is quite accurate with bias of the estimates for $\boldsymbol{D}[1, 1]$ to be -0.008 (ESE: 0.036), $\boldsymbol{D}[2, 2]$ to be -0.005 (ESE: 0.051), and $\boldsymbol{D}[1, 2]$ to be -0.001 (ESE: 0.026).

We carried out simulations to compare the computational speed and estimation performance of the proposed method with the R package coxme (*Therneau*, 2018) un-

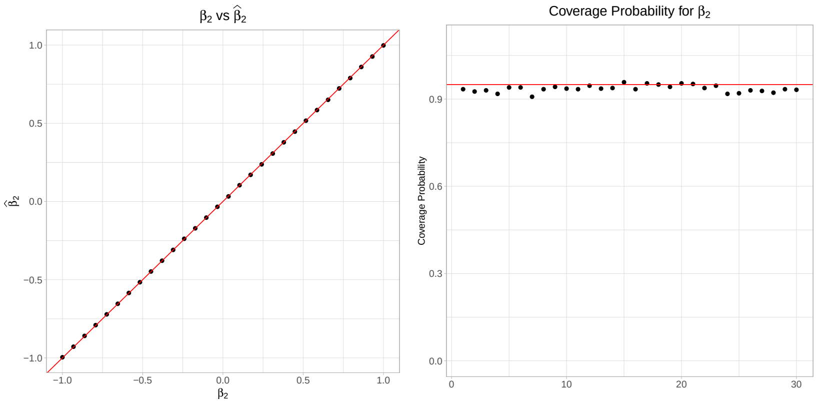


Figure 2.5: For event type 2, we have 30 parameters ranging from -1 to +1. The sub-figure on the left has shown that entries of $\boldsymbol{\beta}_1$ were plotted against the average value their estimates, which is quite close to the red reference line $y = x$; On the right, we show the CPs for all the regression parameters form event 2, and the red line denotes the nominal value 0.95.

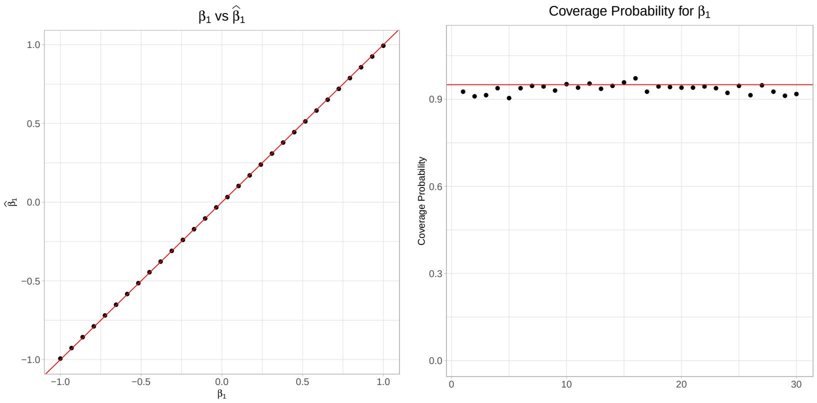


Figure 2.6: For event type 1, we have 30 parameters ranging from -1 to +1. The sub-figure on the left has shown that entries of $\boldsymbol{\beta}_1$ were plotted against the average value of their estimates, which is quite close to the red reference line $y = x$; On the right, we show the CPs for all the regression parameters form event 1, and the red line denotes the nominal value 0.95.

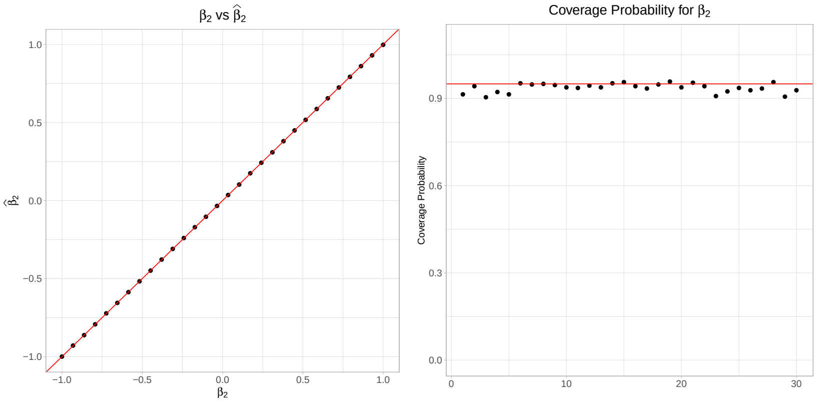


Figure 2.7: For event type 2, we have 30 parameters ranging from -1 to +1. The sub-figure on the left has shown that entries of $\boldsymbol{\beta}_1$ were plotted against the average value their estimates, which is quite close to the red reference line $y = x$; On the right, we show the CPs for all the regression parameters form event 2, and the red line denotes the nominal value 0.95.

der different settings. Though we have inhibited Hessian matrix sparsening in coxme and adjusted our codes to check the parameter convergence similarly, it is impossible for us to make it an entirely fair comparison: coxme can fit models with random effects (interaction terms with covariates), correct approximation bias in its likelihoods calculation, and conduct a comprehensive quality control on the input data. Therefore, the computation time (in seconds) is listed in Table 2.3 to demonstrate that the improved flexibility from the proposed method does not adversely affect the computation efficiency. For each sample, there were n individuals generated, each with a median of $\widetilde{m}_i$ complete event pairs. The regression parameters were estimated comparably well with both methods, and coxme has less bias than the proposed method in the variance matrix estimation, wherein additional information about the correlation sign of the two events needs to be provided.

The proposed likelihood ratio test was tested under different sizes, event frequencies, and covariance matrices. We fixed the regression parameters to be $\beta_1 = 1$ and $\beta_2 = -1$ for simplicity. We focuses on a big family of variance matrices whose variance components $\boldsymbol{D}[1,1] = \boldsymbol{D}[2,2] = 0.5$ and the covariance component varies: $\boldsymbol{D}[1,2] = \pm 0.25$ or ± 0.125 as representatives of strong ($|\rho| = 0.5$) or weak correlated

Table 2.3: Simulation results: Comparing the proposed method and coxme() with respect to computational speed via average run time based on 500 replicates

	True	Proposed method		coxme	
	Value	Bias	ESE	Bias	ESE
$\mathbf{n = 50, \widetilde{m_i} = 9}$	**time cost:**	**6.37s**		**77.2s**	
β_1	1	-0.006	0.061	-0.006	0.061
β_2	-1	-0.003	0.069	0.002	0.069
$\boldsymbol{D}[1,1]$	0.25	-0.011	0.076	-0.010	0.077
$\boldsymbol{D}[2,2]$	0.25	-0.000	0.081	0.001	0.081
$\boldsymbol{D}[1,2]$	-0.125	-0.006	0.057	-0.002	0.057
$\mathbf{n = 50, \widetilde{m_i} = 17}$	**time cost:**	**5.31s**		**136s**	
β_1	1	-0.002	0.046	-0.001	0.046
β_2	-1	-0.001	0.044	-0.001	0.044
$\boldsymbol{D}[1,1]$	0.25	-0.000	0.069	0.002	0.070
$\boldsymbol{D}[2,2]$	0.25	-0.001	0.067	0.001	0.068
$\boldsymbol{D}[1,2]$	-0.125	-0.004	0.052	-0.002	0.053
$\mathbf{n = 100, \widetilde{m_i} = 10}$	**time cost:**	**69.3s**		**745s**	
$\beta_1[1]$	-1	-0.004	0.040	-0.004	0.040
$\beta_1[2]$	-0.5	-0.003	0.032	-0.003	0.032
$\beta_1[3]$	0	-0.001	0.031	-0.001	0.031
$\beta_1[4]$	0.5	-0.000	0.034	-0.000	0.034
$\beta_1[5]$	1	0.005	0.042	0.005	0.042
$\beta_2[1]$	-1	-0.001	0.041	-0.001	0.041
$\beta_2[2]$	-0.5	-0.004	0.034	-0.004	0.031
$\beta_2[3]$	0	0.000	0.031	0.000	0.034
$\beta_2[4]$	0.5	0.002	0.034	0.002	0.034
$\beta_2[5]$	1	0.001	0.041	0.001	0.041
$\boldsymbol{D}[1,1]$	0.25	-0.004	0.051	-0.001	0.053
$\boldsymbol{D}[2,2]$	0.25	-0.008	0.051	-0.003	0.053
$\boldsymbol{D}[1,2]$	0.125	-0.007	0.038	0.001	0.039

Table 2.4: Power and type I error of the proposed likelihood ratio test, with $\lambda_{01} = \lambda_{02}$.

N	$\widetilde{m}_i$	T1E	Power			
$\boldsymbol{D}[1,2]$		0	0.25	-0.25	0.125	-0.125
50	≈ 4	0.066	0.656	0.646	0.236	0.174
100	≈ 4	0.038	0.902	0.918	0.394	0.366
100	≈ 9	0.042	0.976	0.982	0.516	0.482
100	≈ 17	0.056	0.994	0.992	0.612	0.542
100	≈ 34	0.054	1	1	0.648	0.686

($|\rho| = 0.25$) event pairs for power calculations, and $\boldsymbol{D}[1,2] = 0$ for type 1 error (T1E) evaluations. The proposed LRT test starts to performs well when the sample size and the event frequencies are not too small. In addition, if the magnitude of the correlation coefficient (ρ) is low, the proposed LRT is less likely to correctly detect the existence of a non-zero covariance. According to our simulations, T1Es are well controled ($\alpha = 0.05$) in all cases.

2.5 Application

The Dialysis Outcomes and Practice Patterns Study (DOPPS) is a well-known prospective, longitudinal, international study of hemodialysis patients. This study aims to improve the understanding of dialysis practices that are associated with better outcomes for end-stage renal disease patients. Details regarding the DOPPS study can be found in several reports (*Young et al.*, 2000; *Pisoni et al.*, 2004; *Robinson et al.*, 2012). Mortality, hospital admission and inpatient stay are important indicators of quality of life, and morbidity-related outcomes have arguably been under-utilized in the DOPPS and other studies of ESRD patients.

We applied our proposed methods to jointly analyze the time-to-readmission and time-to-discharge (from admission) alternating gap time sequence. Our objective was to determine the important predictors for each recurrent event process, and to quantify the correlation between the two processes. Our study population ($n = 6,032$)

included DOPPS Phase-5 adult patients (age ≥ 18) who entered the DOPPS within 3 months of initiating hemodialysis. Each member of the study cohort was followed for a maximum of 3 years, with the database closing on 12/31/2015. The study population included patients from 11 different countries, including Belgium, Canada, China, Gulf Coast Consortium, Germany, Italy, Japan, Spain, Sweden, the United Kingdom and the United States. The median age among DOPPS patients was 67, with 39.5% being female.

Our primary goal was to compare the hospital admission and the discharge event rates among dialysis patients by country. In particular, Belgium, Canada, China, Gulf Coast Consortium, Germany, Italy, Japan, Spain, Sweden, U.K., Asian-American and African-American are compared to the U.S. Caucasians (reference). Adjustment covariates included age, sex, height, vascular access (arteriovenous (AV) graft, central venous catheter, with AV fistula as the reference), and the following comorbid condition indicators: coronary artery disease (CAD), cancer, cerebral vascular disease (CVD), congestive heart failure symptoms (CHF), chronic obstructive pulmonary disease (COPD), peripheral vascular disease (PVD), stroke, diabetes, hypertension, neurological disorder, psychological disorder, and cellulitis. Table 2.5 lists results based on our model (2.2). DOPPS patients from Belgium ($e^{0.386} = 1.47$), Germany ($e^{0.98} = 2.66$), Italy ($e^{0.360} = 1.43$), Japan ($e^{0.842} = 2.32$), Sweden ($e^{0.507} = 1.66$) and U.K. ($e^{0.534} = 1.71$) had significantly higher covariate-adjusted hospital admission rates than U.S. Caucasians. In contrast, the hospital admission rates for patients in China ($e^{-0.621} = 0.537$) was approximately half that of U.S. Caucasians. With respect to length of hospital stay, patients from Canada ($e^{-0.807} = 0.446$), China ($e^{-1.198} = 0.302$), Germany ($e^{-0.433} = 0.649$), Italy ($e^{-0.667} = 0.513$), Japan ($e^{-0.624} = 0.526$), Spain ($e^{-0.456} = 0.634$) and the U.K. ($e^{-0.470} = 0.625$) had lower discharge rates (implying longer hospital stay) than U.S. Caucasians. We did not observe significant differences in the U.S. among races for either hospital admission

Table 2.5: Application of the proposed method to DOPPS data: Estimated regression parameters (bolded when $p < 0.05$).

	Admission			Discharge		
	Estimate	$\widehat{\text{SE}}$	P-value	Estimate	$\widehat{\text{SE}}$	P-value
Age (per 5 years)	**-0.026**	0.010	0.007	**-0.050**	0.010	<0.001
Height (per 5 cm)	-0.028	0.017	0.103	-0.005	0.018	0.804
Female	0.019	0.069	0.782	-0.065	0.073	0.373
Vascular access						
Arteriovenous graft	**0.524**	0.156	0.001	-0.034	0.149	0.822
Central venous catheter	**0.783**	0.059	<0.001	0.033	0.060	0.583
Comorbid conditions						
CAD	**0.447**	0.069	<0.001	-0.130	0.068	0.056
Cancer	**0.214**	0.082	0.009	**-0.208**	0.082	0.011
CVD	**0.177**	0.076	0.020	-0.070	0.075	0.345
Stroke	**0.190**	0.090	0.034	-0.004	0.088	0.968
CHF	0.078	0.068	0.251	0.028	0.070	0.692
Diabetes	0.053	0.056	0.347	-0.072	0.060	0.228
Hypertension	0.017	0.068	0.802	0.111	0.076	0.143
COPD	**0.264**	0.090	0.003	-0.075	0.087	0.387
Neurological disorder	**0.373**	0.101	<0.001	**-0.341**	0.096	<0.001
Psychological disorder	**0.293**	0.090	0.001	-0.076	0.088	0.389
PVD	0.111	0.078	0.156	0.124	0.079	0.115
Cellulitis	0.169	0.131	0.198	**-0.454**	0.126	<0.001
Countries						
Belgium	**0.386**	0.127	0.002	-0.041	0.125	0.741
Canada	0.234	0.125	0.060	**-0.807**	0.124	<0.001
China	**-0.621**	0.231	0.007	**-1.198**	0.253	<0.001
Gulf	-0.069	0.131	0.596	-0.054	0.138	0.693
Germany	**0.980**	0.100	<0.001	**-0.433**	0.097	<0.001
Italy	**0.360**	0.127	0.005	**-0.667**	0.130	<0.001
Japan	**0.842**	0.100	<0.001	**-0.624**	0.107	<0.001
Spain	-0.138	0.127	0.274	**-0.456**	0.135	0.001
Sweden	**0.507**	0.130	<0.001	-0.115	0.132	0.382
UK	**0.534**	0.135	<0.001	**-0.470**	0.140	0.001
USA: Asian	-0.148	0.305	0.628	-0.214	0.354	0.546
USA: African-American	-0.035	0.089	0.693	-0.020	0.102	0.848
USA: Caucasian	0	-	-	0	-	-

or discharge rates.

Comorbid conditions were generally positively associated with hospital admission and negatively associated with hospital discharge. Common significant predictors for both episodes include cancer and neurological disorder. CAD, CVD, stroke, COPD and psychological disorder were associated with significantly increased hospital admission rates, while cellulitis was significantly associated with increased discharge rates. Note that CAD (p-value= 0.056) was marginally significantly associated with discharges. The impact of age was found to be negatively associated with both the admission and discharge, though the difference for every 5-year increment was small (3-5%). Every 5 cm increment in height was associated with a 3% decrease in the hospitalization risk. In comparison to AV fistula (the most commonly adopted vascular access approach) AV graft and central venous catheter increased the hospital admission rate by 1.69 and 2.19 times, respectively. Note that each of the regression parameters should be interpreted as a conditional effect, given the unobserved frailties.

The estimated variance for time to admission (0.819) was larger than the the length of stay (0.375). The estimated covariance -0.139 implied that the two events were negatively correlated with a correlation coefficient $\widehat{\rho} = -0.251$. The LRT for the dependence between two event processes was 5.41 with a p-value of 0.02. These results indicate that those who had more frequent admission to a hospital would (through a lower discharge rate) tend to have a longer inpatient stay, and the association is significant.

Note that only 21.3% of patients experienced more than one hospitalization. The average length of stay in hospital was 8.8 days (range= $1 \sim 331$; median= 5). We tested the validity to implement the proposed estimating method on huge data sets with a small proportion of subjects that experienced multiple recurrences. The proposed methods appear to work well in data structures resembling DOPPS with respect to event rates and the variance-covariance matrix (data not shown). The convergence tolerance for parameter estimation is 10^{-6}.

2.6 Laplace Approximation

2.6.1 Introduction

Laplace approximation is widely used to compute both posterior distribution or moments under the Bayesian framework, and marginal likelihood of mixed models. This method is mainly proposed to solve likelihood integration problems using second-order Taylor expansion(*Solomon and Cox*, 1992; *Liu and Pierce*, 1993). The penalized survival model we proposed for the alternating recurrent events is an implementation example of this approximation method. This approximation method was found to work satisfactorily when there are frequent observations within groups, or when the dimension of integral is relatively small in comparison with the total number of observations. The estimation based on the marginal likelihood using Laplace approximation will be biased when the dimension of integral is high or the sample size is small, or when the shape of the integrand function departures from that of the Gaussian distribution. The ways to correct the bias has been studied extensively but not exclusively in the literature (*Shun and McCullagh*, 1995; *Breslow and Lin*, 1995; *Lin and Breslow*, 1996; *Ruli et al.*, 2016). In this section, we are investigating deeply on the inherent bias of Laplace approximation, comparing different types of Laplace approximation, and derive a correction term for our approximate marginal likelihood.

2.6.2 Different Laplace Approximation Methods

There are various ways to solve intractable integrals using Laplace approximation. I will summarize 3 major approaches (*Shun and McCullagh*, 1995; *Solomon and Cox*, 1992; *Liu and Pierce*, 1993) in this subsection and how their biases are corrected.

All the methods were re-summarized according to the very important works done by *Breslow and Lin* (1995), *Lin and Breslow* (1996) and *Shun and McCullagh* (1995).

2.6.2.1 Liu & Pierce Method

This is the method we employed in our proposed approach. Assume that we only consider a shared frailty model here. Let l_i be the log-likelihood of individual $i = 1, \dots, n$ and n is the total observations. Each individual has a random effect $b_i \sim N(0, \theta)$ and is independent to the others from other samples. There are multiple observations from each subject, thus we have $l_i = \sum_{j=1}^{n_i} l_{ij}$ and the log joint likelihood $l_i - 1/(2\theta)b_i^2 - \log(\theta)/2$. Through a Taylor expansion of b_i around $\widehat{b}_i$, which solves $l_i^{(1)}(b_i) - b_i/\theta = 0$, and let $\widehat{l}_i^{(k)} = l_i^k|_{b_i=\widehat{b}_i}$, its marginal likelihood can be approximated by

$$\begin{aligned} l_{mi} &= -\frac{1}{2}\log(\theta) + \log\int \exp\left\{l_i - \frac{b_i^2}{2\theta}\right\} db_i \\ &\approx -\frac{1}{2}\log(1-\theta\widehat{l}_i^{(2)}) + \widehat{l}_i - \frac{\widehat{b}_i^2}{2\theta} + \log\int\left(\frac{1}{\theta}-\widehat{l}_i^{(2)}\right)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\frac{1}{\theta}-\widehat{l}_i^{(2)}\right)(b_i-\widehat{b}_i)^2\right\} \\ &\qquad \left\{1+\frac{1}{6}\widehat{l}_i^3(b_i-\widehat{b}_i)^3+\frac{1}{24}\widehat{l}_i^{(4)}(b_i-\widehat{b}_i)^4\right\} db_i \\ &\approx \underbrace{\underbrace{-\frac{1}{2}\log(1-\theta\widehat{l}_i^{(2)}) + \widehat{l}_i - \frac{\widehat{b}_i^2}{2\theta}}_{LP_{1i}} + \overbrace{\frac{1}{8}\frac{\theta^2\widehat{l}_i^{(4)}}{\left(1-\theta\widehat{l}_i^{(2)}\right)^2}}^{Correction}}_{LPi2} \end{aligned} \tag{2.31}$$

Note that $LP_1 = \sum_i^n LP_{1i}$ is the usual approximate marginal likelihood, and $LP_2 = \sum_i^n LP_{2i}$ is the corrected version and the correction term which accounts for the bias.

The correction term can be re-written into

$$\frac{1}{8}\frac{\theta^2\widehat{l}_i^{(4)}}{\left(1-\theta\widehat{l}_i^{(2)}\right)^2} = \frac{1}{8}\frac{\widehat{l}_i^{(4)}}{\left(1/\theta-\widehat{l}_i^{(2)}\right)^2} = \frac{1}{8n_i}\frac{\theta^2\widehat{l}_i^{(4)}/n_i}{\left(1/n_i-\theta\widehat{l}_i^{(2)}/n_i\right)^2}.$$

When θ is small, the correction term approaches 0, while when θ is large, the bias can also be non-negligible. If the cluster size n_i becomes large, the bias or the correction term can become trivial, while when n_i is small, and especially when θ is large, we will need to take the bias into consideration. In general , it is found that the correction term will improve the estimation of the variance component (θ) while the fixed effect parameters seem not change much (*Breslow and Lin*, 1995). In the next subsection, I will derive the correction or bias term for our correlated frailty model.

2.6.2.2 Solomon & Cox Method

Solomon and Cox (1992) approximated l_{mi} in a similar way by expanding the integrand about the true mean (0) of the random effect. Let $l_{i0}^{(k)} = l_i^{(k)} |_{b_i=0}$, the approximate integral should become

$$\begin{aligned} l_{mi} &= -\frac{1}{2}\log(\theta) + \log\int \exp\left\{l_i - \frac{b_i^2}{2\theta}\right\} db_i \\ &\approx l_{i0} - \frac{1}{2}\log(1-\theta l_{i0}^{(2)}) + \frac{\theta l_{i0}^{(1)2}}{2(1-\theta l_{i0}^{(2)})} + \log\int\left(\frac{1}{\theta}-\widehat{l}_i^{(2)}\right)^{\frac{1}{2}} \\ &\qquad \exp\left[-\frac{1}{2}\left\{\frac{1}{\theta} - l_{i0}^{(2)}(b_i - \frac{l_{i0}\theta}{2(1-\theta l_{i0}^{(2)})})^2\right\}\right]\left\{1+\frac{1}{6}l_{i0}^{(3)}b_i^3 + \frac{1}{24}l_{i0}^{(4)}b_i^4\right\} db_i \\ &\approx \underbrace{\underbrace{l_{i0} - \frac{1}{2}\log(1-\theta l_{i0}^{(2)}) + \frac{\theta l_{i0}^{(1)2}}{2(1-\theta l_{i0}^{(2)})}}_{SC_{1i}} + \overbrace{\frac{\theta^2}{2}\left(l_{i0}^{(1)}l_{i0}^{(3)} + \frac{1}{4}l_{i0}^{(4)}\right)}^{correction}}_{SC_{2i}}. \end{aligned} \tag{2.32}$$

2.6.2.3 Shun & McCullagh Method

Shun and McCullagh (1995) studied the bias and its correction from the Laplace approximation of high dimensional integrals. For this class of Laplace approximation, an one-to-one transformation is conducted for the integrand before a Taylor expansion on its Jacobian. In order to demonstrate the high dimensional integrals, we rewrite the likelihood by a g function as $g(\boldsymbol{b}) = -n^{-1}\sum_{i=1}^{n}\{l_i - b_i^2/(2\theta)\}$, and the random vector $\boldsymbol{b} = [b_1, b_2, \ldots, b_n]$. The mapping $\boldsymbol{b} \mapsto \boldsymbol{u}(\boldsymbol{b})$ is sought to ensure $g(\boldsymbol{b}) - g(\widehat{\boldsymbol{b}}) = 1/2\boldsymbol{u}' g^{(2)}(\widehat{\boldsymbol{b}})\boldsymbol{u}$, where $\widehat{\boldsymbol{b}}$ achieves the minimum of $g(\boldsymbol{b})$. The Jacobian of the transformation $\boldsymbol{J}(\boldsymbol{b})$ is $1 + \boldsymbol{O}(\boldsymbol{u})$ in the neighborhood of $\boldsymbol{u} = \boldsymbol{0}$. The approximate marginal likelihood is based on the Taylor expansion of $\boldsymbol{J}(u)$ around the origin, and $\boldsymbol{J}^{(k)}(\boldsymbol{u})/k!$ is the kth coefficient for the Taylor expansion of $\boldsymbol{J}(u)$, we obtain the approximate likelihood with its correction term as

$$\begin{aligned} l_m &= \log \int_{\mathbb{R}^n} \exp\{-ng(\boldsymbol{b})\} d\boldsymbol{b} \\ &= -ng(\widehat{\boldsymbol{b}}) + \log \int_{\mathbb{R}^n} \exp\left\{-\frac{n}{2}\boldsymbol{u}'\boldsymbol{g}^{(2)}(\widehat{\boldsymbol{b}})\boldsymbol{u}\right\} \boldsymbol{J}(u) d\boldsymbol{u} \\ &\approx -ng(\widehat{\boldsymbol{b}}) - \frac{1}{2} log \left| n\boldsymbol{g}^{(2)}(\widehat{\boldsymbol{b}}) \right| + \underbrace{\frac{\boldsymbol{J}^{(2)}(u)}{2n\boldsymbol{g}^{(2)}(\widehat{b})}}_{correction} \end{aligned} \tag{2.33}$$

2.6.3 Bias Correction in Our Model

Now we return to our model with two correlated random intercepts for each subject. Let $\boldsymbol{\gamma}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$, and $\boldsymbol{\gamma}_i = [\gamma_{i1}, \gamma_{i2}]'$. Following LP type Laplace approximation (*Liu and Pierce*, 1993) and we can obtain its second-order estimator $\boldsymbol{LP}_2$ with its correction term. For simplicity, let $l_i^{(u,v)} = \partial^{(u+v)} l_i / \partial\boldsymbol{\gamma}_{i1}^{u} \partial\boldsymbol{\gamma}_{i2}^{v}$, and the combination

$C_k^u = k!/\{u!(k-u)!\}$.

$$
\begin{aligned}
l_{mi} &= -\frac{1}{2}\log|\boldsymbol{D}| + \log\int \exp\left\{l_i - \frac{1}{2}\boldsymbol{\gamma}_i'\boldsymbol{D}^{-1}\boldsymbol{\gamma}_i\right\} d\boldsymbol{\gamma}_i \\
&\approx \underbrace{-\frac{1}{2}\log|\boldsymbol{D}| - \frac{1}{2}\log\left|\boldsymbol{D}^{-1} - \widehat{\boldsymbol{l}}_i^{(2)}\right| + \widehat{l}_i - \frac{1}{2}\widehat{\boldsymbol{\gamma}}_i'\boldsymbol{D}^{-1}\widehat{\boldsymbol{\gamma}}_i}_{LP_{i1}} + \\
&\underbrace{\begin{aligned}
&\log\int \left|\boldsymbol{D}^{-1} - \widehat{\boldsymbol{l}}_i^{(2)}\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i)'\{\boldsymbol{D}^{-1} - \widehat{\boldsymbol{l}}_i^{(2)}\}(\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i)\right\} \\
&\qquad \left[1 + \frac{1}{6}\left\{\sum_{u=0}^{3} C_3^u \widehat{l}_i^{(u,3-u)}(\gamma_{i1} - \widehat{\gamma}_{i1})^u(\gamma_{i2} - \widehat{\gamma}_{i2})^{3-u}\right\}\right. \\
&\qquad\qquad \left. + \frac{1}{24}\left\{\sum_{u=0}^{4} C_4^u \widehat{l}_i^{(u,4-u)}(\gamma_{i1} - \widehat{\gamma}_{i1})^u(\gamma_{i2} - \widehat{\gamma}_{i2})^{4-u}\right\}\right] d\boldsymbol{\gamma}_i\,.
\end{aligned}}_{correction}
\end{aligned}
\tag{2.34}
$$

The correction term in (2.34) is to calculate the third and forth moments for a Gaussian distribution with mean $\widehat{\boldsymbol{\gamma}}_i$ and variance $[\boldsymbol{D}^{-1} - \widehat{\boldsymbol{l}}_i^{(2)}]^{-1} = \left[\begin{smallmatrix} d_{11} & d_{12} \\ d_{12} & d_{22}\end{smallmatrix}\right]$. The conditional distribution of γ_{i1} given γ_{i2} is univariate Gaussian with mean $\widehat{\gamma}_{i1} + d_{12}d_{22}^{-1}(\gamma_{i2} - \widehat{\gamma}_{i2})$ and variance $d_{11} - d_{12}d_{22}^{-1}d_{12}$. Thus $E\{(\gamma_{i1} - \widehat{\gamma}_{i1})(\gamma_{i2} - \widehat{\gamma}_{i2})^2\} = d_{12}d_{22}^{-1}E\{(\gamma_{i2} - \widehat{\gamma}_{i2})^3\} = 0$. Thus (2.34) can be simplified to

$$
\begin{aligned}
LP_{i2} = &LP_{i1} + \log\int \left|\boldsymbol{D}^{-1} - \widehat{\boldsymbol{l}}_i^{(2)}\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i)'\{\boldsymbol{D}^{-1} - \widehat{\boldsymbol{l}}_i^{(2)}\}(\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}_i)\right\} \\
&\qquad \left[1 + \frac{1}{24}\left\{\sum_{u=0}^{4} C_4^u \widehat{l}_i^{(u,4-u)}(\gamma_{i1} - \widehat{\gamma}_{i1})^u(\gamma_{i2} - \widehat{\gamma}_{i2})^{4-u}\right\}\right] d\boldsymbol{\gamma}_i \\
= &\, LP_{i1} + \frac{1}{8}\left\{\widehat{l}_i^{(4,0)}d_{11}^4 + \widehat{l}_i^{(0,4)}d_{22}^4 + 4\widehat{l}_i^{(3,1)}d_{12}d_{11}^3 + 4\widehat{l}_i^{(1,3)}d_{12}d_{22}^3 + \right. \\
&\qquad \underbrace{\left. 2\widehat{l}_i^{(2,2)}(d_{11}^2 + d_{12}^4 d_{22}^{-2} - 2d_{11}d_{12}^2d_{22}^{-1} + 3d_{12}^2d_{22}^2)\right\}}_{correction}
\end{aligned}
\tag{2.35}
$$

2.7 An Extension: More than Alternating Recurrent Events

The estimating approach we proposed in this chapter can be smoothly implemented to multiple event types. For example, in multitype event data, the multiple types of recurrent events could represent different levels of severity, different physiological processes, and different type of health problems (*Cook and Lawless*, 2007). Another example would be clustered different types of failure time data from the same hospital, or competing risks from patients in the same facility. We only consider clustering from facilities here, without individual-level heterogeneity in models for this type of example. Our proposed correlated frailty model 2.2 can easily be extended to accommodate multiple types of events by including correlated frailties. Like many other Laplace approximation based methods, the proposed approach tends to give less estimation bias, especially in terms of the variance components, for data with larger clusters. Simulations and theoretical proofs can be found in Section 2.6 and Appendix A.2. We revise the model for alternating recurrent events in (2.2) to accommodate competing risks and clustered multiple failure times with any $K > 1$:

$$\lambda_{ijk}(t \mid \boldsymbol{Z}_i, \boldsymbol{R}_{ik}, \boldsymbol{\gamma}) = \lambda_{0k}(t)\exp(\boldsymbol{\beta}'_k\boldsymbol{Z}_{ijk} + \boldsymbol{\gamma}'\boldsymbol{R}_{ik}), \tag{2.36}$$

where $i = 1, \ldots, I$ and I is the number of clusters (facilities); $j = 1, \ldots, J_i$ and J_i is the number of individuals in cluster i; $k = 1, \ldots, K$ and K is the number of event types we are considering. Through a similar estimating procedure in Section 2.3, the estimator should be

$$\widehat{\boldsymbol{D}}^{\#}_{K\times K} = \frac{1}{n}\sum_{i=1}^{n}\left\{\left[\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})^{-1}\right]_{blk_i} + \widehat{\boldsymbol{\gamma}}_i\widehat{\boldsymbol{\gamma}}'_i\right\}, \tag{2.37}$$

whose dimension is $K \times K$. It can also be shown to be positive-definite following Appendix A.1.

2.7.1 Simulation Results

We generate $\boldsymbol{I} = 50$ facilities and each facility has a median of $\widetilde{m}_i$ events. We assumed a similar simulation setting as we did for alternating recurrent events by letting $\lambda_{0k} = 1.5$, but there are $\boldsymbol{K} = 3$ event types. $\boldsymbol{\beta}_k$ is the regression parameters for event type k, and $\boldsymbol{\beta}_k[c]$ is its effect for the cth covariate. We let the stacked frailty vector from facility $\boldsymbol{\gamma}_i \sim \boldsymbol{MVN}(\mathbf{0}_3, \boldsymbol{D}_(3 \times 3))$ where its entries can be located by a pair of indices [a,b], where a indicates row and b indicates the column. We let symmetric variance matrix $\boldsymbol{D}$ satisfy $\boldsymbol{D}[1,1] = \boldsymbol{D}[2,2] = \boldsymbol{D}[3,3] = 0.5$ and $\boldsymbol{D}[1,2] = \boldsymbol{D}[1,3] = -\boldsymbol{D}[2,3] = 0.25$. Our simulation results are recorded in Table 2.6.

Table 2.6: Simulation results for an extension of clustered events with more than two event types

	True	Case 1				True	Case 2			
	Value	Bias	ESE	ASE	CP	Value	Bias	ESE	ASE	CP
	$\widetilde{m}_i = 13$		time cost:	138 s		$\widetilde{m}_i = 14$		time cost:	120 s	
$\beta_1[1]$	1	0.003	0.050	0.049	0.936	1	0.001	0.048	0.048	0.948
$\beta_1[2]$	-	-	-	-	-	0.3	0.004	0.039	0.039	0.950
$\beta_2[1]$	-1	-0.002	0.051	0.050	0.942	0.5	0.004	0.041	0.041	0.968
$\beta_2[2]$	-	-	-	-	-	-0.5	-0.000	0.043	0.041	0.930
$\beta_3[1]$	-0.5	-0.001	0.043	0.043	0.946	-0.5	-0.000	0.042	0.042	0.948
$\beta_3[2]$	-	-	-	-	-	0.1	-0.005	0.039	0.039	0.956
$\boldsymbol{D}[1,1]$	0.5	-0.008	0.119	-	-	0.5	-0.008	0.115	-	-
$\boldsymbol{D}[2,2]$	0.5	0.007	0.124	-	-	0.5	0.004	0.126	-	-
$\boldsymbol{D}[3,3]$	0.5	0.003	0.124	-	-	0.5	0.005	0.135	-	-
$\boldsymbol{D}[1,2]$	0.25	-0.006	0.091	-	-	0.25	-0.015	0.090	-	-
$\boldsymbol{D}[2,3]$	-0.25	-0.010	0.092	-	-	-0.25	-0.010	0.096	-	-
$\boldsymbol{D}[1,3]$	0.25	0.018	0.094	-	-	0.25	-0.007	0.091	-	-

2.8 Discussion

In this chapter, we propose a correlated bivariate frailty model for alternating recurrent event (gap time) processes. The regression parameter estimation proceeds through penalized partial likelihood. We also derived a variance-covariance estimator of the bivariate frailty in a recursive estimating formula. Through simulations, the methods were demonstrated to work well for both regression parameters and variance components. The proposed methods were applied to simultaneously analyze hospitalization admissions and discharges among end-stage renal disease patients.

The proposed estimating approach does not require the sign of the correlation between the two recurrent processes when building the model. In the context of our motivating example, it is possible that the longer length-of-stay tends to accompany hospitalization for a more severe episode which, in turn, can be associated with shorter time to readmission. On the other hand, longer hospital stay probably may indicate better care and, as such, be associated with longer time to readmission. Our method would end up with a more flexible estimation on the variance components without assuming whether the admission and discharge events are positively or negatively correlated. Moreover, the proposed LRT would provide information whether a joint model with two procedures is necessary or not. This is important in practice, since fitting two separate frailty models requires less computation than fitting a joint frailty model.

To be able to analyze large data sets, part of our program is written using RcppArmadillo (*Eddelbuettel and Sanderson*, 2014). Besides, the proposed computing algorithm yields very reasonable computation times. When the sample size is large, we recommend using the "sparse" Hessian matrix option to reduce the memory cost.

Moreover, we spent a section studying Laplace approximation. We compared different types of Laplace approximation, evaluated the approximation bias and proposed a correction term for the approximate marginal likelihood of our model.

Last but not least, the methods in this report is open to extended to other settings, e.g. clustered survival outcomes and competing risks. Technically, the method can accommodate more than two event types. Our preliminary simulation results also have demonstrated its potential for a wider use in facility-level clustered event data. Furthermore, although the frailties we consider primarily in this chapter represent subject-specific intercepts, one could also include frailties corresponding to one of more covariates.

CHAPTER III

An Estimating Equation Framework for a Flexible Class of Semiparametric Frailty Models

3.1 Introduction

Recurrent event data are commonly encountered in both experimental and observational studies. Examples of recurrent events include repeated hospital admissions, recurrent infections, tumor incidences. Frequently used methods to analyze recurrent event data include extended Cox models (intensity functions) (*Prentice et al.*, 1981; *Andersen and Gill*, 1982) and rate functions (*Pepe and Cai*, 1993; *Lawless and Nadeau*, 1995; *Lin et al.*, 2000). Recurrent events are almost always clustered within-individual, and there may exist some additional hierarchical grouping or other complex correlations; e.g., individuals treated in the same facility or hospital. As another example, subjects may be at risk for different types of events simultaneously.

It is usually the case that heterogeneity and dependence within subjects (and/or clusters of subjects) exists; this results, in part, from the observed covariates not fully capturing the set of factors predicting the event process. As a remedy, random effects or frailties have been extensively adopted into the modeling of such processes in order to account for such heterogeneity. For simple heterogeneity within subjects, one may incorporate a shared frailty and assume the independence within-subject conditional on the frailty. For dependence between various event types, one may introduce correlated frailties and assume that, conditional on the frailties, the events are independent. If hierarchical structures are present, multiple levels of frailties can also be included under the assumption that conditional independence is also satisfied within each subcluster given the frailties.

Fully parametric models (typically a continuous distribution) are usually assumed for the frailties. Popular options have mostly been restricted to the gamma, log-normal, and positive-stable distributions, largely motivated by computational convenience. Different choices of the frailty distribution may lead to discrepant inferential conclusions, since usually the regression parameter estimation procedure is at least somewhat dependent on the variance component estimation. Since the frailties are unobserved, it is not possible to verify (or even assess empirically) the assumed frailty distribution. Correspondingly, among the common choices for the frailty distribution, none has emerged as being most likely to hold. Most of the time, this decision is based on the practitioner's preference considering computational and interpretational convenience. For example, gamma distributed frailties are much more compatible with the EM algorithm than other distributions. Log-normal frailties align well with various numerical techniques (e.g., Laplace approximations) and confer easier interpretation. Moreover, there is a scarcity of reliable diagnostic approaches to check the validity of the chosen frailty distribution.

Non-parametric frailties, or discretely distributed frailties, have been proposed as a flexible alternative for modeling correlated time-to-event data (*Guo and Rodriguez*, 1992; *Li et al.*, 1998; *Caroni et al.*, 2010; *Gasperoni et al.*, 2018). These methods, which can be viewed as a finite mixture of parametric survival models (*Laird*, 1978; *Heckman and Singer*, 1982, 1984), basically assume that each cluster has an unobserved frailty. They are quite similar to assigning a fixed effect for each cluster with a multiplicative indicator to show which effect is present. The nonparametric frailty methods liberate the frailties from assuming any known type of distributions. However, the computational burden grows dramatically when many clusters are present. *Ma* (1999) and *Ma et al.* (2003) developed an iterative two-step approach of calculating orthodox best linear unbiased predictors (BLUP) for frailties without distribution restrictions, and estimating the baseline, regression parameters, and variance components. As was discussed by *Moreno* (2008), this proposal estimates the baseline nonparametrically, and BLUPs are updated at each iteration, which would easily cause an explosion of computational burden when the sample size is large. Indeed, most existing non-parametric methods, including those developed under the Bayesian framework (*Walker and Mallick*, 1997), suffer from intensive computation when the sample size or number of clusters increases. This is especially true for recurrent event data, wherein each subject typically represents a cluster.

To alleviate the intensive computational burden in *Xue and Brookmeyer* (1996) for multivariate event outcomes, *Xue* (1998) developed an estimating-equation approach for bivariate correlated frailty models without imposing distributional assumptions on the frailties; the procedure required the baseline hazards to be fully parametric. *Wang et al.* (2001) also proposed an approach treating the frailties nonparametrically and as nuissance parameters. The method of *Wang et al.* (2001) assumes a subject-specific non-stationery Poisson process and a shape function (the conditional density) which is invariant across subjects. In contrast to *Xue* (1998), *Wang et al.* (2001) developed procedures which estimate the shape of the baseline hazards nonparametrically; the methods did not provide any estimation approach for the variance components. The method of *Wang et al.* (2001) was extended to accommodate joint models with terminal events (*Huang and Wang*, 2004) and longitudinal observations (*Sun et al.*, 2007).

The majority of the aforementioned methods were developed under specific hierarchical data structures (e.g., subject-level clustered; hierarchically clustered; and correlated multivariate failure times), without setting a universal framework to accommodate the various types of clustering. In this report, we propose a general framework for analyzing recurrent events through semiparametric frailty models. Models assumed under the proposed methods feature both baseline rates and frailty distribution that are left unspecified. The frailty models we consider in this report include shared frailty models, correlated frailty models, and nested frailty models; collectively, this covers a considerable proportion of the correlated failure times settings frequently encountered in practice. The proposed approaches provide consistent estimation of baseline, regression parameters, and variance components, with a relatively fast computational speed and an asymptotic normality property.

The remainder of this chapter is organized as follows. We introduce the notations and different models in Section 3.2. Their estimating approaches are developed in Section 3.3. Simulations to evaluate the proposed methods in finite sample sizes can be found in Section 3.5. Asymptotic properties are summarized in Section 3.4. We apply the proposed approaches to an analysis on end-stage renal disease patients of the Dialysis Outcomes and Practice Patterns Study (DOPPS) in Section 3.6. Some concluding remarks are provided in Section 3.7.

3.2 Notation and Class of Models

Suppose we focus on the occurrence of events within the time interval $[0, \tau]$. In most practical applications, $N^*(t)$ would be the number of recurrent events occurring up to t, for $t \in [0, \tau]$. $N^*(t)$ is naturally clustered by subject; i.e., recurrent events from the same subject are often positively associated. We propose three general classes of frailty models. We assume that $N^*(t)$ follows a nonstationery Poisson process given the frailties, γ, and covariates, $\boldsymbol{Z}$. The Poisson process assumption is important here since it endows the equality between the event rate and intensity. Let $N^*(t)$ be subject to right censoring, with censoring time $\boldsymbol{C}^*$ assumed independent of $N^*(t)$, conditional on the $\boldsymbol{Z}$ and γ. Let the observed censoring time be $C = \min(\tau, C^*)$, then $Y(t) = I(t \leq C)$ denotes the at-risk process, and thus the observed event process $N(t) = \int_0^t Y(s)dN^*(s)$ with $dN^*(t) = N^*(t) - N^*(t^-)$. The number of uncensored event is then denoted by $N(t) = N^*(t \wedge C)$. Moreover, as is typical of frailty modeling, we assume that the frailties are independent of the explanatory covariates for identifiability.

In the subsections that follow, we describe three types of data structures that can be accommodated by the proposed methods.

3.2.1 Model A: Shared Frailty Model

Let γ be a frailty (nonegative, latent) variable, shared among all the events from the same subject. The distribution of γ is left unspecified, with the only moment restrictions being $E(\gamma) = 1$ and $E(\gamma^4) < \infty$. As aforementioned, the frailty γ is independent of $\boldsymbol{Z}$ and given $(\boldsymbol{Z}, \gamma)$, $N^*(t)$ follows a nonstationary Poisson process with event rate

$$E(dN^*(t)|\boldsymbol{Z}, \gamma) = \lambda_0(t)\gamma e^{\boldsymbol{\beta}'\boldsymbol{Z}}, \tag{3.1}$$

where the baseline rate $\lambda_0(t)$ is assumed to be a continuous function. We assume that the observed data represent n replicates of independent and identically distributed random draws indexed by i, $\{N_i(t), C_i, \boldsymbol{Z}_i, \gamma_i;\ t \in [0, C_i]\}$. Note that the event process can also be described with event times $0 \leq t_{i,1} \leq t_{i,2} \leq \cdots \leq t_{i,m_i} \leq C_i$, where $M_i = N_i(C_i)$ is the number of observed events for subject i.

3.2.2 Model B: Correlated Frailty Model

It is not uncommon to see multivariate event outcomes, or different types of events from one subject in clinical studies. For instance, hospitalizations may result from different causes (e.g., infection, comorbidity, etc). The multiple hospitalizations can be viewed as multivariate recurrent events; associations can be captured through a multivariate frailty variate with correlated elements. For simplicity, we focus on the bivariate case; we let $N_1^*(t)$ and $N_2^*(t)$ denote the counting processes recurrent events of type 1 and type 2 respectively. We let $\boldsymbol{\gamma} = [\gamma_1, \gamma_2]'$ be the correlated frailty vector. Their variances $Var(\gamma_1)$ and $Var(\gamma_2)$ account for the clustering dependence within the two event types from the same subject, and covariance $Cov(\gamma_1, \gamma_2)$ accounts for the association between the two events. We denote the correlation (after conditioning on the covariates) between the two events by $\rho = cor(\gamma_1, \gamma_2) \in (-1, 1)$. The model is given by

$$E(dN_j^*(t)|\boldsymbol{Z}, \gamma_j) = \lambda_{0j}(t)\gamma_j e^{\boldsymbol{\beta}_j'\boldsymbol{Z}}, \; j = 1, 2. \tag{3.2}$$

The notation is analogous to that from model (3.1), except for the additional index j to distinguish event types. Note that the covariate vector $\boldsymbol{Z}$ need not be identical for different event models, which can be easily controlled via setting part of $\boldsymbol{\beta}_j$ to be zero. Given the both frailties (γ_1 and γ_2) and covariates, the paired $N_j^*(t)$ are assumed to follow independent Poisson processes within each subject. The independent identically distributed draws are realized in observations $\{N_{1i}(t), N_{2i}(t), C_i, \boldsymbol{Z}_i, \boldsymbol{\gamma}_i; \; t \in [0, C_i]\}$, $i = 1, \ldots, n$.

3.2.3 Model C: Nested Frailty Model

In many settings, one level of clustering (subject-level) may not be sufficient to capture the correlation structure of the data. For instance, different patients can be treated in the same hospital, in which case they are subject to two different sources of heterogeneity: one from the hospital (level 1) and the other from within-subject (level 2). To fully describe the clustering structure, one could introduce two frailties; one for the hospital (denoted by ϵ_k) where $k = 1, \ldots, K$; and the second for the patient (denoted by γ_{ki}), where $i = 1, \ldots, I_k$, i.e., I_k subjects nested under the cluster or hospital k. For identifiability, the two levels of frailties are assumed to be mutually independent; i.e., $\epsilon_k \perp \gamma_{ki}$. The nested model (Model C) is thus formulated as follows,

$$E(dN^*_{ki}(t)|\mathbf{Z}_{ki}, \epsilon_k, \gamma_{ki}) = \lambda_0(t)\epsilon_k\gamma_{ki}e^{\boldsymbol{\beta}'\mathbf{Z}_{ki}}, \tag{3.3}$$

for $k = 1, \ldots, K$ and $i = 1, \ldots, I_k$. One can view subjects from the shared first level of clustering (e.g., hospitals) as iid draws $(N_{ki}(t), C_{ki}, \epsilon_k, \gamma_{ki}, \mathbf{Z}_{ki})$, indexed by $i = 1, \ldots, I_k$ and $k = 1, \ldots, K$. Moreover, the number of subjects in each hospital or I_k is assume to be independent of $(N_{ki}(t), C_{ki}, \epsilon_k, \gamma_{ki}, \mathbf{Z}_{ki})$. The number of subjects is given by $\sum_{k=1}^{K} I_k$. Note that this model can be extended to accommodate more than two levels, but the estimation accuracy will greatly decline as the complexity of the hierarchical data structure increases.

3.3 Estimation

3.3.1 Baseline Shape

One can describe the shape of baseline rate through the function $f(t) = \lambda_0(t)/\Lambda_0(\tau)$, where $\Lambda_0(\tau) = \int_0^\tau \lambda_0(s)ds$ is the cumulative baseline rate at the ending time τ; this quantity has been termed the baseline “size” (*Wang and Huang*, 2014). In the absence of time-varying covariates/frailties, the shape function$f()$ is invariant across subjects. For example, in shared frailty model (Model A), we have

$$f(t) = \frac{\lambda(t|z,\gamma)}{\Lambda(\tau|z,\gamma)} = \frac{\lambda_0(t)\exp(\boldsymbol{\beta}'z)\gamma}{\Lambda_0(\tau)\exp(\boldsymbol{\beta}'z)\gamma} = \frac{\lambda_0(t)}{\Lambda_0(\tau)}, t \in [0,\tau]. \tag{3.4}$$

Note that both Model B and Model C satisfy (3.4), since the covariate and frailty components cancel out. Note that there are different baseline rates for correlated frailty models (Model B), and thus their shape functions need to be defined separately for each event type.

The cumulative shape function is defined as $F(t) = \int_0^t f(s)ds = \Lambda_0(t)/\Lambda_0(\tau)$. Estimation of the baseline shape is based on a nonparametric maximization of a conditional likelihood,

$$\begin{aligned} L_c &= \prod_{i=1}^{n} p(t_{i,1}, t_{i,2}, \cdots, t_{i,m_i} \mid z_i, c_i, \gamma_i, m_i) \\ &= \prod_{i=1}^{n} m_i! \prod_{j=1}^{m_i} \frac{f(t_{ij})}{F(c_i)}, \end{aligned} \tag{3.5}$$

which is free of regression parameters, covariates, and frailties. According to *Wang et al.* (1986), the nonparametric maximum likelihood estimator (NPMLE) of $F(t)$ has the following product-limit representation,

$$\hat{F}(t) = \prod_{s_{(l)}>t} \left(1 - \frac{d_{(l)}}{R_{(l)}}\right), \tag{3.6}$$

where $\{s_{(l)}\}$ are the ordered and distinct values of the event times $\{t_{ij}\}$, $\{d_{(l)}\}$ is the number of observed events at $s_{(l)}$, and $\{R_{(l)}\}$ is the total number of events with event time and observation censoring time satisfying $\{t_{ij} \leq s_{(l)} \leq c_i\}$. The quantity in (3.6) is the analog of the Kaplan-Meier estimator, if treating the ending time τ to be the origin then estimating backwards in time.

3.3.2 Regression Parameters and the Baseline Size

Estimation of the regression parameters $\boldsymbol{\beta}$ and the size parameter $\Lambda(\tau)$ follows the first moment condition of the Poisson process (*Wang et al.*, 2001; *Huang and Wang*, 2004). The observed event counts can also be expressed in $M = N(C)$ satisfying the moment condition

$$E(MF(C)^{-1} \mid z) = E\{E(N(C)F(C)^{-1} \mid z, \gamma, C) \mid z\} = \exp(\boldsymbol{\beta}' z)\Lambda_0(\tau). \tag{3.7}$$

Note that (3.7) does not depend on the distribution of C, since $E(N(C)F(C)^{-1} \mid \boldsymbol{z}, \gamma, C) = \exp(\boldsymbol{\beta}'\boldsymbol{z})\Lambda_0(\tau)\gamma$ which is free of C. Based on the independence property $E(\gamma \mid \boldsymbol{z}) = E(\gamma) = 1$ and the NPMLE in (3.5), we obtain the unbiased estimating equation for $\boldsymbol{\theta} = [log(\Lambda_0(\tau)), \boldsymbol{\beta}']'$ as follows,

$$\sum_{i=1}^{n} \bar{z}_i(m_i \widehat{F}^{-1}(c_i) - e^{\boldsymbol{\theta}'\bar{z}_i}) = 0, \tag{3.8}$$

where $\bar{z}_i = [1, \boldsymbol{z}_i']'$. The nested Model C also satisfies (3.7) and can be estimated through (3.8), as the two layers of frailties are assumed to be independent of each other and covariates; note that the frailties can be rescaled here to have unit mean.

Correlated Model B has multiple event types and consequently different batches of regression parameters and baseline sizes. Each event process $N_j(t)$ would satisfy the moment condition in (3.7), such that parameters $\boldsymbol{\theta}_j$ can be estimated through separate estimating equations analogous to (3.8). This also implies that, unlike many other existing frailty methods (*Klein*, 1992; *Nielsen et al.*, 1992; *Xue and Brookmeyer*, 1996; *Ripatti and Palmgren*, 2000), the estimation of the regression parameters (and the baseline size) treats the variance components as nuisance parameters, which greatly reduces the computational burden.

3.3.3 Variance Components

We propose to estimate the variance components through a second moment condition derived from the fact the variance of a Poisson process is identical to its mean. This equality suggests the following moment condition for Model A,

$$\begin{aligned} E((M^2 - M)F^{-2}(C) \mid \boldsymbol{z}) =& E[E\{(M^2 - M)F^{-2}(C) \mid C, \gamma\} | \boldsymbol{z}] \\ =& \Lambda_0(\tau)^2 e^{2\boldsymbol{\beta}'\boldsymbol{z}} E(\gamma^2). \end{aligned} \tag{3.9}$$

Given the baseline shape, size, and regression parameters having already been estimated following Subsections 3.3.1 and 3.3.2, the corresponding estimating equation for the variance component is

$$\sum_{i=1}^{n} \left\{ (m_i^2 - m_i)\widehat{F}^{-2}(c_i) - \widehat{\Lambda}_0(\tau)^2 \exp(2\widehat{\boldsymbol{\beta}}' \boldsymbol{z}_i) E(\gamma^2) \right\} = 0. \tag{3.10}$$

Note that estimating equation (3.10) can produce a similar batch of estimating equations for all the parameters $\boldsymbol{\theta}$ by treating all the parameters in 3.3.2 unknown and taking the first-order derivative. To that end, if the estimating equation in (3.8) is also considered, we will end up with a larger number of estimating equations than the number of entries in $\boldsymbol{\theta}$. In order to find optimal solutions, we evaluated both generalized method of moments (GMM) and empirical likelihood (EL), which have been extensively studied (*Hansen*, 1982; *Hansen et al.*, 1996; *Smith*, 1997). After carrying out a lengthy series of simulations (Appendix B.2), we found that including the second moment restriction for the parameter estimation tend to largely increase instability, that is, both GMM and EL produced biased estimators, although the estimating standard errors were slightly reduced in comparison with the estimating equation (EE) method (solely using (3.10) for variance components). It appears that the only case that GMM can beat the proposed EE method is when the sample size is sufficiently large (e.g. n=10,000). Possible reasons include that the high correlation between the two moment conditions makes the improvement from the additional second moment condition negligible, and that the relatively large variation of the second moment condition causes a large instability of general estimation procedure. Therefore, in the rest of this chapter, we avoid using the second moment condition for regression parameter estimation.

For correlated frailty Model B, the variances $Var(\gamma_j)$ can be estimated through estimating equations analogous to (3.10) by adding an index to distinguish the different event types in Model B (3.2). The covariance between the two event types can be estimated based on the following moment condition,

$$
\begin{aligned}
E((M_1 M_2 F_1^{-1}(C_1) F_2^{-1}(C_1) \mid z) =& E[E\{M_1 M_2 F_1^{-1}(C_1) F_2^{-1}(C_2) \mid z, C_1, C_2, \gamma_1, \gamma_2\} \mid z] \\
=& \exp(\beta_1' z + \beta_2' z)\Lambda_{01}(\tau)\Lambda_{02}(\tau)E(\gamma_1\gamma_2).
\end{aligned}
\tag{3.11}
$$

Note that $M_j = N_j(C_j)$ denotes the observed counts by the censoring time C_j. Suppose m_{1i} and m_{2i} denote the observed number of the two recurrent event types from subject i, the estimating equation for the covariance $E(\gamma_1\gamma_2)$ hence can be obtained as

$$
\sum_{i=1}^{n} \left\{(m_{1i} m_{2i})\widehat{F}_1^{-1}(c_{1i})\widehat{F}_2^{-1}(c_{2i}) - \widehat{\Lambda}_{10}(\tau)\widehat{\Lambda}_{20}(\tau)\exp(\widehat{\boldsymbol{\beta}}_1' z_i + \widehat{\boldsymbol{\beta}}_2' z_i)E(\gamma_1\gamma_2)\right\} = 0. \tag{3.12}
$$

The second moments for nested Model C with the two levels of clustering frailties cannot be identified via

$$
E((M^2 - M)F^{-2}(C) \mid z) = \Lambda_0(\tau)^2 e^{2\beta' z} E(\gamma^2)E(\epsilon^2). \tag{3.13}
$$

We have $E(M \mid z, c, \epsilon) = \Lambda_0(c)\exp(\boldsymbol{\beta}' z)\epsilon$, and the revised "borrow-strength" estimators of Model A in *Huang and Wang* (2004)

$$
\widehat{\epsilon_k} = \frac{\sum_{i=1}^{I_k} m_{ki}\widehat{F}(c_{ki})^{-1}}{\widehat{\Lambda}_0(\tau)\sum_{i=1}^{I_k} e^{\widehat{\beta}' z_{ki}}}. \tag{3.14}
$$

Thus the estimating equations for the variance components, based on the borrow-strength concept, are given by

$$
\begin{aligned}
&\sum_{k=1}^{K}\sum_{i=1}^{I_k}\left\{(m_{ki}^2 - m_{ki})\widehat{F}^{-2}(c_{ki}) - \exp(2\widehat{\boldsymbol{\beta}}' z_{ki})\widehat{\Lambda}_0(\tau)^2 E(\epsilon^2\gamma^2)\right\} = 0 \\
&\sum_{k=1}^{K}\sum_{i=1}^{I_k}\left\{(m_{ki}^2 - m_{ki})\widehat{F}^{-2}(c_{ki}) - \exp(2\widehat{\boldsymbol{\beta}}' z_{ki})\widehat{\Lambda}_0(\tau)^2 \widehat{\epsilon}_k^2 E(\gamma^2)\right\} = 0.
\end{aligned}
\tag{3.15}
$$

Note that, in the context of clustered subjects, only facilities with more than one event observed will be counted, with accuracy improving substantially when the cluster sizes (I_k) are large. We denote the resulting estimators as $\widehat{E}(\epsilon^2\gamma^2)$ and $\widehat{E}(\gamma^2)$, and $\widehat{E}(\epsilon^2) = \widehat{E}(\epsilon^2\gamma^2)/\widehat{E}(\gamma^2)$. We will show in Section 3.4 that the proposed estimators are consistent for $K \to \infty$.

Alternatively, we propose a "U-statistic" method without predicting the frailty values. Suppose that, within cluster k, with $I_k \geq 2$, the expectation holds that $E(M_{ki}M_{kj}F^{-1}(C_{ki})F^{-1}(C_{kj}) \mid z_{ki}, z_{kj}) = E(\epsilon^2)\Lambda_0(\tau)^2 \exp(\boldsymbol{\beta}'(z_{ki} + z_{kj}))$ for any $i \neq j$. The U-statistic estimating equation for $E(\epsilon^2)$ is given by

$$\sum_{k=1}^{K} I(I_k \geq 2) \sum_{(i \neq j) \in \boldsymbol{C}_{2,I_k}} \left\{ m_{ki}m_{kj}\widehat{F}^{-1}(c_{ki})\widehat{F}^{-1}(c_{kj}) - \exp(\widehat{\boldsymbol{\beta}}'(z_{ki} + z_{kj}))\widehat{\Lambda}_0(\tau)^2 E(\epsilon^2) \right\}, \tag{3.16}$$

where $\boldsymbol{C}_{2,I_k}$ is the set of all possible combinations of two subjects selected from facility k. The second moment estimator of ϵ is represented by $\widetilde{E}(\epsilon^2)$, with a tilde to distinguish it from the borrow-strength one, which instead is denoted as $\widehat{E}(\epsilon^2)$. The second moment estimator of γ is given by the ratio $\widetilde{E}(\gamma^2) = \widehat{E}(\epsilon^2\gamma^2)/\widetilde{E}(\epsilon^2)$. According to Section 3.4, the U-statistic method does not require large facility sizes to establish its consistency, hence it still produces accurate results when many small facilities are present.

3.3.4 Estimation Procedure

Unlike most Expectation–maximization (EM) or BLUP-type estimating approaches that require two iterative steps of parameter estimation and random effect prediction (*Klein*, 1992; *Nielsen et al.*, 1992; *Xue and Brookmeyer*, 1996; *Ma*, 1999; *Ripatti and Palmgren*, 2000; *Ma et al.*, 2003) and Chapter II, the procedure of the proposed estimating method can be described via an assembly line using marginal moment conditions as depicted in the flowchart (Figure 3.1): 1) the baseline shape function $F(t)$ is nonparametrically estimated following (3.6), and this is the only step that all the ordered event and censoring times are utilized; 2) the estimated baseline shape function is input to the estimating equations (3.8) for the regression parameters $\boldsymbol{\beta}$ and the baseline size $\Lambda_0(\tau)$, only event counts and the first moment condition of the nonstationery Poisson process are considered here; 3) the final step is to estimate the variance components using the second moment conditions following (3.10)-(3.16).

We also propose to obtain the standard error for the parameter estimation via bootstrapping. For shared and correlated frailty models, since each subjects are independent of each other, one may conduct a nonparametric random draw of the same number of subjects with replacement and estimate the parameters repeatedly to obtain the estimated standard error. According to our experience, letting the number of replicates be $B = 100$ would provide reliable results. For nested frailty Model C, simple random draws on independent facilities may result in an apparent underestimation of the standard errors for the variance components, thus adding another round of random-draw within facilities tend to provide better coverage probabilities, which was also called "two-step bootstrap" (*Xiao and Abrahamowicz*, 2010).

3.4 Asymptotic Properties

To establish the asymptotic properties, we assume the following regularity conditions:

(A1) $\Lambda_0(\tau) > 0$.

(A2) $P(C \geq \tau, \gamma > 0) > 0$; and $P(C \geq \tau, \gamma > 0, \epsilon > 0) > 0$ for Model C.

(A3) Elements of $\boldsymbol{Z}$ are uniformly bounded.

(A4) $E(\gamma^4) < \infty$; for Model C, $E(\epsilon^4) < \infty$.

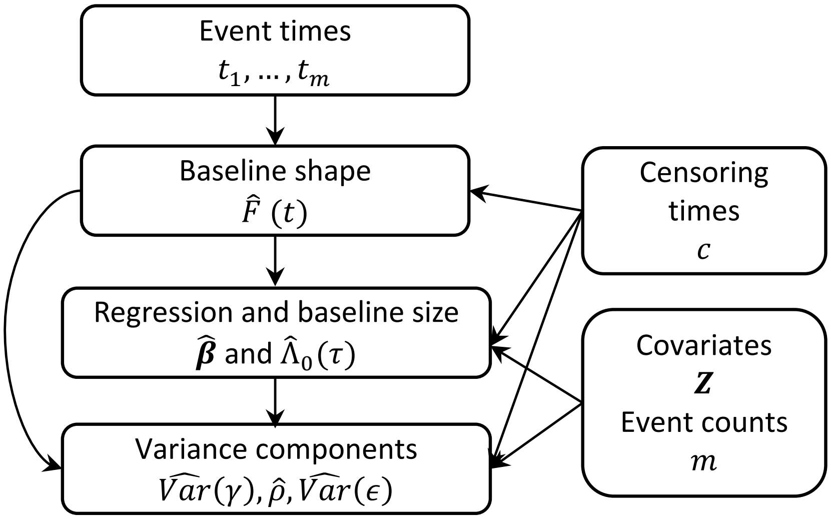

Event times
$t_1, \dots, t_m$
Baseline shape
$\widehat{F}(t)$
Censoring times
c
Regression and baseline size
$\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_0(\tau)$
Covariates
$\boldsymbol{Z}$
Event counts
m
Variance components
$\widehat{Var}(\gamma), \hat{\rho}, \widehat{Var}(\epsilon)$

Figure 3.1: Flow chart of the proposed estimating procedure.

(A5) $G(t) = E[\gamma \exp(\boldsymbol{\beta}'\boldsymbol{Z})I(C \geq t)]$ is a continuous function for $t \in [0, \tau]$; and $G_c(t) = E[\epsilon\gamma \exp(\boldsymbol{\beta}'\boldsymbol{Z})I(C \geq t)]$ for Model C.

For Model A, we established the asymptotic properties of regression parameters and baseline rates following *Wang et al.* (2001). Let $G(t) = E[\gamma \exp(\boldsymbol{\beta}'\boldsymbol{Z})I(C \geq t)]$. Then we use $G(t)$ to define $R(t) = G(t)\Lambda_0(t)$ and $Q(t) = \int_0^t G(u)d\Lambda_0(u)$. The product-limit type estimator (3.6) for the shape function $\widehat{F}(t)$ has an iid representation in $\sqrt{n}(\widehat{F}(t) - F(t)) = (F(t)/\sqrt{n})\sum_{i=1}^n b_i(t) + o_p(1)$ for $t \in [\tau_0, \tau]$, where $\tau_0 > \inf\{t : \Lambda_0(t) > 0\}$; and for $i = 1, \dots, n$,

$$b_i(t) = \sum_{l=1}^{m_i} \left\{ \int_t^\tau \frac{I(t_{il} \leq u \leq c_i)dQ(u)}{R(u)^2} - \frac{I(t < t_{il} \leq \tau)}{R(t_{il})} \right\}.$$

The regression parameters $\boldsymbol{\beta}$ and the baseline size $\Lambda_0(\tau)$ can be estimated through solving the estimating equation in (3.8). Hence the $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = n^{-1/2}\sum_{i=1}^n \boldsymbol{\psi}^{-1}\boldsymbol{e}_i + o_p(1)$, where $\boldsymbol{\psi} = E\left[-\partial \boldsymbol{e}_i/\partial \boldsymbol{\theta}\right]$ and

$$\boldsymbol{e}_i = -\int \frac{\bar{\boldsymbol{x}}_1 m b_i(c)}{F(c)} dV(z, m, c) + w_{1i}\bar{z}_i\left(\frac{m_i}{F(c_i)} - \exp(\boldsymbol{\theta}'\bar{z}_i)\right).$$

For convenience, we use $[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1$ and $[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_{-1}$ to denote the first entry and the remaining entries (beyond the first entry) of $\boldsymbol{\psi}^{-1}\boldsymbol{e}_i$ respectively. Thus we have the following iid representation, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = n^{-1/2}\sum_{i=1}^{n}[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_{-1} + o_p(1)$, implying weak convergence to a mean-0 normal distribution with variance-covariance matrix $E([\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_{-1}[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]'_{-1})$. Moreover, the baseline size also has an iid representation, given by $\sqrt{n}(\widehat{\Lambda}_0(\tau) - \Lambda_0(\tau)) = n^{-1/2}\Lambda_0(\tau)\sum_{i=1}^{n}[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1 + o_p(1)$. Thus, for $\widehat{\Lambda}_0(t) = \widehat{\Lambda}_0(\tau)\widehat{F}(t)$ we have

$$\sqrt{n}(\widehat{\Lambda}_0(t) - \Lambda_0(t)) = (n^{-1/2}F(t)\Lambda_0(\tau)\sum_{i=1}^{n}\left([\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1 + b_i\right) + o_p(1).$$

Following the proofs in Appendix B.5, we obtain the iid representation for the second moment of the frailty, $\sqrt{n}(\widehat{E}(\gamma^2) - E(\gamma^2)) = n^{-1/2}\sum_{i=1}^{n} s_i + o_p(1)$, where s_i are defined as

$$s_i = \left\{\frac{g_i}{\Lambda_0^2(\tau)E\left\{\exp(2\boldsymbol{\beta}'\boldsymbol{Z})\right\}} - 2E(\gamma^2)[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1 - \frac{2E(\gamma^2)h_i}{E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]}\right\},$$

with g_i and h_i both have zero mean.

Theorem III.1. *Model A under the regularity conditions, as $n \to \infty$, has an iid representation of the shape function $\sqrt{n}(\widehat{F}(t) - F(t)) = (F(t)/\sqrt{n})\sum_{i=1}^{n} b_i(t) + op(1)$ for $t \in [\tau_0, \tau]$; the baseline size $\sqrt{n}(\widehat{\Lambda}_0(\tau) - \Lambda_0(\tau))$ converges weakly to a mean-0 normal distribution with variance $\Lambda_0^2(\tau)E([\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1^2)$; the baseline rate $\sqrt{n}(\widehat{\Lambda}_0(t) - \Lambda_0(t))$ converges weakly to a mean-0 normal distribution with variance $F^2(t)\Lambda_0^2(\tau)E[\{[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1 + b_i\}^2]$; the regression parameters $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges weakly towards a mean-0 normal distribution with variance-covariance matrix $E([\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_{-1}[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]'_{-1})$; the second moment estimator $\sqrt{n}(\widehat{E}(\gamma^2) - E(\gamma^2))$ converges weakly to a mean-0 normal distribution with* *variance $E(s_i^2)$, so is the variance estimator.*

Due to the fact that the estimating equations are marginally given for each event type, the asymptotic properties for Model B can be obtained similarly following the proofs for Theorem III.1 in Appendix B. We define $G_1(t)$ and $G_2(t)$ functions for two different event types separately. The estimates for the regression parameters, baseline sizes, and baseline rates have comparable iid representations to the ones for Model A, and the corresponding iid values like b_{ji}, $\boldsymbol{e}_{ji}$, $\boldsymbol{\psi}_j = E[-\partial \boldsymbol{e}_{ji}/\partial \boldsymbol{\theta}_j]$ for event type $j \in \{1, 2\}$ can be found in the Appendix B. The correlation between event types, i.e. ρ, is an additional type of value under estimation in Model B. It has been proved in the Appendix B.5 that $\sqrt{n}(\widehat{\rho} - \rho) = \sum_{i=1}^{n} r_i + o_p(1)$, where

$$r_i = \frac{q_i}{\sqrt{Var(\gamma_1)Var(\gamma_2))}} - \frac{s_{1i}}{2Var(\gamma_1)^{\frac{3}{2}}Var(\gamma_2)^{\frac{1}{2}}} - \frac{s_{2i}}{2Var(\gamma_1)^{\frac{1}{2}}Var(\gamma_2)^{\frac{3}{2}}}.$$

Note that s_{1i} and s_{2i} construct the iid representations for $\sqrt{n}(\widehat{E}(\gamma_1^2) - E(\gamma_1^2)) = n^{-1/2}\sum_{i=1}^{n} s_{1i} + o_p(1)$ and $\sqrt{n}(\widehat{E}(\gamma_2^2) - E(\gamma_2^2)) = 1/\sqrt{n}\sum_{i=1}^{n} s_{2i} + o_p(1)$. The detailed formulation of q_i can be found in Appendix B.5, which together with s_{1i} and s_{2i}, has mean zero.

Theorem III.2. *Model B under the regularity conditions, with $n \to \infty$, has all the iid representation and asymptotic properties given in Theorem III.1 by adding an index $j \in \{1, 2\}$ to distinguish the values for the different event types. Specifically, we have an iid representation for the shape functions of each event type $\sqrt{n}(\widehat{F}_j(t) - F_j(t)) = F_j(t)/\sqrt{n}\sum_{i=1}^{n} b_{ji}(t) + o_p(1)$; the baseline sizes $\sqrt{n}(\widehat{\Lambda}_{0j}(\tau) - \Lambda_{0j}(\tau))$ converge weakly to a mean-0 normal distribution with variance $\Lambda_{0j}^2(\tau)E([\boldsymbol{\psi}_j^{-1}\boldsymbol{e}_{ij}]_1^2)$; the baseline rates $\sqrt{n}(\widehat{\Lambda}_{0j}(t) - \Lambda_{0j}(t))$ converge weakly to a mean-0 normal distribution with variance $F_j^2(t)\Lambda_{0j}^2(\tau)E[\{[\boldsymbol{\psi}_j^{-1}\boldsymbol{e}_{ji}]_1 + b_{ji}\}^2]$; the regression parameters $\sqrt{n}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j)$ converges weakly to a mean-0 normal distribution with variance-covariance matrix $E([\boldsymbol{\psi}_j^{-1}\boldsymbol{e}_{ji}]_{-1}[\boldsymbol{\psi}_j^{-1}\boldsymbol{e}_{ji}]'_{-1})$; the second moment estimator $\sqrt{n}(\widehat{E}(\gamma_j^2) - E(\gamma_j^2))$ converges weakly to a mean-0 nor-*

mal distribution with variance $E(s_{ji}^2)$, so as the variance estimator; the correction coefficient $\sqrt{n}(\widehat{\rho}-\rho)$ also converges weakly towards a mean-0 normal distribution with variance $E(r_i^2)$.

Model C has K independent clusters (facilities), and conditional independence holds within each facility. We treat the number of subjects I_k within facility k as a random variable satisfying $I_k \perp (N_{ki}(t), C_{ki}, \epsilon_k, \gamma_{ki}, \boldsymbol{Z}_{ki})$. Using subscript c to distinguish those estimators from the other two models, we define the new time-dependent function

$$G_c(t) = E\left\{\sum_{i=1}^{I_1} \epsilon_1 \gamma_{1i} \exp(\boldsymbol{\beta}'\boldsymbol{Z}) I(C_{1i} \geq t)\right\} = \nu \int_0^\tau \epsilon\gamma \exp(\boldsymbol{\beta}'\boldsymbol{z}) I(c \geq t) dW(c, \epsilon, \gamma, \boldsymbol{z}),$$

where $\nu = E(I_1) = E(I_k)$. Then we define $R_c(t) = G_c(t)\Lambda_0(t)$ and $Q_c(t) = \int_0^t G_c(u)d\Lambda_0(u)$. In Appendix B.3, we derive the iid representation for the shape function $\sqrt{K}(\widehat{F}(t) - F(t)) = \frac{F(t)}{\sqrt{K}} \sum_{i=1}^K b_{ck}(t) + o_p(1)$ as $K \to \infty$, where we have $b_{ck}(t)$ as

$$b_{ck}(t) = \sum_{i=1}^{I_k} \sum_{l=1}^{m_{ki}} \left\{\int_t^\tau \frac{I(t_{kil} \leq u \leq c_i) dQ_c(u)}{R_c^2(u)} - \frac{I(t < t_{kil} \leq \tau)}{R_c(t_{kil})}\right\}.$$

The iid representation of the parameters is $\sqrt{K}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = 1/\sqrt{K} \sum_{k=1}^K \boldsymbol{\psi}_c^{-1} \boldsymbol{e}_{ck}$, where $\boldsymbol{\psi}_c = E\left[-\frac{\partial \boldsymbol{e}_{ck}}{\partial \boldsymbol{\theta}}\right]$, and $\boldsymbol{e}_{ck}$ is given by

$$\boldsymbol{e}_{ck} = -\nu \int \frac{m\bar{\boldsymbol{z}} b_{ck}(c)}{F(c)} dV(\boldsymbol{z}, m, c) + \sum_{i=1}^{I_k} \bar{\boldsymbol{z}}_{ki} \left(\frac{m_{ki}}{F(c_{ki})} - \exp(\boldsymbol{\theta}'\bar{\boldsymbol{z}}_{ki})\right).$$

The consistency of borrow-strength estimators (3.14) is provided in Appendix B.5. It follows that when the cluster sizes I_k are large, $\widehat{\epsilon}_k$ converges in probability to ϵ_k, and thus the estimation equation (3.15) provides consistent estimators for $E(\gamma^2)$ and $E(\epsilon^2)$ when $K \to \infty$. Both the consistency and asymptotic normality for U-statistic estimators have been established following Appendix B.5, where we have

$1/\sqrt{K}\left(\widehat{E}(\epsilon^2\gamma^2) - E(\epsilon^2\gamma^2)\right) = 1/\sqrt{K}\sum_{k=1}^{K} s_{ck} + o_p(1)$ and

$$s_{ck} = \left\{\frac{g_{ck}}{\nu\Lambda_0^2(\tau)E\left\{\exp(2\boldsymbol{\beta}'\boldsymbol{Z})\right\}} - 2E(\epsilon^2\gamma^2)[\boldsymbol{\psi}_c^{-1}\boldsymbol{e}_{ck}]_1 - \frac{E(\epsilon^2\gamma^2)h_{ck}}{\nu E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]}\right\}.$$

Note that g_{ck} and h_{ck} are analogues of g_i and h_i in Model A. Moreover, the iid representation for the U-statistic estimator of $E(\epsilon^2)$ is $\sqrt{K}(\widetilde{E}(\epsilon^2) - E(\epsilon^2)) = \frac{1}{\sqrt{K}}\sum_{i=1}^{n} w_k + o_p(1)$ where w_k is given by

$$w_k = \frac{u_k}{\omega\Lambda_0^2(\tau)E\left\{\exp(\boldsymbol{\beta}'\boldsymbol{Z})\right\}^2} - 2E(\gamma^2)[\boldsymbol{\psi}_c^{-1}\boldsymbol{e}_{ck}]_1 - \frac{E(\gamma^2)v_k}{\omega E\{\exp(\boldsymbol{\beta}'\boldsymbol{Z})\}^2}.$$

Note that both u_k and v_k have mean zero, and their definitions can be found in Appendix B.5. Following the delta method, we have $\sqrt{K}\left(\widetilde{E}(\gamma^2) - E(\gamma^2)\right) = \sum_{k=1}^{K} y_k + o_p(1)$, where

$$y_k = \frac{1}{E(\epsilon^2)}s_{ck} - \frac{E(\gamma^2)}{E(\epsilon^2)}w_k.$$

Theorem III.3. *As $K \to \infty$ and under the regularity conditions, Model C has an iid representation of the shape function $\sqrt{K}(\widehat{F}(t) - F(t)) = (F(t)/\sqrt{K})\sum_{k=1}^{K} b_{ck}(t) + o_p(1)$ for $t \in [\tau_0, \tau]$; the baseline size $\sqrt{K}(\widehat{\Lambda}_0(\tau) - \Lambda_0(\tau))$ converges weakly to a mean-0 normal distribution with variance $\Lambda_0^2(\tau)E([\boldsymbol{\psi}_c^{-1}\boldsymbol{e}_{ck}]_1^2)$; the baseline rate $\sqrt{K}(\widehat{\Lambda}_0(t) - \Lambda_0(t))$ converges weakly to a mean-0 normal distribution with variance $F^2(t)\Lambda_0^2(\tau)E[\{[\boldsymbol{\psi}_c^{-1}\boldsymbol{e}_k]_1 + b_{ck}\}^2]$; the regression parameters $\sqrt{K}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges weakly towards a mean-0 normal distribution with variance-covariance matrix $E([\boldsymbol{\psi}_c^{-1}\boldsymbol{e}_{ck}]_{-1}[\boldsymbol{\psi}_c^{-1}\boldsymbol{e}_{ck}]'_{-1})$; when I_k are large, the borrow-strength estimators are consistent, i.e. $\widehat{E}(\epsilon^2) \to_p E(\epsilon^2)$ and $\widehat{E}(\gamma^2) \to_p E(\gamma^2)$; the U-statistic estimators do not require large facility sizes and still ensure the asymptotic normality as $K \to \infty$, i.e. $\sqrt{K}(\widetilde{E}(\epsilon^2) - E(\epsilon^2))$ and $\sqrt{K}(\widetilde{E}(\gamma^2) - E(\gamma^2))$ converge weakly towards mean-0 normal distributions with variance $E(w_k^2)$ and $E(y_k^2)$ respectively, so are the variance estimators.*

3.5 Simulation Studies

Simulations under different scenarios for the three frailty models have been carried out to evaluate the finite-sample properties of the proposed estimating approaches. We first checked the estimation of Model A. The shared frailty is set to follow either a gamma or a log-normal distribution with unit mean and variance equal to 0.5 or 1, representing moderate to strong within-subject association. For each simulation, we generated 1000 data sets of $n = 1000$ independent subjects with event rate $\lambda_0 \exp(\boldsymbol{\beta}'\boldsymbol{z})\gamma$. Note that the two entries of the bivariate covariate vector $\boldsymbol{z}$ follow two distributions: a mean-0 normal distribution N(0,0.25) and a Bernoulli distribution with probability 0.5. Their regression parameters are $\boldsymbol{\beta} = [0.5, -0.3]'$. Let the baseline rate be $\lambda_0 = 0.1$, and the stopping time of the study be $\tau = 10$. Hence, the baseline size is $\Lambda_0(\tau) = 1$ and the cumulative baseline shape follows a linear function $F(t) = \Lambda_0(t)/\Lambda(\tau) = 0.1t$. We consider the censoring time to follow a uniform distribution $C \sim U(2, 10)$, with about 60% of subjects censored. Note that the "censored" subjects we define here are those without any observed events. Asymptotic standard errors (ASE) were obtained via nonparametric bootstrap with $B = 100$, and the average computation time (user time) is around 8 seconds. For all simulations, we set the convergence tolerance for parameter estimation to be 10^{-8}. According to Table 3.1, the estimation of regression parameters ($\boldsymbol{\beta}$), baseline size ($\Lambda(\tau)$) and the frailty variance $Var(\gamma)$ is quite quite accurate. Coverage probabilities (CP) are generally close to the nominal 95%, except for the variance component $Var(\gamma)$, for which the ASE tends to slightly underestimate the empirical standard error (ESE), especially for the log-normal frailty settings. Underestimation of the ASEs for cluster-level parameters using nonparametric bootstrap has been reported widely in literature (*Massonnet et al.*, 2006; *Xiao and Abrahamowicz*, 2010). By comparing the cases with $Var(\gamma) = 0.5$ and $Var(\gamma) = 1$, we also notice that increasing the dependence within subjects, decreases the CP.

Table 3.1: Model A: the estimating results for the regression parameters, baseline sizes, and the variance components for the shared frailty.

	True	Model A: Gamma				Model A: Log-normal			
	Value	Bias	ESE	ASE	CP	Bias	ESE	ASE	CP
$\boldsymbol{\beta}[1]$	0.5	0.001	0.103	0.106	0.952	-0.004	0.112	0.107	0.939
$\boldsymbol{\beta}[2]$	-0.3	0.005	0.109	0.107	0.947	0.006	0.109	0.107	0.936
$Var(\gamma)$	0.5	-0.008	0.152	0.150	0.929	-0.003	0.173	0.163	0.910
$\Lambda_0(\tau)$	1	0.007	0.113	0.107	0.949	0.006	0.112	0.105	0.935
$\boldsymbol{\beta}[1]$	0.5	-0.007	0.116	0.117	0.947	-0.006	0.118	0.116	0.943
$\boldsymbol{\beta}[2]$	-0.3	0.003	0.119	0.117	0.935	-0.002	0.117	0.117	0.946
$Var(\gamma)$	1	-0.009	0.215	0.203	0.914	-0.019	0.294	0.247	0.881
$\Lambda_0(\tau)$	1	0.011	0.117	0.110	0.947	0.004	0.107	0.110	0.944

We also tested the estimating approach for the correlated frailty model, Model B. In consistence to the simulations of Model A, we sampled 1000 datasets, each of which had $n = 1000$ subjects. Bivariate covariates were included following the same distributions as those of Model A. Regression parameters for the two event types were set to be $\boldsymbol{\beta}_1 = [0.5, -0.3]'$ and $\boldsymbol{\beta}_2 = [-0.5, 0.3]'$. The two baseline rate are $\lambda_{01} = \lambda_{02} = 0.2$; the stopping time is $\tau = 10$, such that the baseline sizes for two event types are $\Lambda_{01}(\tau) = \Lambda_{02}(\tau) = 2$. We generated bivariate frailties $\boldsymbol{\gamma} = [\gamma_1, \gamma_2]'$ following either a bivariate gamma simulated by R package lcmix (*Dvorkin*, 2012), or a log-normal distribution simulated by R package compositions (*van den Boogaart et al.*, 2018). Their variances are $Var(\gamma_1)$ and $Var(\gamma_2)$, and their correlation coefficient is ρ. We simulated the event times following event rates $\lambda_{0j} \exp(\boldsymbol{\beta}_j' z_i)\gamma_j$, $j = 1, 2$. Censoring times follow $C_j \sim U(2, 10)$. Over 40% of the subjects have at least one of the two events censored. We obtained the ASE through nonparametric bootstraping. The average computation time for each iteration is about 35 seconds. According to Table 3.2, the estimation of the regression parameters, baseline sizes and variance components is quite accurate. Similar to the results of Model A in Table 3.1, the ASE for the variance estimators tends to be underestimated and thus reduces the

CP; and this underestimation seems to be severer for log-normal distributed frailties, and/or when the variance components are more predominant.

Table 3.2: Model B: the estimating results for the regression parameters, baseline sizes, and the variance components for the correlated frailties.

	True	Model B: Gamma				Model B: Log-normal			
	Value	Bias	ESE	ASE	CP	Bias	ESE	ASE	CP
$\boldsymbol{\beta}_1[1]$	0.5	0.001	0.093	0.090	0.938	0.000	0.089	0.089	0.952
$\boldsymbol{\beta}_1[2]$	-0.3	0.003	0.088	0.089	0.951	-0.002	0.091	0.090	0.946
$\boldsymbol{\beta}_2[1]$	-0.5	0.001	0.080	0.081	0.951	-0.004	0.085	0.081	0.936
$\boldsymbol{\beta}_2[2]$	0.3	0.001	0.081	0.081	0.943	-0.001	0.082	0.080	0.943
$Var(\gamma_1)$	0.5	-0.015	0.116	0.110	0.909	-0.014	0.135	0.121	0.886
$Var(\gamma_2)$	0.5	-0.008	0.097	0.090	0.914	-0.006	0.116	0.103	0.909
ρ	0.5	-0.010	0.143	0.151	0.948	-0.028	0.145	0.161	0.940
$\Lambda_{01}(\tau)$	2	0.014	0.188	0.172	0.942	0.023	0.172	0.172	0.944
$\Lambda_{02}(\tau)$	2	0.003	0.149	0.155	0.944	0.007	0.154	0.157	0.957
$\boldsymbol{\beta}_1[1]$	0.5	-0.001	0.102	0.101	0.949	-0.001	0.106	0.101	0.941
$\boldsymbol{\beta}_1[2]$	-0.3	0.002	0.103	0.101	0.945	-0.002	0.099	0.101	0.950
$\boldsymbol{\beta}_2[1]$	-0.5	0.001	0.096	0.094	0.942	0.000	0.094	0.093	0.953
$\boldsymbol{\beta}_2[2]$	0.3	0.002	0.094	0.093	0.952	-0.006	0.092	0.093	0.944
$Var(\gamma_1)$	1	-0.023	0.162	0.156	0.918	-0.020	0.264	0.204	0.854
$Var(\gamma_2)$	1	-0.013	0.143	0.134	0.904	-0.014	0.246	0.184	0.859
ρ	0.5	-0.041	0.094	0.092	0.908	-0.069	0.105	0.104	0.861
$\Lambda_{01}(\tau)$	2	0.014	0.204	0.182	0.941	0.015	0.186	0.181	0.949
$\Lambda_{02}(\tau)$	2	0.009	0.179	0.172	0.948	0.011	0.170	0.171	0.946

To evaluate the estimating approach for Model C on nested data structures, we considered three different pairs of the number of facilities (K) and the facility sizes (I_k): $K = 50$ and $I_k = 40$ (Pair 1), $K = 100$ and $I_k = 40$ (Pair 2), or $K = 100$ and $I_k = 10$ (Pair 3). Bivariate covariates $\boldsymbol{z}_i$ for subject i were similarly drawn to those in the simulations of Model A and Model B, and their regression parameters are $\boldsymbol{\beta} = [0.5, -0.3]'$. Event times follow a Poisson process with rate $\lambda_0 \exp(\boldsymbol{\beta}'\boldsymbol{z}_{ki})\epsilon_k\gamma_{ki}$, where $\lambda_0 = 0.1$. The two levels of frailties, ϵ_k and γ_{ki}, are random draws following either two independent gamma or log-normal distributions, with unit means and xsvariances given in Table 3.3. Censoring times follow a uniform distribution $C \sim U(8, 10)$ with stopping time $\tau = 10$. Over 50% of the subjects are censored without any observed events. The average computation time for the three setting pairs (Pair 1-3) is 25, 120 and 8 seconds respectively. According to the results in Table 3.3, the regression parameters and the baseline sizes are quite accurate as expected. Let Var_b denote the variance estimates using the borrow-strength method, with Var_u denoting those derived through U-statistics. The performance of the variance component estimators varies according to the combination of K and I_k. In general, the borrow-strength method tends to work well when I_k is relatively large. For example, the estimation bias for the variance estimates of the two frailties can reach 50% of the true values when $I_k = 10$, but only 10% when $I_k = 40$. Since we obtain $\widehat{E}(\epsilon^2)$ by taking the ratio of the product estimator $\widehat{E}(\epsilon^2\gamma^2)$ and $\widehat{E}(\gamma^2)$, the slightly underestimated $\widehat{E}(\gamma^2)$ will cause an overestimated $\widehat{E}(\epsilon^2)$. When I_k is small, this bias can decrease the CP considerably for the variance estimates (Pair 3). Conversely, the U-statistic method is more robust to the size of I_k, and its performance improves when K increases (Pair 1 vs. Pair 2). Consistent with Model A and Model B, the estimation performance of the variance components is in general better for the gamma distributed frailties than the log-normal distributed frailties, as the ASEs are more severely underestimated for log-normal frailties. Note that the regression parameters and baseline rate are not affected by which particular methods are used for the variance components. The underestimation of ASE can be compensated to some degree by adding another layer of bootstrap within facilities to obtain higher CP; corresponding results are parenthesized in Table 3.3. We increase the degree of heterogeneity, especially in the facility level, and summarize the estimating results in Table 3.4. In the presence of more predominant heterogeneity, CP decline dramatically, especially among the variance component estimates. The estimating bias of the facility-level variance decreases dramatically for the borrow-strength estimator, but increases for the U-statistic estimator, in terms of the relative magnitude (with respect to $Var(\epsilon)$). The estimating bias of the subject-level variance seems to be quite comparable to those with less hierarchical dependence.

To assess the accuracy of NPMLE for the baseline shape estimation, we tested a linear shape function ($F(t) = 0.1t$) and a cubic baseline shape function ($F(t) = (t-5)^3/250 + 0.5$) following the simulations of Model C with their baselines modified accordingly. Four different seeds have been tried. Their plots with the estimated curves (dashed lines) and 95% confident intervals (dotted lines) are presented in Figures 3.2-3.3, superimposed by the true shape function curves (solid lines). All the plots demonstrate that the NPMLE estimates the shape functions quite accurately.

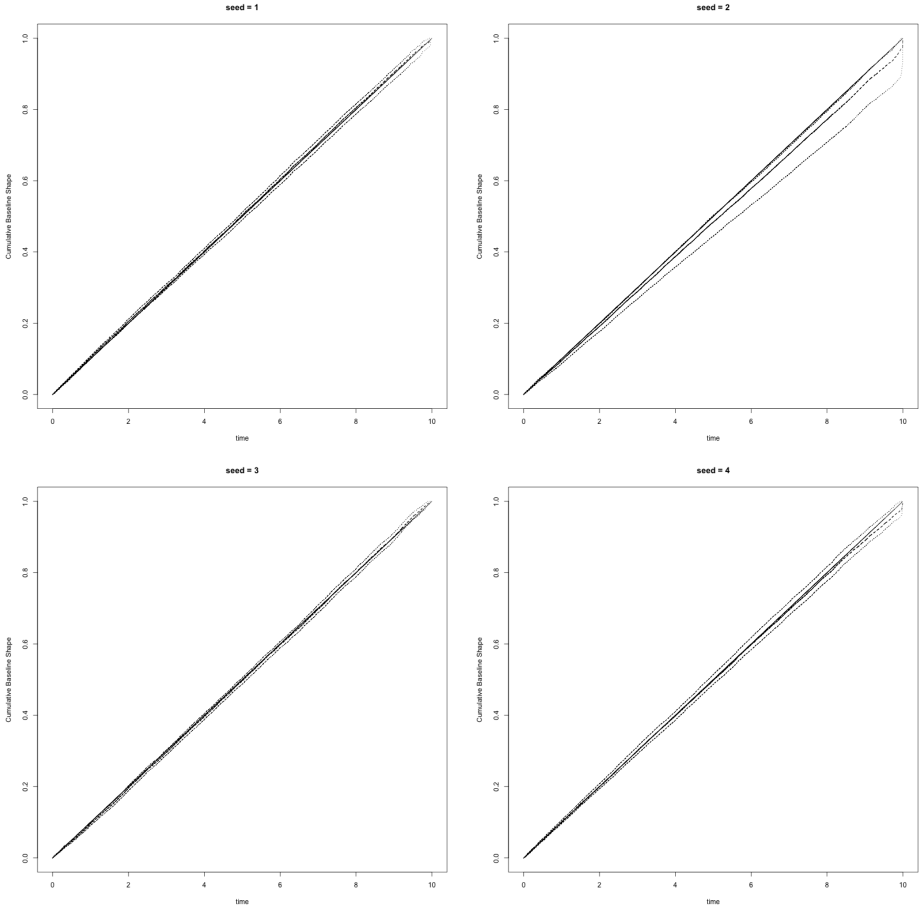


Figure 3.2: Linear baseline shape ($F(t) = 0.1t$) generated with different seeds (1-4). Note that the solid line is the true cumulative baseline shape, the dashed line is the estimated cumulative baseline shape, and two dotted lines are 95% confidence intervals obtained via bootstrapping.

Table 3.3: Model C with moderate heterogeneity: the estimating results for the regression parameters, baseline sizes, and the variance components for the nested frailties.

	True	Model C: Gamma				Model C: Log-normal			
	Value	Bias	ESE	ASE	CP	Bias	ESE	ASE	CP
Pair 1:		K = 50		I_k = 40					
$\boldsymbol{\beta}[1]$	0.5	-0.001	0.066	0.062	0.933	0.000	0.065	0.062	0.931
$\boldsymbol{\beta}[2]$	-0.3	0.001	0.063	0.062	0.940	0.001	0.064	0.061	0.937
$Var_b(\epsilon)$	0.3	0.035	0.076	0.068	0.932	0.032	0.105	0.077	0.889
				(0.081)	(0.972)			(0.089)	(0.947)
$Var_u(\epsilon)$	0.3	-0.007	0.075	0.067	0.864	-0.010	0.104	0.077	0.806
				(0.079)	(0.935)			(0.088)	(0.876)
$Var_b(\gamma)$	0.3	-0.044	0.054	0.051	0.793	-0.047	0.062	0.056	0.776
				(0.069)	(0.920)			(0.076)	(0.904)
$Var_u(\gamma)$	0.3	-0.003	0.058	0.055	0.923	-0.006	0.066	0.059	0.920
				(0.073)	(0.981)			(0.081)	(0.977)
$\Lambda_0(\tau)$	1	0.002	0.094	0.090	0.930	0.002	0.090	0.089	0.940
Pair 2:		K = 100		I_k = 40					
$\boldsymbol{\beta}[1]$	0.5	0.001	0.046	0.045	0.929	-0.002	0.045	0.044	0.939
$\boldsymbol{\beta}[2]$	-0.3	-0.001	0.046	0.044	0.941	0.002	0.045	0.044	0.942
$Var_b(\epsilon)$	0.3	0.037	0.056	0.051	0.918	0.034	0.076	0.060	0.934
				(0.059)	(0.954)			(0.069)	(0.961)
$Var_u(\epsilon)$	0.3	-0.004	0.056	0.050	0.888	-0.007	0.075	0.060	0.834
				(0.058)	(0.927)			(0.068)	(0.894)
$Var_b(\gamma)$	0.3	-0.041	0.039	0.037	0.739	-0.042	0.046	0.041	0.733
				(0.050)	(0.877)			(0.055)	(0.870)
$Var_u(\gamma)$	0.3	0.000	0.042	0.040	0.926	-0.001	0.049	0.044	0.918
				(0.053)	(0.984)			(0.058)	(0.980)
$\Lambda_0(\tau)$	1	0.002	0.065	0.064	0.946	-0.001	0.062	0.063	0.951
Pair 3:		K = 100		I_k = 10					
$\boldsymbol{\beta}[1]$	0.5	0.002	0.092	0.088	0.933	-0.002	0.089	0.089	0.943
$\boldsymbol{\beta}[2]$	-0.3	-0.003	0.090	0.088	0.944	-0.004	0.091	0.088	0.938
$Var_b(\epsilon)$	0.3	0.156	0.079	0.072	0.414	0.155	0.102	0.081	0.551
				(0.107)	(0.862)			(0.118)	(0.905)
$Var_u(\epsilon)$	0.3	-0.008	0.076	0.070	0.890	-0.009	0.099	0.079	0.849
				(0.104)	(0.981)			(0.116)	(0.968)
$Var_b(\gamma)$	0.3	-0.156	0.067	0.062	0.321	-0.155	0.074	0.067	0.349
				(0.073)	(0.429)			(0.076)	(0.419)
$Var_u(\gamma)$	0.3	-0.009	0.087	0.081	0.911	-0.007	0.094	0.087	0.903
				(0.091)	(0.930)			(0.094)	(0.917)
$\Lambda_0(\tau)$	1	0.006	0.081	0.085	0.955	0.006	0.092	0.086	0.934

Table 3.4: Model C with high heterogeneity: the estimating results for the regression parameters, baseline sizes, and the variance components for the nested frailties.

	True	Model C: Gamma				Model C: Log-normal			
	Value	Bias	ESE	ASE	CP	Bias	ESE	ASE	CP
Pair 1:		$K = 50$		$I_k = 40$					
$\boldsymbol{\beta}[1]$	0.5	-0.004	0.084	0.079	0.922	0.003	0.085	0.077	0.928
$\boldsymbol{\beta}[2]$	-0.3	0.002	0.079	0.078	0.942	0.002	0.081	0.076	0.929
$Var_b(\epsilon)$	1	0.008	0.290	0.229	0.874	-0.042	0.549	0.254	0.687
				(0.254)	(0.904)			(0.288)	(0.737)
$Var_u(\epsilon)$	1	-0.052	0.281	0.224	0.820	-0.102	0.518	0.249	0.622
				(0.248)	(0.852)			(0.282)	(0.679)
$Var_b(\gamma)$	0.5	-0.045	0.085	0.070	0.787	-0.048	0.122	0.088	0.752
				(0.094)	(0.900)			(0.112)	(0.851)
$Var_u(\gamma)$	0.5	0.000	0.090	0.075	0.908	-0.002	0.130	0.094	0.850
				(0.100)	(0.970)			(0.119)	(0.944)
$\Lambda_0(\tau)$	1	-0.001	0.154	0.148	0.919	0.008	0.163	0.145	0.904
Pair 2:		$K = 100$		$I_k = 40$					
$\boldsymbol{\beta}[1]$	0.5	-0.001	0.060	0.058	0.923	-0.001	0.059	0.056	0.938
$\boldsymbol{\beta}[2]$	-0.3	-0.002	0.060	0.056	0.930	-0.004	0.059	0.055	0.935
$Var_b(\epsilon)$	1	0.047	0.216	0.180	0.911	0.005	0.411	0.234	0.751
				(0.200)	(0.931)			(0.253)	(0.783)
$Var_u(\epsilon)$	1	-0.011	0.212	0.176	0.866	-0.052	0.406	0.231	0.703
				(0.196)	(0.894)			(0.248)	(0.714)
$Var_b(\gamma)$	0.5	-0.045	0.060	0.053	0.755	-0.056	0.081	0.065	0.694
				(0.070)	(0.900)			(0.090)	(0.831)
$Var_u(\gamma)$	0.5	-0.003	0.064	0.056	0.911	-0.013	0.087	0.069	0.852
				(0.074)	(0.984)			(0.097)	(0.951)
$\Lambda_0(\tau)$	1	0.005	0.110	0.108	0.934	0.005	0.112	0.105	0.935
Pair 3:		$K = 100$		$I_k = 10$					
$\boldsymbol{\beta}[1]$	0.5	-0.001	0.117	0.114	0.937	-0.003	0.119	0.111	0.928
$\boldsymbol{\beta}[2]$	-0.3	-0.004	0.120	0.114	0.924	-0.006	0.119	0.111	0.940
$Var_b(\epsilon)$	1	0.208	0.277	0.222	0.927	0.170	0.525	0.271	0.866
				(0.286)	(0.978)			(0.348)	(0.956)
$Var_u(\epsilon)$	1	-0.022	0.267	0.209	0.841	-0.064	0.473	0.255	0.671
				(0.276)	(0.936)			(0.336)	(0.802)
$Var_b(\gamma)$	0.5	-0.171	0.102	0.084	0.406	-0.169	0.148	0.101	0.440
				(0.096)	(0.481)			(0.112)	(0.518)
$Var_u(\gamma)$	0.5	-0.013	0.135	0.111	0.880	-0.002	0.207	0.134	0.845
				(0.123)	(0.923)			(0.143)	(0.884)
$\Lambda_0(\tau)$	1	0.005	0.133	0.126	0.916	0.004	0.132	0.123	0.916

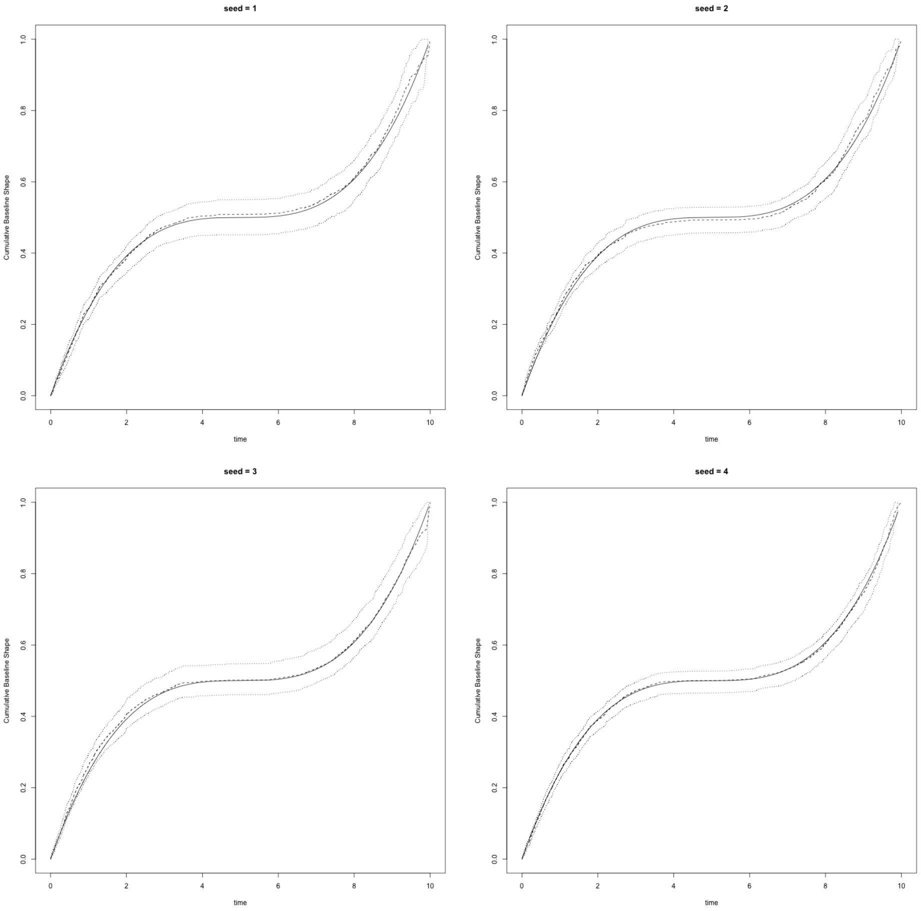


Figure 3.3: Cubic baseline shape ($F(t) = (t - 5)^3/250 + 0.5$) generated with different seeds (1-4). Note that the solid line is the true cumulative baseline shape, the dashed line is the estimated cumulative baseline shape, and two dotted lines are 95% confidence intervals obtained via bootstrapping.

3.6 Application

The Dialysis Outcomes and Practice Patterns Study (DOPPS) is a longitudinal prospective study of hemodialysis patients across different countries. The ultimate goal of the DOPPS is to improve the understanding of dialysis practices that are associated with better event outcomes for end-stage renal disease patients (*Young et al.*, 2000; *Pisoni et al.*, 2004; *Robinson et al.*, 2012). Our objective in this analysis was to determine the important predictors of recurrent hospitalizations. The study sample used for analysis consists of $n = 6,031$ patients from 495 facilities (size ranges from 1 to 74) across 11 different countries: Belgium, Canada, China, Gulf Coast Consortium, Germany, Italy, Japan, Spain, Sweden, the United Kingdom and the United States. We focus on DOPPS Phase-5 adult patients (age ≥ 18) who entered the DOPPS within three months of initiating hemodialysis. With only one subject deleted for its very short censoring ($C_i = 1$), all other subjects of the study cohort have been followed for a maximum of three years by the end of the observation period, 12/31/2015.

The median age among DOPPS patients is 67, with 39.5% being female. We also want to compare the hospital admission rate among dialysis patients by different countries, i.e. Belgium, Canada, China, Gulf Coast Consortium, Germany, Italy, Japan, Spain, Sweden, U.K., Asian-American and African-American are compared to the U.S. Caucasians (reference). Adjustment covariates included age, sex, height, vascular access (arteriovenous (AV) graft, central venous catheter, with AV fistula as the reference), and the comorbid condition indicators like coronary artery disease (CAD), cancer, cerebral vascular disease (CVD), congestive heart failure symptoms (CHF), chronic obstructive pulmonary disease (COPD), peripheral vascular disease (PVD), stroke, diabetes, hypertension, neurological disorder, psychological disorder, and cellulitis.

We analyzed the DOPPS data with Model A treating each subject as a cluster and

Model C with facility-level clustering to be added. Model B is equivalent to fitting Model A twice for two event types and adding a covariance to the final estimation. Note that, these different models provide identical regression parameter estimates, since their variance components are treated as nuisance parameters when using the first moment condition in the first part of the estimation. The variance components, however, are estimated differently according to the design of the model, and the product of the two levels of variance components estimates $\widehat{E}(\epsilon^2\gamma^2)$ in Model C is identical to $\widehat{E}(\gamma^2)$ in Model A. Since the bootstrap in Model A is simply resampling subjects, while in Model C is resampling independent facilities, the ASE of the regression parameters are also different between the two models. Note that an additional round of resampling within subjects among Model C would also compensate (to some extent) the underestimation of the ASE for the variance components.

We considered two different event outcomes, the first is the time to hospital admission and the second is the days hospitalized. Note that the former is only considering the frequency of hospitalizations, while the latter is considering both the frequency and the length of stay. We summarized all the estimation results of Model A and C for both outcomes in Table 3.5 and Table 3.6. Note that the parenthesized ASE and CP in Table 3.6 were computed with additional bootstrapping in facilities. The convergence tolerance for parameter estimation was 10^{-8}. Note that among these analyses, we set $B = 500$ to enjoy a better estimation of the distribution of the parameter estimates via bootstrapping.

Model A assumes independence between subjects. DOPPS patients from Germany ($e^{0.870} = 2.39$), Japan ($e^{0.614} = 1.85$), and U.K. ($e^{0.429} = 1.54$) had significantly higher (p-value< 0.05) covariate-adjusted hospital admission rates than U.S. Caucasians; in contrast, the hospital admission rates for patients in China ($e^{-1.122} = 0.326$), Gulf ($e^{-0.389} = 0.678$), and Spain ($e^{-0.379} = 0.684$) were significantly lower than U.S. Caucasians. Comorbid conditions were generally positively associated with hospi-

Table 3.5: Application of the proposed method to DOPPS data using Model A: estimates are highlighted in bold when $p < 0.05$.

	Admission			Hospitalization days		
	Estimate	$\widehat{\text{SE}}$	P-value	Estimate	$\widehat{\text{SE}}$	P-value
Age (per 5 years)	**-0.046**	0.016	0.004	0.026	0.016	0.104
Height (per 5 cm)	-0.029	0.023	0.208	0.038	0.029	0.185
Female	0.015	0.098	0.880	0.036	0.120	0.765
Vascular access						
Arteriovenous graft	**0.527**	0.209	0.012	-0.110	0.257	0.669
Central venous catheter	**0.797**	0.096	<0.001	**0.430**	0.106	<0.001
Comorbid conditions						
CAD	**0.515**	0.111	<0.001	**0.325**	0.113	0.004
Cancer	**0.204**	0.100	0.040	0.029	0.117	0.804
CVD	**0.284**	0.108	0.008	0.027	0.151	0.859
Stroke	0.165	0.106	0.118	0.110	0.181	0.545
CHF	0.003	0.093	0.977	0.162	0.121	0.181
Diabetes	-0.018	0.078	0.816	-0.096	0.107	0.371
Hypertension	-0.031	0.102	0.759	**-0.471**	0.104	<0.001
COPD	**0.273**	0.100	0.006	0.102	0.160	0.524
Neurological disorder	**0.384**	0.113	0.001	**0.669**	0.169	<0.001
Psychological disorder	**0.349**	0.116	0.003	0.231	0.171	0.178
PVD	0.132	0.099	0.183	**-0.305**	0.117	0.009
Cellulitis	0.213	0.140	0.129	**0.433**	0.169	0.010
Countries						
Belgium	0.111	0.196	0.570	-0.122	0.314	0.697
Canada	0.189	0.159	0.236	0.193	0.179	0.281
China	**-1.122**	0.266	<0.001	**-1.268**	0.192	<0.001
Gulf	**-0.389**	0.175	0.026	**-0.693**	0.194	<0.001
Germany	**0.870**	0.126	<0.001	**0.414**	0.183	0.024
Italy	0.244	0.162	0.133	0.226	0.225	0.315
Japan	**0.614**	0.134	<0.001	**0.970**	0.152	<0.001
Spain	**-0.379**	0.171	0.027	-0.207	0.199	0.297
Sweden	0.204	0.182	0.262	**-0.531**	0.173	0.002
UK	**0.429**	0.160	0.007	0.275	0.221	0.214
USA: Asian	-0.472	0.472	0.318	0.065	0.408	0.874
USA: African-American	-0.111	0.137	0.419	0.058	0.169	0.729
USA: Caucasian	0	-	-	0	-	-
$\Lambda_0(\tau)$	**0.482**	0.067	<0.001	67.384	51.494	0.191
$Var(\gamma^2)$	**1.384**	0.220	<0.001	**2.116**	0.483	<0.001

Table 3.6: Application of the proposed method to DOPPS data using Model C: estimates are highlighted in bold when $p < 0.05$.

	Admission			Hospitalization days		
	Estimate	$\widehat{\text{SE}}$	P-value	Estimate	$\widehat{\text{SE}}$	P-value
Age (per 5 years)	**-0.046**	0.017	0.009	0.026	0.019	0.183
Height (per 5 cm)	-0.029	0.022	0.189	0.038	0.029	0.185
Female	0.015	0.102	0.885	0.036	0.105	0.731
Vascular access						
Arteriovenous graft	**0.527**	0.207	0.011	-0.110	0.258	0.670
Central venous catheter	**0.797**	0.136	<0.001	**0.430**	0.145	0.003
Comorbid conditions						
CAD	**0.515**	0.112	<0.001	**0.325**	0.107	0.002
Cancer	0.204	0.107	0.055	0.029	0.117	0.804
CVD	**0.284**	0.109	0.009	0.027	0.143	0.851
Stroke	0.165	0.102	0.108	0.110	0.190	0.564
CHF	0.003	0.087	0.975	0.162	0.120	0.178
Diabetes	-0.018	0.076	0.812	-0.096	0.132	0.468
Hypertension	-0.031	0.118	0.791	**-0.471**	0.133	<0.001
COPD	**0.273**	0.109	0.012	0.102	0.147	0.490
Neurological disorder	**0.384**	0.114	0.001	**0.669**	0.190	<0.001
Psychological disorder	**0.349**	0.116	0.003	0.231	0.187	0.219
PVD	0.132	0.091	0.148	**-0.305**	0.117	0.009
Cellulitis	0.213	0.157	0.174	**0.433**	0.154	0.005
Countries						
Belgium	0.111	0.264	0.673	-0.122	0.425	0.774
Canada	0.189	0.194	0.331	0.193	0.277	0.486
China	**-1.122**	0.365	0.002	**-1.268**	0.212	<0.001
Gulf	-0.389	0.226	0.085	**-0.693**	0.233	0.003
Germany	**0.870**	0.178	<0.001	**0.414**	0.208	0.046
Italy	0.244	0.272	0.370	0.226	0.239	0.344
Japan	**0.614**	0.241	0.011	**0.970**	0.254	<0.001
Spain	-0.379	0.278	0.173	-0.207	0.273	0.448
Sweden	0.204	0.208	0.326	**-0.531**	0.219	0.015
UK	0.429	0.260	0.099	0.275	0.463	0.553
USA: Asian	-0.472	0.608	0.437	0.065	0.401	0.872
USA: African-American	-0.111	0.185	0.550	0.058	0.152	0.700
USA: Caucasian	0	-	-	0	-	-
$\Lambda_0(\tau)$	**0.482**	0.093	<0.001	**67.384**	17.516	<0.001
$Var_b(\epsilon^2)$	**0.875**	0.142	<0.001	1.247	0.426	0.003
		(0.232)	(< 0.001)		(0.692)	(0.071)
$Var_u(\epsilon^2)$	**0.482**	0.117	<0.001	**1.004**	0.238	<0.001
		(0.180)	(0.007)		(0.393)	(0.011)
$Var_b(\gamma^2)$	0.271	0.125	0.030	0.387	0.383	0.313
		(0.153)	(0.076)		(0.389)	(0.320)
$Var_u(\gamma^2)$	**0.609**	0.183	0.001	0.555	0.398	0.163
		(0.195)	(0.002)		(0.405)	(0.171)

tal admissions. Common significant positive predictors include CAD, cancer, CVD, COPD, neurological disorder, psychological disorder were associated. The impact of age was found to be negatively associated with hospital admissions, though the difference for every 5-year increment was small (-4.5%). In comparison to AV fistula (the most commonly adopted vascular access approach) AV graft and central venous catheter increased the hospital admission rate by 1.69 and 2.22 times, respectively. The baseline size is 0.482, implying an average event counts for $\boldsymbol{Z} = \mathbf{0}$ (without any comorbidity, male, average age and height, AV fistula, U.S. Caucasians) at the end of the study $\tau = 1487$ (days). The baseline shape was plotted in Figure 3.4. The estimated variance of the shared frailty within each subject was 1.384 (p-value< 0.001).

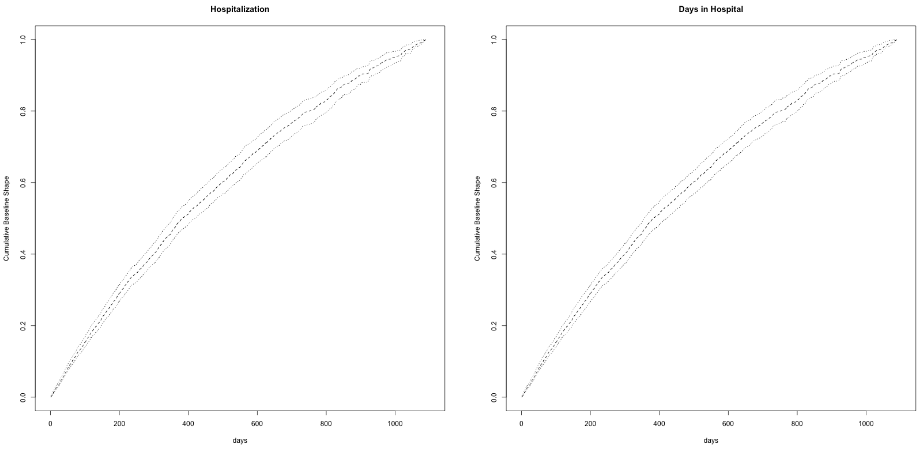


Figure 3.4: Hospitalizations and days in hospital: estimated baseline shape functions $\widehat{F}(t)$ and their 95% bootstrap confidence intervals.

The significant predictors of hospitalization admission rates using Model C were quite similar to those using Model A. Note that in Model C, though the mean estimates in both models were identical, the ASE in Model C were in general larger than those using Model A. There we observed fewer significant predictors: Cancer, Gulf, Spain were no longer significant in Model C. Moreover, the variance component was decomposed into the facility level and the subject level, which were both significant. We presented both the borrow-strength and U-statistic methods. Note that over 80% were small sized facilities ($I_k < 20$), thus we would expected that the variance estimates using the U-statistic method were more accurate. Inference results based on bootstrapping with correction (in parenthesis) and without both indicated the significance of the two levels of heterogeneity based on the U-statistic method.

Hospitalization days were treated as another outcome for analysis. Because we coded each day in hospital as an event, this outcome would count for both the event frequency and the length of stay. Possibly due to its composite information, the significant predictors were quite different from those of the hospitalization admission rates. Assumed to be independent subjects in Model A, DOPPS patients from Germany ($e^{0.414} = 1.51$) and Japan ($e^{0.970} = 2.64$) had significantly higher hospitalization days than U.S. Caucasians; in contrast, patients in China ($e^{-1.268} = 0.28$), Gulf ($e^{-0.693} = 0.50$), and Sweden ($e^{-0.531} = 0.59$) were significantly lower than U.S. Caucasians.

Comorbid conditions CAD, neurological disorder and cellulitis were positively associated with the hospitalization days, while hypertension and PVD were negatively associated. Moreover, in comparison to AV fistula, only central venous catheter significantly increased the hospitalization days by 1.54 times. Due to the very large ASE, the baseline size is not significant in Model A. The estimated variance of the shared frailty within each subject was 2.116 (p-value< 0.001). Model C shared the identical significant predictors. Possibly due to the more stable bootstrapping on facilities than subjects, however, our ASE for the baseline size in Model C was much smaller. In addition, according to Table 3.6, clustering effect from facilities was more significant than that within a subject in presence of all the covariates, which further implies that the facility clustering effect cannot be neglected in DOPPS hospitalization days. The estimated baseline shape using Model C can also be found in Figure 3.4.

3.7 Discussion

In this report, we propose three different frailty models, including a shared frailty model, a correlated frailty model and a nested frailty model to accommodate a variety of clustered event data. A nonstationary Poisson process is assumed and there is no distributional restriction put on the random effects. Though not necessarily required, dependent censoring can be circumvented neatly under the assumption of conditional independence between censoring and event processes of interest given the observed covariates and random effects.

The general estimating framework summarized in Figure 3.1 establishes fast and accurate estimation on regression effects, baseline rate shape function and sizes, and the variance components, accompanied by proved asymptotic properties. In comparison with the regular frailty models with random effects of known distributions, the proposed approach has a slightly lower estimating efficiency (*Ye et al.*, 2007). However, the marginal and sequential estimating procedure endows our estimation with a much faster computational speed via avoiding intensive iterations between the estimation steps for the regression parameters and variance components. Their standard errors are conveniently obtained through bootstrapping. Currently, all the codes were written in `R` (*R Core Team*, 2018a), implying a considerable potential for improvement in computational speed by transferring (part of) the codes to `C++` (e.g. `Rcpp`). Therefore, the proposed models and their estimating framework can adapt well to relatively large data sets with a lot subjects or clusters present.

Unlike the other estimating equation methods (*Lin et al.*, 2000; *Xue*, 1998; *Kalbfleisch et al.*, 2013), as discussed by *Wang et al.* (2001) for shared frailty models, the proposed framework does not accommodate time-varying covariates. Time-varying covariates are interesting but will destroy the invariant feature of the shape function, and thus its nonparametric estimation. Alternatively, for those who are still interested in estimating the variance components, one may assume parametrically distributed baselines

(*Xue*, 1998) or frailties (*Kalbfleisch et al.*, 2013) to incorporate time-varying covariates, which need to be external to the event process (*Kalbfleisch and Prentice*, 2002)

CHAPTER IV

Multistate Rate Models to Assess the Impact of Exposure to Lead on Children Behaviors Using Accelerometer Data

4.1 Introduction

Exposures to environmental toxicants and their detrimental effects on childhood development have been widely studied in a vast epidemiological literature in the past several decades. For example, prenatal and/or postnatal exposure to lead and other environmental toxic agents have been found significantly associated with attention deficit hyperactivity disorder (ADHD). ADHD is the most common neurodevelopmental disorder in children and adolescents, with an estimated prevalence around 5% (*Polanczyk et al.*, 2007). There have been a large number of studies to assess the association between lead exposure and ADHD in different countries around the world; (*Bellinger et al.* (1987); *Needleman et al.* (1990); *Cummins and Goldman* (1992); *Braun et al.* (2006); *Wang et al.* (2008); *Boucher et al.* (2012); *Hong et al.* (2015)). As was pointed out by *Thapar et al.* (2013), although there exists a significant association between lead exposure and ADHD, such effect is intertwined with other factors such as genetic and familial risks as well as unobserved confounders, implying a lack of evidence of causality for this association.

The Early Life Exposure in Mexico to ENvironmental Toxicants (ELEMENT) Project has established a longitudinal cohort since the mid-1990s with an international multi-institutional partnership. One main objective of this project is twofold: (i) to understand whether exposures to prenatal and postnatal lead would elevate risks on the neurodevelopment among children, and (ii) whether a supplementation of maternal calcium would help suppress the detrimental effects from the lead exposures. In addition, many biomarkers, including genetic and epigenetic polymorphisms for metabolism, and other environmental toxicants like phthalates, metals, pesticides and fluoride, have been also included in the ELEMENT data (*Perng et al.*, 2019). For example, recently researchers utilized the ELEMENT data to examine the effects of prenatal exposure to fluoride on ADHD symptoms (*Bashash et al.*, 2017, 2018). Both self-reported ADHD diagnoses and/or measured Conners' Continuous Performance Test (CPT) (*Conners et al.*, 2000) are recorded as part of the ELEMENT cohorts. Note that ADHD is often diagnosed by two domains of traits: 1) difficulties to sustain attention and 2) fidgeting and persistent pattern of motor activity (*Thapar et al.*, 2012; *Thapar and Cooper*, 2016).

In this chapter, we focus on physical activity among children and adolescents in the ELEMENT cohorts, which is measured by wrist-worn ActiGraph GT3X+, a wearable accelerometer device that provides objective measurements of movements and high-resolution activity profiles. With a high resolution of sampling, physical movement signals are recorded and processed to be activity counts for analysis. Activity counts are useful to the evaluate the intensity of physical movements in a time window or epoch (e.g., one minute). For GT3X+ device, tri-axial measurements of body movements at the default frequency of 30Hz (30 measurements per second) are captured in three dimensions, which are commonly aggregated by vector magnitudes of $\mathrm{VM} = \sqrt{\mathrm{axis}_1^2 + \mathrm{axis}_2^2 + \mathrm{axis}_3^2}$ for each epoch. Furthermore, other interpretable activity metrics, e.g. energy expenditure (MET) and activity index (AI), have been proposed to enhance signal detection, enable fair comparisons, and reduce the dimension of raw data for more efficient data processing procedures (*Welk*, 2005; *Colley et al.*, 2011; *Harrington et al.*, 2011; *Van Hees et al.*, 2013; *Hildebrand et al.*, 2014; *Bai et al.*, 2014, 2016). Several studies have been conducted to map physical activity counts to different activity intensity categories using accelerometer measurements (*Freedson et al.*, 1998; *Puyau et al.*, 2002; *Freedson et al.*, 2005; *Troiano et al.*, 2008; *Sasaki et al.*, 2011; *Evenson et al.*, 2015; *Chandler et al.*, 2016). In addition to methods to derive multiple measures and metrics, *Zhang et al.* (2019) present an overview of various models to analyze the accelerometer measurements, including longitudinal data analysis approaches like mixed-effects models (*Fitzmaurice et al.*, 2012; *Li et al.*, 2017) and functional data analysis methods (*Li et al.*, 2014; *Goldsmith et al.*, 2015, 2016). Moreover, to address excessive zero measurements in accelerometer data, *Bai et al.* (2018) proposed a two-stage model with a model fitting step and a smoothing step, while *Li et al.* (2018) fitted three categories of activity jointly under ordinal transitions.

Almost all the aforementioned statistical models that have been employed for the analysis of accelerometer data are grounded upon certain modeling assumptions. Some existing models require heavy computational burden due to the use of multiple steps to obtain parameter estimations in order to handle very noisy observations. In this chapter we first convert the physical activity counts into ordinal categorical variables (*Chandler et al.*, 2016) and then focus on analyzing the time-dependent frequencies of activity-state transitions. The categorical variables define four physical activity statuses ranging from being sedentary, slightly active, moderately active to vigorously active. We propose to use a family of multistate rate models in that the moments-based modeling of the time-to-event rates (*Lin et al.*, 2000) is reformulated by crude hazards under the framework of competing risks. Similar to the classical relative risk models (*Cox*, 1972), a multistate rate model consists of a non-parametric baseline rate and a proportional multiplier of covariate effects. This formulation allows us to examine the association between lead exposure and physical activity profiles among the children in ELEMENT cohorts. Moreover, the baseline rates can be stratified by activity transition types and/or other categorical variable and are assumed to renew every day, implying a semi-Markov renewal property that the average daily changing patterns of different transition rates occur repeatedly. This renewal property imposes a difficulty to obtain proper filtration to form the partial likelihoods since each subject can contribute multiple times to the at-risk set of a stratum (e.g., a transition type). The proposed multistate rate models, however, avoid using partial likelihoods and are fully founded upon moment conditions. The proposed model is also quite flexible so that we can model proportional effects of covariates to be shared among some strata. In addition, fitting the proposed multistate rate models is computationally easy due to the available R package survival (*Therneau*, 2015). With the software, the regression parameters can be estimated consistently, as well as the the inference for correlated transitions corrected by using robust sandwich variance estimators.

The rest of the chapter is organized as follows. The accelerometer data are summarized in Section 4.2. The proposed multistate models are discussed in Section 4.3. The analysis results using the proposed multistate rate models are listed in Section 4.4. Some concluding remarks are included in Section 4.5.

4.2 Accelerometer Data

The ELEMENT project consists of three birth cohorts (cohort 1 in 1994-1997, cohort 2 in 1997-2000, cohort 3 in 2001-2005) of over 2000 children from Mexico City whose mothers were initially recruited from clinics. The accelerometer data sets used in this paper were collected as part of an ELEMENT follow-up study conducted between 2015 and 2018 from 554 children (*Perng et al.*, 2019). Among the total 554 children, 519 of them have complete 7-day (10080-minute) accelerometer measurements without interruptions. We choose an epoch of one minute to aggregate activity counts, from which VMs are calculated. MVs are the primary evaluation that we use to analyze children physical behaviors.

We proposed to label children physical activity states by classifying their one-minute VMs according to *Chandler et al.* (2016) rules, where the original 5-second cutoff values are transformed to one-minute cutoff values. The resulting cutoff values of activity counts are 3660, 9804, and 23628 per minute, which divide the individual physical activity into four categories: sedentary, slightly, moderately, to vigorously active. They are recorded in ordinal variables coded as 0,1,2 and 3, respectively in our data analyses below. To visualize the daily changing patterns, Figure 4.1 displays daily time series VMs and the transformed activity states from two subjects (#252 and #320) over a period of 7 days. The x-axes denote clock-time, and 7 daily time series are stacked (semitransparent colors), that are superimposed by the daily mean (black) and median curves (red) curves. Note that both curves of median and mean physical activity categories are derived from the median and mean VMs, respectively. The gray lines in the left panels of Figure 4.1 denote the Chandler's cutoff values for the VM classification. Figure 4.1 shows that these two children experience daily changing patterns over 7 days, suggesting a certain daily renewal mechanism. More details of the renewal property and the proposed event rate models will be discussed in Section 4.3.

The frequency distributions of individual states and their average daily transition counts among the 554 subjects are summarized in Figure 4.2, where notation $j \to k$ denotes a transition from state j to state k, $j, k = 0, 1, 2, 3$ and $j \neq k$. We noticed that children predominantly stayed in sedentary status with very low physical activities, such as sleeping or being seated. The highest frequency of transitions occurred between sedentary state and slightly active state ($0 \leftrightarrow 1$), and the second most frequent one occurred between slightly active state and moderately active state ($1 \leftrightarrow 2$), where $j \leftrightarrow k$ denotes both directions of transitions between two states j and k, namely $j \rightarrow k$ and $k \rightarrow j$, where $j, k = 0, 1, 2, 3$ and $j \neq k$. The transitions from and to vigorously active state ($j \leftrightarrow 3$ for $j \in \{0, 1, 2\}$) were generally rare. For the ease of visualization, we colored the two types of transitions, increased-activity transitions ($j \rightarrow k$, $k > j$) and decreased-activity transitions ($j \rightarrow k$, $k < j$) differently in salmon versus cyan in Figure 4.2. Since the occurrence of vigorous active state was relatively rare, it would be of interest to look at the conditional proportions which motivate the multistate models conditional on risk subgroups. The conditional proportions of individual states (y-axis) given their previous states (x-axis) are shown in Figure 4.3. It is evident that the ELEMENT children were likely to stay in low-level activities (sedentary and slightly active), while tended to move to a lower activity state if they were in a high-level activity state (moderately active or vigorously active).

In order to study potential impact of lead exposure on the physical activity behavior, we considered the blood lead concentration (Pb in ug/dL) that was measured at their previous follow-up visits, and included adjusting covariates like age, gender and children's Z-score body mass index for age (Zbfa). Due to the missingness (see Figure 4.4), we end up with a dataset of 333 children (170 boys and 163 girls) with no missing values in all variables in the analysis in Section 4.4. The state and transition plots for these non-missing subjects in Appendix C, in correspondence to Figures 4.2-4.3, display little changes after deleting the missing observations, which implies that the missing data mechanism is at random. The summary statistics of continuous explanatory variables are given in Table 4.1. Age and Pb exposure are both centered prior to their use in model fittings, and the boys are always the reference group in the analysis.

There exist some differences between boys and girls in terms of their distributions in the explanatory variables (Table 4.1) and activity state transitions (Figure 4.5).

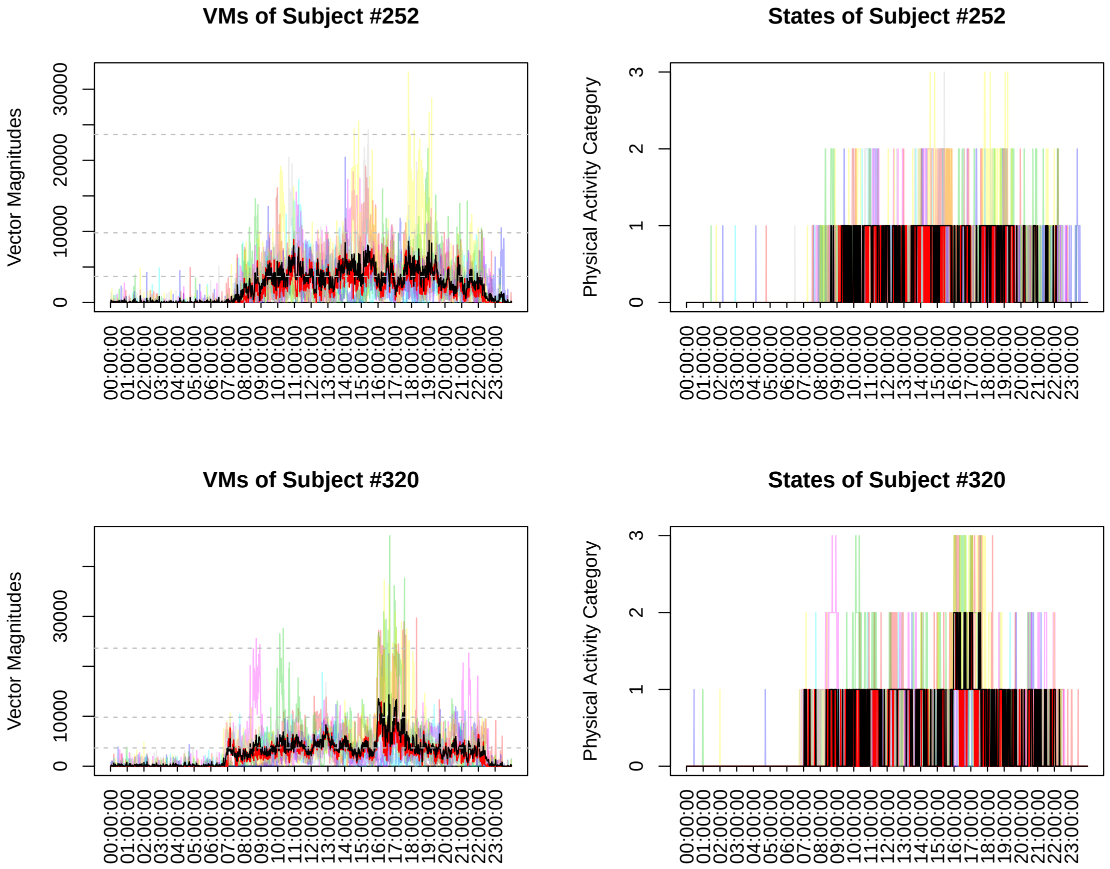


Figure 4.1: Daily time series of vector magnitudes (left) and transformed states (right) over a week were stacked, respectively. Different semi-transparent colors indicate different days, red and black colors represent median and mean curves. Note that the median and mean physical activity categories were derived from the median and mean VMs. For the left panels, gray dashed lines denote the three cut-off values. The activity states under investigation include sedentary (0), slightly active (1), moderately active (2), and vigorously active (3).

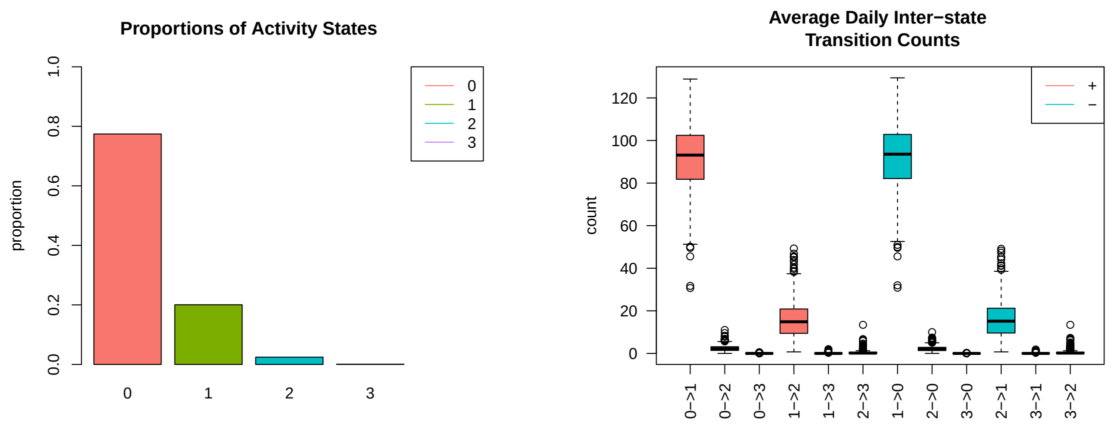


Figure 4.2: Marginal proportions of individual activity states (left) and the distribution of average daily transitions counts for each subject (right). Note that in the right panel, the salmon boxes denote transitions with increased activities (labeled by "+"), while the cyan boxes denote transitions with decreased activities (labeled by "-"). The activity states include sedentary (0), slightly active (1), moderately active (2), and vigorously active (3) statuses.

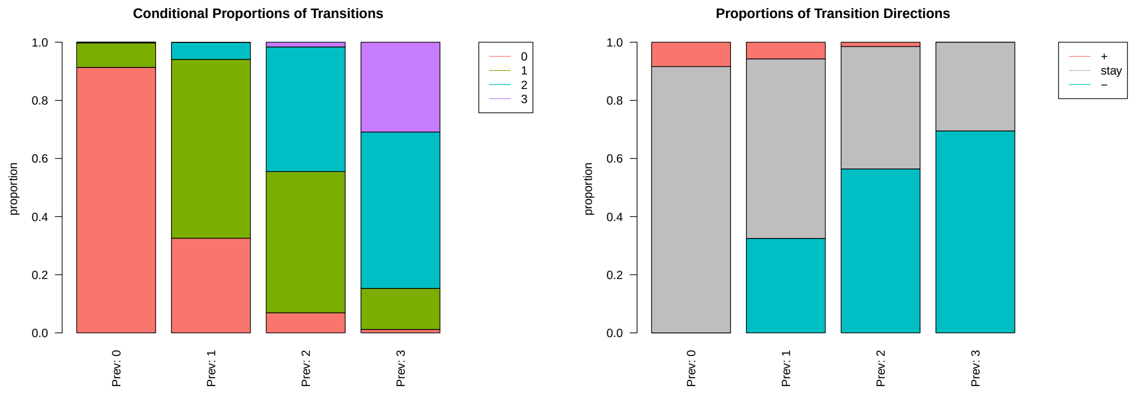


Figure 4.3: The left penal shows the conditional proportions of transitions while the right panel describes the conditional proportions of the transition directions. Their x-axes correspond to previous states and the y-axes represent the transition proportions. The activity states include sedentary (0), slightly active (1), moderately active (2), and vigorously active (3). The transition directions include increased ("+"), decreased("-"), and remained ("stay").

As given in Table 4.1, the girls have a narrower range of Pb ([1 ug/dL, 13 ug/dL]) than that of the boys ([1 ug/dL, 41 ug/dL]), and their standard deviations are 4 and 2.3 respectively. Moreover, Zbfa is slightly higher among girls than boys with their means (and ranges) to be 0.66 ([-2.13, 3.45]) and 0.42 ([-2.90, 3.21]). Age is similarly distributed for both gender groups. Figures 4.5 shows gender-stratified transition proportions (the top panel) and their average event counts (the bottom panel). It is interesting to notice that the boys tended to have higher proportions and frequencies for transitions from or to vigorous activity state (i.e., $j \leftrightarrow 3$ for $j \in \{0, 1, 2\}$), as well as for the transitions between sedentary state and moderately active state ($0 \leftrightarrow 2$). In contrast, transitions between low activity states tended less frequently to occur among the boys than the girls. This implies that the girls tended to more often experience low-level activities than the boys. Note that the frequencies in the bottom panel are in a logarithmic scale thus the true differences between values on the plot should be more substantial than what has been shown in the plot. The summary counts of each transition stratified by gender to produce Figure 4.5 are listed in Appendix Table C.1.

Table 4.1: The summary statistics of explanatory variables for the 333 children.

				Boys (n = 170)		Girls (n = 163)	
Variable	Mean	SD	Range	Mean	SD	Mean	SD
Age	13.59	1.77	[10.77, 17.46]	13.73	1.74	13.45	1.80
Pb (ug/dL)	3.23	3.28	[1, 41]	3.40	4.0	3.05	2.3
Zbfa	0.53	1.24	[-2.90, 3.45]	0.42	1.32	0.66	1.15

4.3 Multistate Rate Models

When the modeling of multiple event types is of interest in data analysis, a multistate model is designed to fit observed transition times. Given a transition from state j to k (where $j \neq k$), or $j \to k$, let $N^*_{i(j\to k)}(t)$ denote the cumulative counts of transitions $j \to k$ for subject i by time t if the subject is always at risk for this transition.

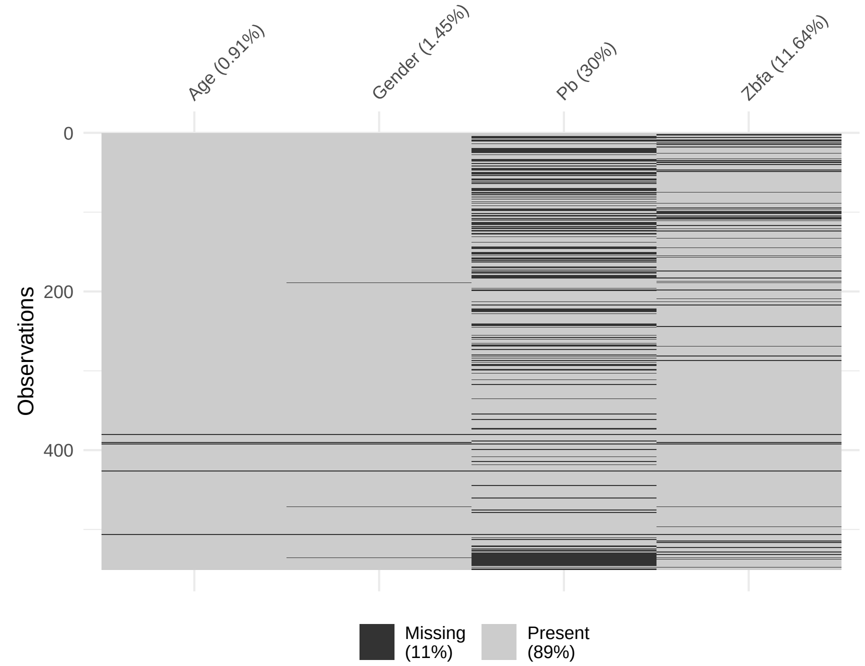


Figure 4.4: Distribution of missing entries for each explanatory variable of interest. Each black line denotes a missing entry for a subject (row). The percentages of missingness are also provided.

In this study of physical activity, there are four states including sedentary (0), slightly active (1), moderately active (2), and vigorously active (3) states, respectively. For $j < k$ transition, $j \to k$ denotes a movement from an activity state to a higher activity state, or *vice versa*. It is worth pointing out that a subject cannot be at risk for every transition at a given time, which is obviously conditional on his/her present state. This is because a subject is at risk for a specific transition $j \to k$ only when his or her instantaneously preceding state is j and before the ending time T. Let a time-varying categorical variable $S_i(t) \in \{0, 1, 2, 3\}$ denote the activity states of subject i at time t. It is natural to introduce an at-risk process $Y_{i(j \to k)}(t) = I(t \in [0, T], S_i(t^-) = j)$ to indicate whether subject i is at risk for transition from state j to k at time t. Note t^- denotes the time instantaneously before t. Let $dN^*_{\cdot}(t) = N^*_{\cdot}(t) - N^*_{\cdot}(t^-)$ be an instant increment of a counting process $N^*_{\cdot}(t)$ at time t. Since the accelerometer data have been aggregated in minutes and followed up for 7 days, we have all the time points

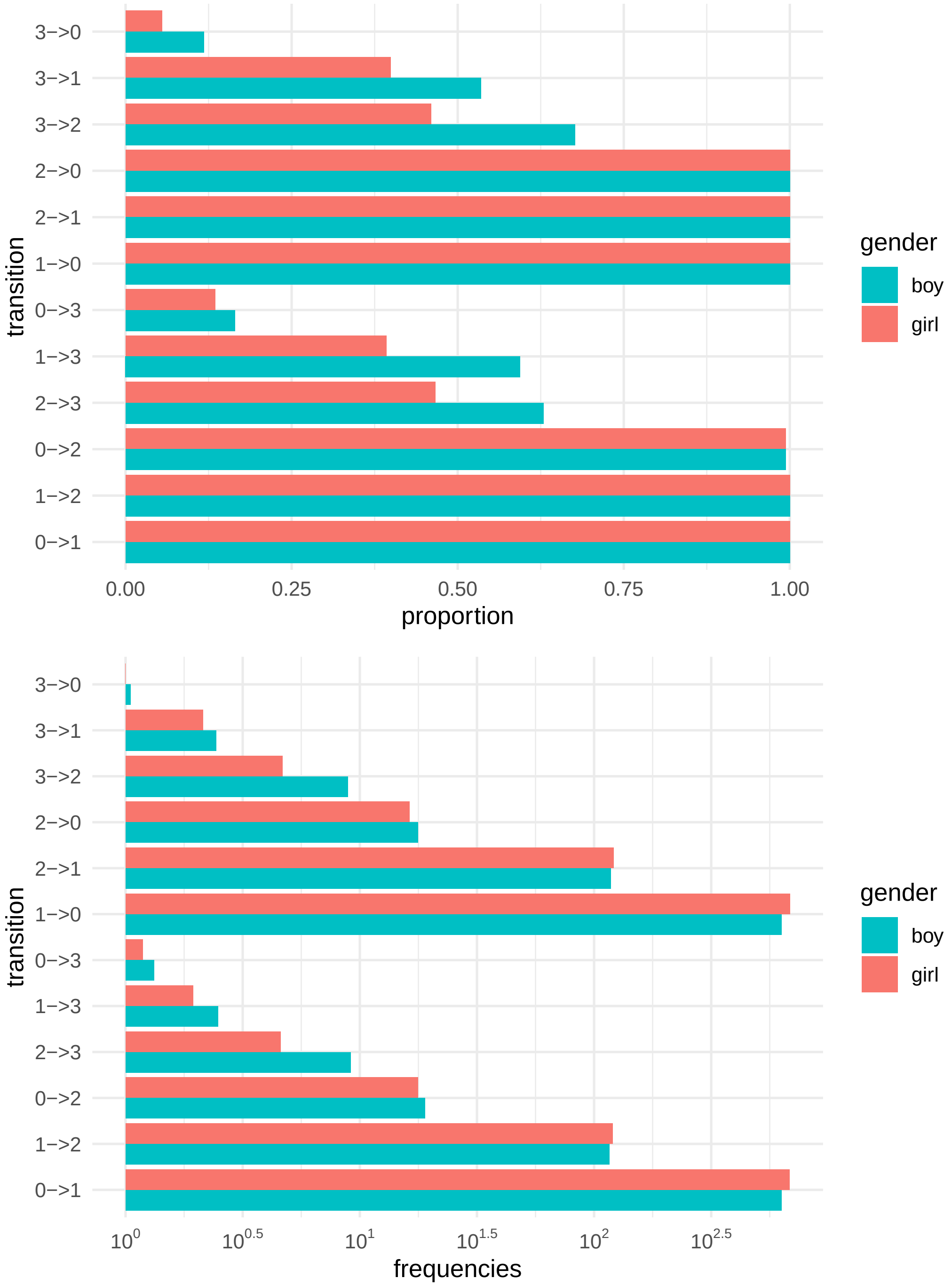


Figure 4.5: Gender-stratified proportions of subjects with at least one corresponding transitions (top) and the corresponding average case counts (bottom). Note that the average counts in the bottom panel are plotted in a logarithmic scale. The activity states under investigation include sedentary (0), slightly active (1), moderately active (2), and vigorously active (3).

$t = 0, 1, \ldots, 10080$ (minutes) and $t^{-} = \max(t-1, 0)$. With the at-risk process taken into consideration, the counting process of observed $j \rightarrow k$ transitions on subject i is $N_{i(j\rightarrow k)}(t) = \int_0^t Y_{i(j\rightarrow k)}(s) dN^{*}_{i(j\rightarrow k)}(s)$ or equivalently $dN_{i(j\rightarrow k)}(t) = Y_{i(j\rightarrow k)}(t) dN^{*}_{i(j\rightarrow k)}(t)$, with $dN.(t) = N.(t) - N.(t^{-})$.

As an extension of the relative risk models (*Cox*, 1972) for univariate event times, multistate models have been studied extensively in a vast literature of survival analysis, e.g. *Andersen and Gill* (1982), *Andersen et al.* (1992), *Kalbfleisch and Prentice* (2002), *Therneau and Grambsch* (2000). Similar to the classical Cox model, multistate models are formulated with a nonparametric baseline hazard or intensity function, and a multiplicative relative risk term that accounts for the contributions from explanatory variables of interest. The property of proportional hazard is satisfied in the absence of time-varying covariate effects $\beta(t)$. The multistate or Cox model can conveniently incorporate time-dependent external covariates $\boldsymbol{Z}(t)$ (*Kalbfleisch and Prentice*, 2002). The formulation of partial likelihood (*Cox*, 1975) requires the event history or filtration, defined by a $\sigma-$field $\mathcal{F}_{t^{-}} = \sigma\{N_{i(j\rightarrow k)}(u^{-}), Y_{i(j\rightarrow k)}(u), \boldsymbol{Z}_i(u) : 0 \leq u \leq t, i \in 1, \ldots, n, j \neq k$ and $j, k \in \{0, 1, 2, 3\}\}$ for a total of n subjects and time $t > 0$ (*Kalbfleisch and Prentice*, 2002; *Fleming and Harrington*, 2011). The sequence of filtrations is used to establish the conditional independence between events, both within and between event types for each subject, which is the essence in the construction of the partial likelihood. A regular multistate model for a given type of inter-state transition is given by

$$E\left(dN_{i(j\rightarrow k)}(t) | \mathcal{F}_{t^{-}}\right) = Y_{i(j\rightarrow k)}(t) \lambda_{j\rightarrow k}\left(t, \boldsymbol{Z}_i(t)\right), \;\; j \neq k, \;\; j, k \in \{0, 1, 2, 3\} \qquad (4.1)$$

where $E\left(dN^{*}_{i(j\rightarrow k)}(t) | \mathcal{F}_{t^{-}}\right) = \lambda_{j\rightarrow k}\left(t, \boldsymbol{Z}_i(t)\right)$ denotes the population intensity of transition from state j to k for subject i at time t. Moreover, the population intensity may be specified by a relative risk model:

$$\lambda_{j\to k}\,(t, \boldsymbol{Z}_i(t)) = \lambda_{0(j\to k)}(t)\exp(\boldsymbol{\beta}'_{j\to k}\boldsymbol{Z}_i(t)), \tag{4.2}$$

where $\boldsymbol{\beta}_{j\to k}$ is a p-element vector of effects from covariates $\boldsymbol{Z}(t)$. For example, if the elements of $\boldsymbol{Z}(t)$ include gender, age, lead exposure and Zdfa, then complexity of $\boldsymbol{\beta}_{j\to k}$ is 4. When the covariates are all time-invariant, namely $\boldsymbol{Z}_i(t) \equiv \boldsymbol{Z}_i$, model (4.1) for $N_{i(j\to k)}(t)$ reduces to a nonstationary Poisson process (*Lin et al.*, 2000).

In the cohort of 554 children, 333 of them have complete covariate data and wore the ActiGraph device with no interruption for a week. To account for the human biological circadian rhythms, we consider stratifying the baseline intensities by day (a 24-hour or 1440-minute cycle). Specifically, suppose that we first set the baseline function to refresh at the beginning of each day, i.e. $00 : 01 : 00$ *am* for clock-time Hour : Minute : Second, and then introduce a function $D(t)$ to map a follow-up time t to day $d \in \{1, \ldots, 7\}$ by $t \in [t_{0d}, t_{0(d+1)})$, where t_{0d} is the starting time for day d and $t_{0(d+1)} - t_{0(d)} = 24$ hours is the circadian cycle length. In addition, we use another function $B(t)$ to map time t to the daily clock-time $B(t) = t - t_{0d(t)}$. Consequently, an alternative the relative risk model equivalent to (4.2) that accounts for the circadian cycle may be specified given $D(t) = d$ and $B(t) = v$ as follows:

$$\lambda_{d(j\to k)}\,(v, \boldsymbol{Z}_{id}(v)) = \lambda_{0d(j\to k)}(v)\exp(\boldsymbol{\beta}'_{d(j\to k)}\boldsymbol{Z}_{id}(v)), \tag{4.3}$$

where $\boldsymbol{Z}_{id}(v)$ is the $\boldsymbol{Z}_i(t)$ at day d and clock-time v for subject i. Note that the baseline hazard and regression parameters in model (4.3) are specific to the stratum of an inter-state transition type and the follow-up day. Within the circadian cycle length $\tau = 1440$ minutes (24 hours), similar to above $\boldsymbol{Z}_{id}(v)$, we use a clock-time map to define the corresponding circadian event processes $N^*_{id(j\to k)}(v) = N^*_{i(j\to k)}(t)$, the transition-specific at-risk process $Y_{id(j\to k)}(v) = Y_{i(j\to k)}(t)$ and the observed circadian transition counts $N_{id(j\to k)}(v) = N_{i(j\to k)}(t)$ for $0 \leq t \leq \tau$. Model (4.3) can be analyzed using the regular partial-likelihood due to the fact that each subject contributes to at most one at-risk time interval within each transition-day stratum to give an appropriate filtration according to *Kalbfleisch and Prentice* (2002), Chapters 8-9. In other words, the at-risk states in day $d+1$ will not affect the at-risk sets of day d. It is commonly seen that the covariate effects are consistent across different days, validating the use of shared regression parameters, or $\boldsymbol{\beta}_{d(j\to k)} = \boldsymbol{\beta}_{j\to k}$ in model (4.3), the it is reduced to:

$$\lambda_{d(j\to k)}(v, \boldsymbol{Z}_{id}(v)) = \lambda_{0d(j\to k)}(v)\exp(\boldsymbol{\beta}'_{j\to k}\boldsymbol{Z}_{id}(v)). \tag{4.4}$$

Moreover, reducing the number of parameters of estimation can improve the estimation efficiency. Note that model (4.2) and model (4.4) are basically equivalent, namely $E(dN_{i(j\to k)}(t)|\mathcal{F}_{t^-}) = E(dN_{id(j\to k)}(v)|\mathcal{F}_{t^-}) = Y_{id(j\to k)}(v)\lambda_{d(j\to k)}(v, \boldsymbol{Z}_{id}(v))$ and $\lambda_{0(j\to k)}(t) = \lambda_{0d(j\to k)}(v)$ with identical $\boldsymbol{\beta}_{j\to k}$ estimates under the circadian mapping functions $B(t) = v$ and $D(t) = d$.

As has been discussed in Section 4.2, the subjects experience some daily renewal patterns, which makes it tempting to further simplify the model in (4.4) by resetting the baseline at the beginning of each day or the circadian cycle. For example, as pointed out in *Kalbfleisch and Prentice* (2002) and *Cook and Lawless* (2007), the recurrent event modeling based on sojourns or gap times reset time zero instantaneously after observing an event. However, most gap time models typically require additional (conditional) independence assumptions under, for example, frailty models. Technically we may modify model (4.4) to have a shared baseline hazard within a transition stratum of the following form:

$$\lambda_{d(j\to k)}(v, \boldsymbol{Z}_{id}(v)) = \lambda_{0(j\to k)}(v)\exp\left(\boldsymbol{\beta}'_{j\to k}\boldsymbol{Z}_{id}(v)\right). \tag{4.5}$$

However, the relevant estimation and inference become nontrivial because in (4.5), it is difficult to obtain a proper filtration to formulate the regular partial likelihood. This is due to the fact that each subject may be at risk for a state transition, e.g., $j \rightarrow k$, at the same daily clock time (e.g., v) in different days, and thus contribute to the same at-risk set multiple times. Therefore, in the presence of such shared baseline hazards, no appropriate sequence of filtrations can be defined, and consequently the parameter estimation based on the corresponding partial likelihoods is intractable.

To circumvent the difficulty of obtaining the partial likelihoods for parameter estimation, we propose to obtain the estimating equations via a suitable moment assumption

$$E\left(dN_{id(j\rightarrow k)}(v)|\boldsymbol{Z}_{id}(v), Y_{id(j\rightarrow k)}(v)\right) = Y_{id(j\rightarrow k)}(v)\mu_{d(j\rightarrow k)}\left(v, \boldsymbol{Z}_{id}(v)\right), \qquad (4.6)$$

in which, $\mu_{d,(j\rightarrow k)}\left(v, \boldsymbol{Z}_{id}(v)\right) = E\left(dN^*_{id(j\rightarrow k)(v)} \mid \boldsymbol{Z}_{id}(v), Y_{id(j\rightarrow k)}(v) = 1\right)$ denotes an event rate, rather than the aforementioned event intensity $\lambda_{d,(j\rightarrow k)}\left(v, \boldsymbol{Z}_{id}(v)\right)$ generated from a filtration. We further assume a semiparametric relative-risk structure for the rate model:

$$\mu_{d(j\rightarrow k)}\left(v, \boldsymbol{Z}_{id}(v)\right) = \mu_{0(j\rightarrow k)}(v)\exp\left(\widetilde{\boldsymbol{\beta}}'_{j\rightarrow k}\boldsymbol{Z}_{id}(v)\right). \qquad (4.7)$$

Note taht distinct from those of intensity models (4.5) and (4.4), $\widetilde{\boldsymbol{\beta}}_{j\rightarrow k}$ denotes the covariate effects of the rate model. Similar to classical relative-risk models, the rate model (4.7) also consists of a nonparametric baseline rate and the relative risk part with regression parameter $\widetilde{\boldsymbol{\beta}}_{j\rightarrow k}$. According to *Lin et al.* (2000), an advantage of the rate model is that we are able to obtain some martingale-like residuals, denoted by $M_{id(j\rightarrow k)}(v) = \int_0^v dM_{id(j\rightarrow k)}(s)ds$, that are based on the moment conditions given in (4.6) and their increments:

$$dM_{id(j\rightarrow k)}(v) = dN_{id(j\rightarrow k)}(v) - Y_{id(j\rightarrow k)}(v)\mu_{0d(j\rightarrow k)}\left(v, \boldsymbol{Z}_{id}(v)\right).$$

Though $M_{id(j\to k)}(v)$ is not a martingale here, it still provides a consistent estimation since the the moment condition $E(dM_{id(j\to k)}(v)) \equiv 0$ is satisfied according to (4.6). Based on this assumption, we then follow the similar arguments from *Lin et al.* (2000) to obtain two estimating equations for the baseline rates and the regression parameters $\widetilde{\boldsymbol{\beta}}_{j\to k}$ respectively:

$$\sum_{i}^{n}\sum_{d=1}^{7}\int_{0}^{\tau} dM_{id(j\to k)}(v) = 0, \text{ and} \tag{4.8}$$

$$\sum_{i}^{n}\sum_{d=1}^{7}\int_{0}^{\tau} \left\{\boldsymbol{Z}_{id}(v) - \bar{\boldsymbol{Z}}_{j\to k}(\widetilde{\boldsymbol{\beta}}_{j\to k}, v)\right\} dM_{id(j\to k)}(v) = 0, \tag{4.9}$$

where $\bar{\boldsymbol{Z}}_{j\to k}(\widetilde{\boldsymbol{\beta}}_{j\to k}, v) = S^{(1)}_{j\to k}(\widetilde{\boldsymbol{\beta}}_{j\to k}, v)/S^{(0)}_{j\to k}(\widetilde{\boldsymbol{\beta}}_{j\to k}, v)$. Note that the utility functions are defined as

$$S^{(m)}_{j\to k}(\widetilde{\boldsymbol{\beta}}_{j\to k}, v) = n^{-1}\sum_{i=1}^{n}\sum_{d=1}^{7} Y_{id(j\to k)}(v)\boldsymbol{Z}_{id}(v)^{\otimes m}\exp\left(\widetilde{\boldsymbol{\beta}}'_{j\to k}\boldsymbol{Z}_{id}(v)\right)$$

for $(m = 0, 1, 2)$, where for a vector $\boldsymbol{a}$, $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$ and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}'$.

In addition, the occurrence of transition $j \longrightarrow k$ may censor occurrences of other at-risk transitions, e.g. $j \longrightarrow l$ where $l \neq k$ and $l \neq j$. Therefore, the proposed multistate rate model is indeed to analyze the "crude" rate given the risk process $Y_{id(j\to k)}(v)$ or equivalently $Y_{i(j\to k)}(t)$ for $D(t) = d$ and $B(t) = v$, in an analogue to the concept of crude hazard, under the framework of competing risks, as opposed to the "net" rate (or hazard) that is not identifiable here without additional assumptions of independence between different types of transitions. Note that *Lin et al.* (2000) assumed random censoring given the covariates, which is different from the proposed crude rate model (4.6) here conditional on the at-risk process

Estimates of regression parameters $\hat{\widetilde{\boldsymbol{\beta}}}_{j\to k}$ are obtained by the roots of equation (4.9) which are numerically found by Newton-Raphson algorithm. Since all parameters $\widetilde{\boldsymbol{\beta}}_{j\to k}$ as well as baseline rates in (4.7) differ by transition strata, it is technically equivalent to fitting separate rate models for each transition . Besides, according to the first moment condition (4.8), estimation of the baseline rates follows an Aalen-Breslow-form:

$$\widehat{\mu}_{0(j\to k)}(\nu) = \frac{\sum_{i=1}^{n}\sum_{d=1}^{7} dN_{id(j\to k)}(\nu)}{nS^{(0)}(\hat{\widetilde{\boldsymbol{\beta}}}_{j\to k}, \nu)}. \ \nu \in [0, \tau) \tag{4.10}$$

As discussed in Section 4.2, accelerometer time series data have been aggregated on the basis of one-minute epoch, thus there might be multiple observations of the same transition type at a clock time ν. We plan to apply existing strategies in in literature to deal with any tied transitions (refer to Chapter 3.3 by *Therneau and Grambsch* (2000)). In the implementation, we employed the default setting with the Efron method in the R function `coxph` from package the survival (*Therneau*, 2015).

Following some regularity conditions and the framework of empirical process theory, we can establish large sample properties for the estimation and inference proposed in the multistate rate model (4.7). Due to the interplay between and within different types of transitions from each subject, we apply the robust sandwich variance estimator (*Lin et al.*, 2000). This robust standard error calculation is also called the grouped jackknife that relates to influence diagnostics (*Therneau and Grambsch*, 2000).

Model (4.7) may be further tailored to account for other confounding factors. For example, we could consider a finer stratification by adding additional variable like gender (indexed by g). Then, model (4.7) becomes

$$\mu_{dg(j\to k)}(\nu, \boldsymbol{Z}_{id}(\nu)) = Y_{id(j\to k)}(\nu)\mu_{0g(j\to k)}(\nu)\exp(\widetilde{\boldsymbol{\beta}}'_{g(j\to k)}\boldsymbol{Z}_{id}(\nu)), \tag{4.11}$$

where $\mu_{dg(j\to k)}$ and $\widetilde{\boldsymbol{\beta}}_{g(j\to k)}$ represent the shared baseline rate and the regression parameters respectively for transition $j \to k$ among subjects with gender $g \in \{1, 2\}$. In particular, we use $g = 1$ to denote boys and $g = 2$ to denote girls. In addition, both models (4.7) and (4.11) may be reduced to more parsimonious forms by removing some unimportant covariates and/or allowing some of the regression parameters to be shared across different strata. In the data analysis in Section 4.4, we will start from complicated stratified models with different regression parameters and then explore parsimonious model constructs to gain estimation efficiency.

4.4 Data Analysis

The accelerometer data and explanatory variables are summarized in Section 4.2 from a group of 333 children in the ELEMENT 2015 cohort study with no missing data. We now apply the multistate rate models in Section 4.3 to analyze the physical activity behaviors of these children.

4.4.1 Stratified by State Transition

We begin to apply the multistate rate model (4.7) stratified by transition types with stratum-specific regression parameters. We consider only baseline covariates in the analysis to form $\boldsymbol{Z}(\nu) = \boldsymbol{Z}$, including centered age, centered blood lead (Pb), and age-adjusted BMI (Zbfa). Tables 4.2-4.3 list the estimated effects $\hat{\hat{\boldsymbol{\beta}}}$ and their robust standard errors obtained from the R package. In particular, Table 4.2 presents the results for the increased-activity transitions ($j \rightarrow k$ for $j < k$) and Table 4.3 presents those for the decreased-activity transitions ($j \rightarrow k$ for $j > k$). In the multistate rate model (4.7), the exponential term $\exp(\hat{\hat{\boldsymbol{\beta}}}'\boldsymbol{Z})$ is a proportional multiplier of the baseline rate $\mu_{0(j\rightarrow k)}(\nu)$, representing a relative rate ratio contributed by the explanatory variables, among which the effect of Pb is of clinical interest. According to these results, we make following conclusions that are itemized below.

- As shown in Table 4.2, there exist significant positive effects of Pb on the increased-activity transition rates $0 \rightarrow 3$ and $2 \rightarrow 3$, judged by the corresponding p-values less than the significant level 0.05, which are also highlighted in bold in Table 4.2.

- Age tends to be negatively associated with increased-activity transitions ($j \rightarrow k$ with $j < k$) except for two transitions $0 \rightarrow 3$ and $2 \rightarrow 3$, possibly due to their small event numbers. For decreased-activities transitions, age tends to increase the transition rates significantly in four cases except for the transitions $3 \rightarrow 2$ and $3 \rightarrow 1$.

- The girls tended experience more frequent transitions in the cases of $0 \rightarrow 1$, $2 \rightarrow 1$ and $3 \rightarrow 1$, while less transitions in the cases of $1 \rightarrow 3$ and $2 \leftrightarrow 3$ (both $2 \rightarrow 3$ and $3 \rightarrow 2$). The girls were less vigorously active than boys, evidently by the fact that the girls tended to have less transitions from states of low activities to the vigorous state (state 3), while to experience more low-level increased-activity transitions. Moreover, the physical activity of the girls exhibited a higher frequency of dropping from the vigorous activity state to the lower activity states ($3 \rightarrow 1$ and $3 \rightarrow 2$) than boys.

- The age-adjusted BMI variable Zbfa is a significant positive predictor for the low-activity transition $0 \rightarrow 1$ (between the sedentary and moderately active states), but a significant negative predictor for transitions to the vigorously active state ($1 \rightarrow 3$ and $2 \rightarrow 3$).

- The nonparametrically estimated baseline rates are plotted in Figure 4.6 with the x-axes following daily clock hours. We notice that the time-varying pattern of the event rates at the low-level activity transitions of both directions $0 \leftrightarrow 1$ align well with the human routine daily activity pattern; the transition rates stay low during the sleeping period and elevate during the daytime. The estimated baseline event rates for the other transitions look more noisy due to the lack of regular recorded for these transitions.

- Concordance index or C-index (*Harrell et al.*, 1982) is also reported in the analysis. A C-index is useful to evaluate the model predictability (see Tables 4.2-4.3). The aggregated C-indices for the model of increased-activity transitions and that of decreased-activity transitions are 0.537 and 0.512, respectively.

- The martingale-like residuals for both the increased-activity and decreased-activity transition models are plotted in Figure 4.7. In an agreement with the fact that the multistate rate models for the increased-activity transitions have a higher pooled C-index than the models for the decreased-activity transitions, the residuals for the former are also distributed in a more balanced fashion than the latter.

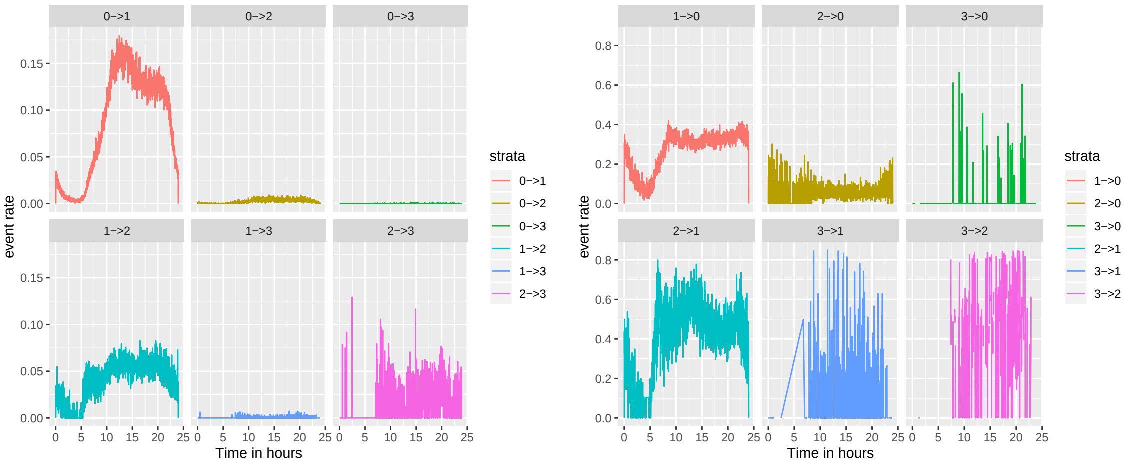


Figure 4.6: Estimated baseline event rates of increased-activity transitions (left) and decreased-activity transitions (right). The x-axes are labeled by daily clock time of $0 - 24$ hours.

4.4.2 Stratified by State Transition and Gender

Next, we consider a finer stratification in model (4.11) by allowing the baseline event rate and regression parameters to be gender-transition specific. This gender stratification incorporates not only a gender-specific baseline rate for each transition type, but also the interaction effects between gender and explanatory variables

Table 4.2: Results of the multistate rate model stratified by type of increased-activity transitions for all children. Significant Pb effects are highlighted in bold.

Stratum	C-index	Variable	Estimation	Robust SE	Z-score	P-value
(0 → 1)	0.536	age	-0.058	0.009	-6.735	< 0.001
		female	0.087	0.028	3.117	0.002
		Zbfa	0.029	0.011	2.634	0.008
		Pb	-0.002	0.004	-0.676	0.499
(0 → 2)	0.563	age	-0.127	0.020	-6.509	< 0.001
		female	-0.085	0.068	-1.257	0.209
		Zbfa	0.005	0.025	0.211	0.833
		Pb	0.006	0.009	0.615	0.539
(0 → 3)	0.583	age	-0.155	0.100	-1.538	0.124
		female	-0.274	0.308	-0.890	0.374
		Zbfa	-0.128	0.104	-1.231	0.218
		Pb	**0.033**	**0.016**	**2.081**	**0.037**
(1 → 2)	0.540	age	-0.069	0.015	-4.505	< 0.001
		female	-0.077	0.049	-1.569	0.117
		Zbfa	-0.008	0.018	-0.470	0.638
		Pb	0.001	0.006	0.157	0.875
(1 → 3)	0.642	age	-0.175	0.062	-2.826	0.005
		female	-0.764	0.179	-4.258	< 0.001
		Zbfa	-0.172	0.059	-2.900	0.004
		Pb	-0.001	0.021	-0.054	0.957
(2 → 3)	0.617	age	-0.024	0.055	-0.435	0.664
		female	-0.683	0.181	-3.781	< 0.001
		Zbfa	-0.142	0.054	-2.644	0.008
		Pb	**0.040**	**0.015**	**2.644**	**0.008**

Table 4.3: Results of the multistate rate model stratified by types of decreased-activity transitions for all children. Significant Pb effects are highlighted in bold.

Stratum	C-index	Variable	Estimation	Robust SE	Z-score	P-value
(1 → 0)	0.511	age	0.019	0.008	2.492	0.013
		female	-0.008	0.027	-0.279	0.781
		Zbfa	-0.001	0.010	-0.062	0.951
		Pb	-0.004	0.007	-0.575	0.565
(2 → 0)	0.533	age	0.056	0.025	2.253	0.024
		female	0.118	0.086	1.383	0.167
		Zbfa	0.014	0.033	0.405	0.686
		Pb	0.005	0.008	0.682	0.496
(3 → 0)	0.584	age	0.261	0.131	1.994	0.046
		female	-0.143	0.476	-0.301	0.763
		Zbfa	-0.032	0.226	-0.140	0.888
		Pb	-0.036	0.081	-0.440	0.660
(2 → 1)	0.570	age	0.056	0.024	2.305	0.021
		female	0.283	0.078	3.616	< 0.001
		Zbfa	0.053	0.029	1.803	0.071
		Pb	0.003	0.008	0.359	0.720
(3 → 1)	0.587	age	-0.056	0.054	-1.044	0.296
		female	0.629	0.173	3.640	< 0.001
		Zbfa	0.005	0.067	0.079	0.937
		Pb	-0.016	0.013	-1.268	0.205
(3 → 2)	0.556	age	-0.040	0.019	-2.113	0.035
		female	-0.221	0.078	-2.816	0.005
		Zbfa	-0.003	0.029	-0.114	0.909
		Pb	0.003	0.006	0.531	0.596

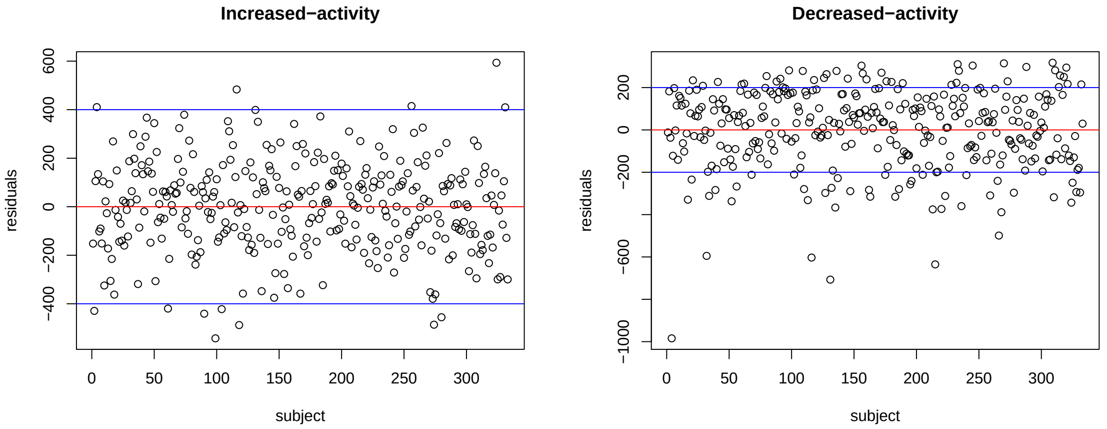


Figure 4.7: Summed residuals within subjects for the increased-activity transition models (left) and the decreased-activity transition models (right)

of interest, which can be time-varying. Thus, $\boldsymbol{Z}$ will exclude the gender variable. The estimation results, including C-index for each transition stratum, are listed in Table 4.4 for boys and Table 4.5 for girls. Their estimated baseline rates and the resulting martingale-like residuals are plotted in Figures 4.8-4.11.

- In contrast to the non-significant effect of Pb on the transition $0 \to 1$ in Table 4.2, the modeling strategy with the stratification of gender detects a significant positive effect of Pb for girls experiencing this transition, while a significant negative effect of Pb for boys. Cancellation of the two opposite Pb effects is responsible for the non-significance of Pb in the previous analysis without stratifying by gender. Pb is once again found to be positively associated with vigorous activity transitions like $2 \to 3$ among boys, but not among girls; and the previously significant positive association of Pb in the $0 \to 3$ transition disappears. Pb is found now a significant positive predictor for decreased-activity transitions $3 \to 0$ and $3 \to 2$ among girls.

- Results of the variables like age and Zbfa are quite consistent in both gender strata and in comparison with the model without a gender stratification (Table 4.2).

- Age is negatively associated with most increased-activity transitions except for vigorous elevations like $0 \rightarrow 3$ and $2 \rightarrow 3$ for boys, and all vigorous activity transitions ($j \rightarrow 3$ for $j \in \{0, 1, 2\}$) among girls. Age is still a significant positive predictor of all decreased-activity transitions among boys except for transitions of $3 \rightarrow 1$ and $3 \rightarrow 2$, while stays non-significant for all decreased-activity transitions among girls.

- Zbfa is significantly positively associated with low-level transition $0 \rightarrow 1$ among boys and significantly negatively associated with the high-level transition $1 \rightarrow 3$ for both genders.

- The gender-transition specific C-index for each stratum can be found in Tables 4.4 and 4.5. For the increased-activity transitions, the aggregated C-index pooled over all the strata of the increased-activity transitions among the boys is 0.544, higher than that of girls, which is 0.526; and for the decreased-activity transitions, the aggregated C-index is still higher among boys (0.517) than girls (0.513). Both findings imply a better prediction of the models for the boys than the girls. Note that the C-index values of the decreased-activity transition models for both genders are slightly higher than the one obtained from the model without stratifying by gender (0.511). The C-index for $3 \rightarrow 1$ among boys is lower than 0.5, implying a worse prediction performance than a random guess. These models can hardly achieve C-indices 0.7 or above with such few explanatory variables used in the analyses.

- Figures 4.8 and 4.9 show the estimated baseline rates for both directions of transitions $0 \leftrightarrow 1$ for both genders, which are quite similar to each other and are comparable to those from the models without gender stratification.

Table 4.4: Results of the multistate rate model stratified by types of activity transitions among boys. Significant Pb effects are highlighted in bold.

Stratum	C-index	Variable	Estimation	Robust SE	Z-score	P-value
$(0 \rightarrow 1)$	0.542	age	-0.075	0.013	-6.005	< 0.001
		Zbfa	0.031	0.015	2.103	0.035
		Pb	**-0.008**	**0.003**	**-2.315**	**0.021**
$(0 \rightarrow 2)$	0.584	age	-0.176	0.027	-6.544	< 0.001
		Zbfa	-0.003	0.034	-0.080	0.937
		Pb	0.002	0.010	0.173	0.862
$(0 \rightarrow 3)$	0.621	age	-0.243	0.130	-1.872	0.061
		Zbfa	-0.140	0.128	-1.094	0.274
		Pb	0.026	0.023	1.128	0.259
$(1 \rightarrow 2)$	0.551	age	-0.100	0.022	-4.487	< 0.001
		Zbfa	-0.021	0.024	-0.854	0.393
		Pb	0.002	0.005	0.341	0.733
$(1 \rightarrow 3)$	0.602	age	-0.194	0.065	-2.973	0.003
		Zbfa	-0.161	0.072	-2.248	0.025
		Pb	-0.002	0.025	-0.077	0.939
$(2 \rightarrow 3)$	0.565	age	0.004	0.063	0.057	0.954
		Zbfa	-0.137	0.063	-2.163	0.031
		Pb	**0.038**	**0.015**	**2.503**	**0.012**
$(1 \rightarrow 0)$	0.516	age	0.026	0.012	2.152	0.031
		Zbfa	-0.012	0.013	-0.924	0.355
		Pb	-0.002	0.009	-0.163	0.871
$(2 \rightarrow 0)$	0.529	age	0.078	0.030	2.591	0.010
		Zbfa	0.034	0.047	0.722	0.470
		Pb	0.004	0.008	0.555	0.579
$(3 \rightarrow 0)$	0.667	age	0.465	0.166	2.802	0.005
		Zbfa	-0.155	0.295	-0.526	0.599
		Pb	-0.217	0.180	-1.206	0.228
$(2 \rightarrow 1)$	0.556	age	0.078	0.023	3.355	0.001
		Zbfa	0.104	0.035	2.931	0.003
		Pb	0.001	0.007	0.092	0.927
$(3 \rightarrow 1)$	0.486	age	0.038	0.058	0.648	0.517
		Zbfa	-0.024	0.080	-0.296	0.767
		Pb	-0.032	0.017	-1.922	0.055
$(3 \rightarrow 2)$	0.563	age	-0.036	0.023	-1.550	0.121
		Zbfa	0.020	0.039	0.512	0.609
		Pb	0.001	0.006	0.176	0.860

Table 4.5: Results of the multistate rate model stratified by types of activity transitions among girls. Significant Pb effects are highlighted in bold.

Stratum	C-index	Variable	Estimation	Robust SE	Z-score	P-value
(0 → 1)	0.525	age	-0.042	0.011	-3.738	< 0.001
		Zbfa	0.021	0.016	1.314	0.189
		Pb	**0.014**	**0.007**	**2.087**	**0.037**
(0 → 2)	0.541	age	-0.076	0.026	-2.883	0.004
		Zbfa	0.011	0.037	0.295	0.768
		Pb	0.026	0.018	1.478	0.139
(0 → 3)	0.611	age	-0.052	0.138	-0.376	0.707
		Zbfa	-0.136	0.167	-0.815	0.415
		Pb	0.116	0.081	1.436	0.151
(1 → 2)	0.525	age	-0.041	0.021	-1.969	0.049
		Zbfa	0.010	0.026	0.384	0.701
		Pb	0.000	0.012	0.017	0.987
(1 → 3)	0.591	age	-0.142	0.122	-1.165	0.244
		Zbfa	-0.207	0.094	-2.194	0.028
		Pb	0.014	0.044	0.313	0.754
(2 → 3)	0.582	age	-0.090	0.106	-0.848	0.396
		Zbfa	-0.162	0.092	-1.754	0.080
		Pb	0.049	0.045	1.089	0.276
(1 → 0)	0.513	age	0.015	0.010	1.478	0.139
		Zbfa	0.018	0.016	1.070	0.284
		Pb	-0.011	0.007	-1.654	0.098
(2 → 0)	0.518	age	0.035	0.039	0.900	0.368
		Zbfa	-0.027	0.044	-0.616	0.538
		Pb	0.009	0.021	0.411	0.681
(3 → 0)	0.636	age	0.208	0.232	0.899	0.369
		Zbfa	1.298	0.707	1.835	0.067
		Pb	**0.617**	**0.284**	**2.172**	**0.030**
(2 → 1)	0.528	age	0.033	0.039	0.841	0.400
		Zbfa	-0.027	0.045	-0.606	0.544
		Pb	0.014	0.018	0.811	0.417
(3 → 1)	0.576	age	-0.135	0.089	-1.516	0.130
		Zbfa	-0.186	0.141	-1.317	0.188
		Pb	-0.023	0.058	-0.401	0.688
(3 → 2)	0.614	age	-0.076	0.041	-1.849	0.064
		Zbfa	-0.101	0.055	-1.845	0.065
		Pb	**0.051**	**0.024**	**2.114**	**0.035**

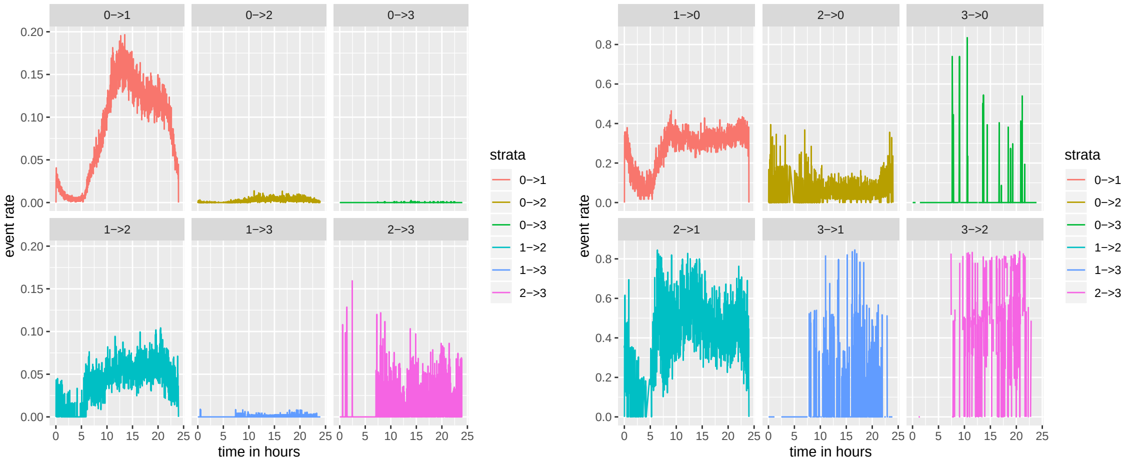


Figure 4.8: The estimated baseline event rates the increased-activity transitions (left) and decreased-activity transitions (right) among the boys. The x-axes are labeled by daily clock time of $0 - 24$ hours.

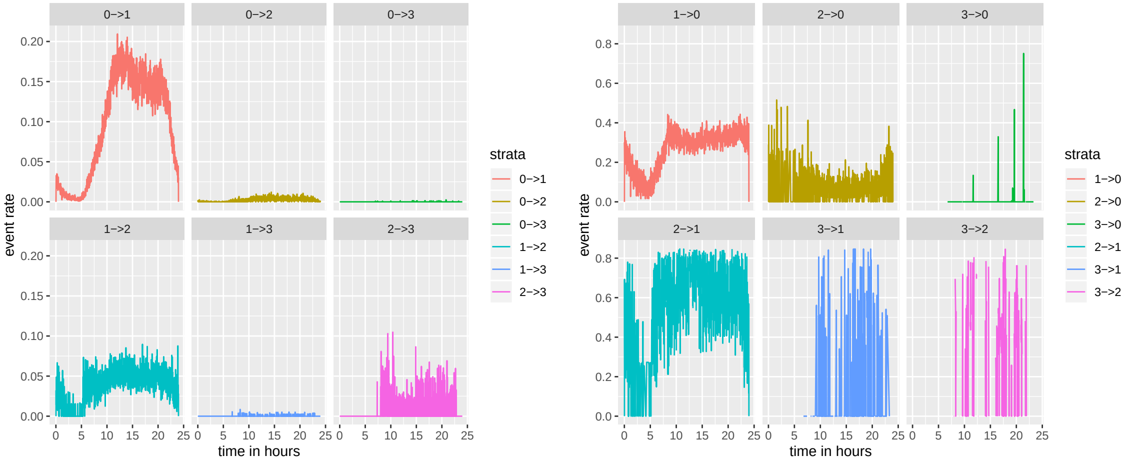


Figure 4.9: The estimated baseline event rates for the increased-activity transitions (left) and the decreased-activity transitions (right) among the girls. The x-axes are labeled by daily clock time of $0 - 24$ hours.

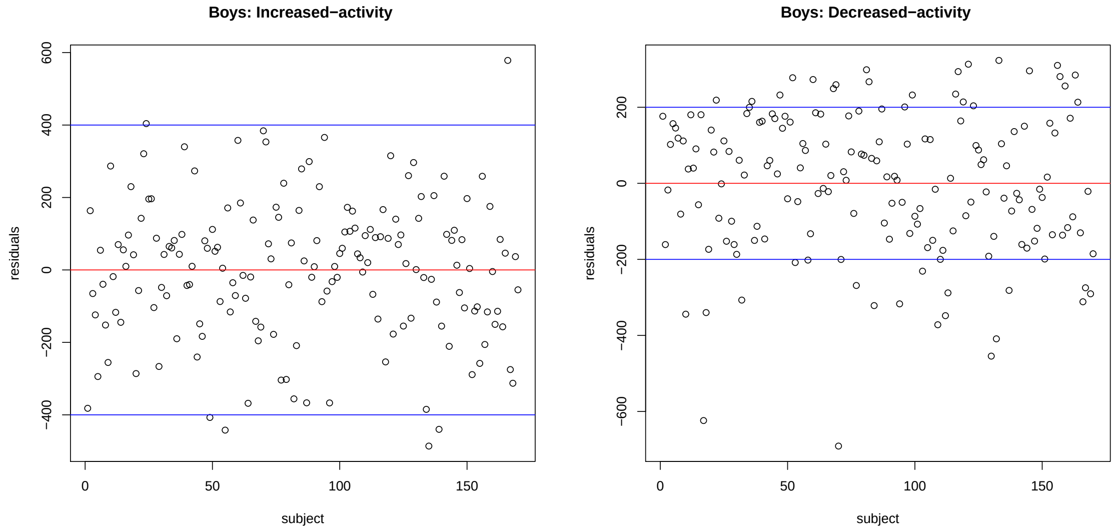


Figure 4.10: Summed residuals within subjects for the increased-activity transition models (left) and the decreased-activity transition models (right) among boys.

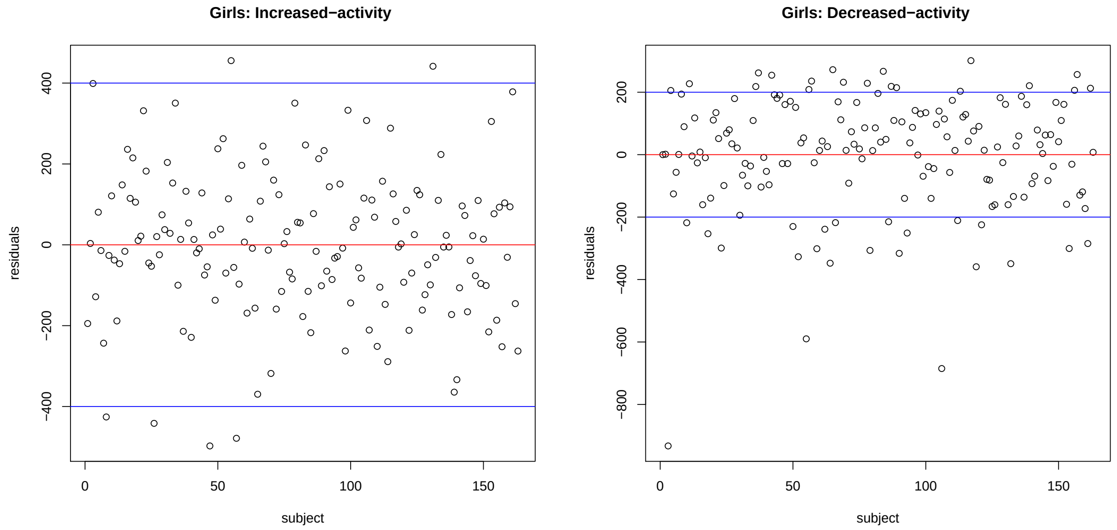


Figure 4.11: Summed residuals within subjects for the increased-activity transition models (left) and the decreased-activity transition models (right) among girls.

4.4.3 Parsimonious Models

We further simplify model (4.11) by allowing some regression parameters to be shared among some strata. We apply the Wald test with the robust sandwich variance matrices to test for the hypothesized parsimony in the framework of estimating functions. Once again, our proposed multistate models are purely based on moment conditions. The proposed models (4.7) and (4.11), as well as their extensions with shared regression effects, are formulated by the event rates rather than the event intensities. To assess the effects of Pb in these models, we do not consider any parsimonious specification of Pb related parameters. With some trials and errors based based on the Walt tests, we obtain the simplified models that share common regression parameters across different gender-transition strata and their estimates are summarized in Tables 4.6-4.9, where the reference groups are set as $0 \rightarrow 1$ and $1 \rightarrow 0$ for the increased-activity transition models and the decreased-activity transition models, respectively. Note that we did not enumerate all possible models, therefore, there may be other parsimonious models providing a comparable or even better prediction performance. In Table 4.6 for example, the models of increased-activity transitions among boys share a common effect of age in all transition strata except for $0 \rightarrow 2$ and $0 \rightarrow 3$. Effects of Zbfa are specified to be common for three groups of transitions, including transitions $1 \rightarrow 3$ and $2 \rightarrow 3$, transition $1 \rightarrow 2$, and the rest, respectively.

We find that direction and significance of the Pb effects in all parsimonious models are almost unchanged. The C-index values are 0.544 and 0.516 for the increased-activity and the decreased-activity transitions among boys, and are 0.526 and 0.513 for those among girls. There is little reduction in the prediction accuracy with the parsimonious models in (4.11) with no shared effects. Models with shared effects lead to better interpretability and accuracy for cases that some transitions display quite comparable behaviors. The statistical power is also expected to improve for the estimation of some shared effects, especially among small strata with fewer transition counts, since their estimation can be improved largely by pooling over the events (transitions) from other strata. In a high agreement to the aforementioned findings, the plots of the estimated baseline rates and martingale residual in Figures 4.12-4.15 obtained from the parsimonious models resemble largely Figures 4.8-4.11. Therefore, we conclude that the stratified parsimonious models produce comparable estimation with better parameter interpretations and higher statistical efficiency.

Table 4.6: Results of the multistate rate model with shared effects among boys for the increased-activity transitions. Significant Pb (interaction) effects are highlighted in bold. Age is shared in all transitions except for $0 \rightarrow 2$ and $1 \rightarrow 3$ transitions; Zbfa has a separate effect for $1 \rightarrow 2$ transition, a shared effect among $1 \rightarrow 3$ and $2 \rightarrow 3$, and a common effect among the rest. Covariates Age, Zbfa and Pb have their (shared) effects for the transition $0 \rightarrow 1$ as the reference.

Variable	Estimation	Robust SE	Z-score	P-value
Age	-0.078	0.011	-6.829	< 0.001
Age*I($0 \rightarrow 2$)	-0.094	0.023	-4.172	< 0.001
Age*I($1 \rightarrow 3$)	-0.115	0.064	-1.794	0.073
Zbfa	0.030	0.015	2.014	0.044
Zbfa*I($1 \rightarrow 2$)	-0.048	0.026	-1.841	0.066
Zbfa*I($1 \rightarrow 3$ or $2 \rightarrow 3$)	-0.178	0.060	-2.943	0.003
Pb	**-0.007**	**0.003**	**-2.275**	**0.023**
Pb*I($0 \rightarrow 2$)	0.010	0.010	1.028	0.304
Pb*I($0 \rightarrow 3$)	0.031	0.024	1.288	0.198
Pb*I($1 \rightarrow 2$)	0.008	0.006	1.394	0.163
Pb*I($1 \rightarrow 3$)	0.006	0.025	0.239	0.811
Pb*I($2 \rightarrow 3$)	**0.050**	**0.015**	**3.281**	**0.001**

4.5 Conclusions

In this chapter, we proposed a family of stratified multistate event rate models to analyze the temperal transitions of physical activity states of the 333 children with complete variables in the ELEMENT cohort from Mexico City. The central goal of scientific interest here is to examine the association between lead exposure and physical activity transitions related to the children's behavior patterns. The

Table 4.7: Results of the multistate rate model with shared effects among boys for decreased-activity transitions. Significant Pb (interaction) effects are highlighted in bold. Age has a shared effect among all transitions except for the 2 → 0 and 2 → 1 transitions, with a shared effect that is different from the reference, or another distinct effect for the 3 → 0 transition; Zbfa has a common effect shared among all transitions except for 2 → 1. Covariates Age, Zbfa and Pb, all have their (shared) effects for the transition 1 → 0 as the reference.

Variable	Estimation	Robust SE	Z-score	P-value
Age	0.026	0.012	2.155	0.031
Age*I(2 → 0 or 2 → 1)	0.052	0.026	2.015	0.044
Age*I(3 → 0)	0.427	0.175	2.446	0.014
Zbfa	-0.011	0.013	-0.856	0.392
Zbfa*I(2 → 1)	0.115	0.039	2.954	0.003
Pb	-0.001	0.009	-0.157	0.875
Pb*I(2 → 0)	0.004	0.012	0.327	0.744
Pb*I(3 → 0)	-0.216	0.185	-1.168	0.243
Pb*I(2 → 1)	0.002	0.015	0.148	0.882
Pb*I(3 → 1)	-0.029	0.017	-1.765	0.078
Pb*I(3 → 2)	-0.004	0.011	-0.314	0.754

Table 4.8: Results of the multistate rate model with shared effects among girls for the increased-activity transitions. Significant Pb (interaction) effects are highlighted in bold. Age is shared in all transitions except for the transition 0 → 2 and the transition 1 → 3, wherein each has a distinct effect; Zbfa has a common effect for all transitions except for the 1 → 3 and 2 → 3 transitions. Covariates Age, Zbfa and Pb have their (shared) effects for the transition 0 → 1 as the reference.

Variable	Estimation	Robust SE	Z-score	P-value
Age	-0.042	0.010	-4.016	< 0.001
Age*I(0 → 2)	-0.034	0.020	-1.699	0.089
Age*I(1 → 3)	-0.097	0.120	-0.811	0.417
Zbfa	0.019	0.015	1.269	0.204
Zbfa*I(1 → 3 or 2 → 3)	-0.186	0.088	-2.113	0.035
Pb	**0.014**	**0.007**	**2.107**	**0.035**
Pb*I(0 → 2)	0.012	0.015	0.788	0.431
Pb*I(0 → 3)	0.094	0.082	1.144	0.253
Pb*I(1 → 2)	-0.014	0.011	-1.217	0.224
Pb*I(1 → 3)	-0.002	0.042	-0.038	0.970
Pb*I(2 → 3)	0.030	0.042	0.727	0.467

Table 4.9: Results of the multistate rate model with shared effects among girls for decreased-activity transitions. Significant Pb (interaction) effects are highlighted in bold. Age has a shared effect except for the transition 3 → 2; Zbfa has distinct effect values for transitions, 3 → 0, 3 → 1 and 2 → 1, respectively, and a shared effect for the rest. Covariates Age, Zbfa and Pb, all have their (shared) effects for the transition 1 → 0 as the reference.

Variable	Estimation	Robust SE	Z-score	P-value
Age	0.018	0.011	1.717	0.086
Age*I(3 → 2)	-0.094	0.043	-2.210	0.027
Zbfa	0.010	0.015	0.679	0.497
Zbfa*I(3 → 0)	0.982	0.487	2.016	0.044
Zbfa*I(3 → 1)	-0.149	0.151	-0.988	0.323
Zbfa*I(3 → 2)	-0.112	0.059	-1.905	0.057
Pb	-0.011	0.007	-1.630	0.103
Pb*I(2 → 0)	0.020	0.018	1.119	0.263
Pb*I(3 → 0)	**0.518**	**0.195**	**2.660**	**0.008**
Pb*I(2 → 1)	0.025	0.016	1.616	0.106
Pb*I(3 → 1)	-0.024	0.055	-0.443	0.658
Pb*I(3 → 2)	**0.062**	**0.026**	**2.396**	**0.017**

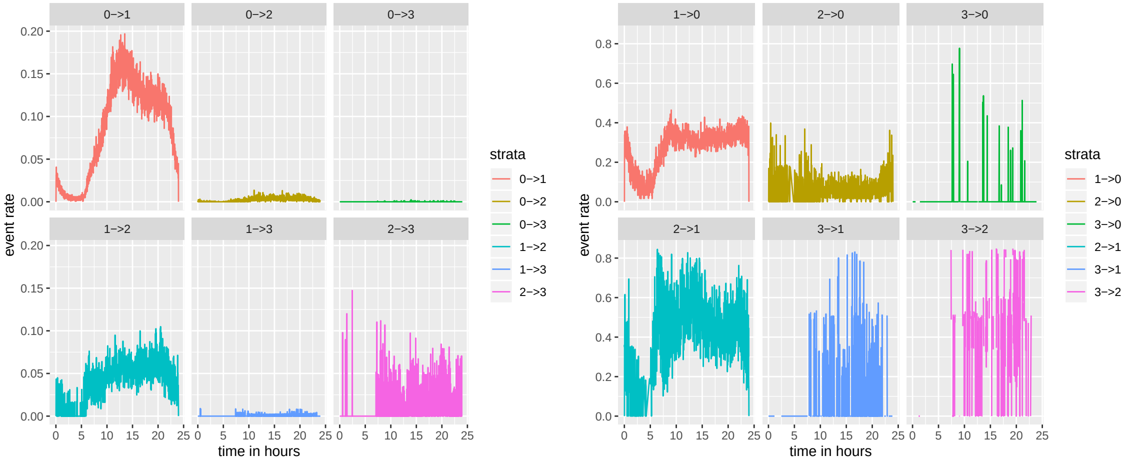


Figure 4.12: Estimated baseline event rates for the increased-activity transitions (left) and the decreased-activity transitions (right) among boys with shared regression parameters across different transitions (see Tables 4.6 and 4.7). The x-axes are labeled by daily clock time of $0 - 24$ hours.

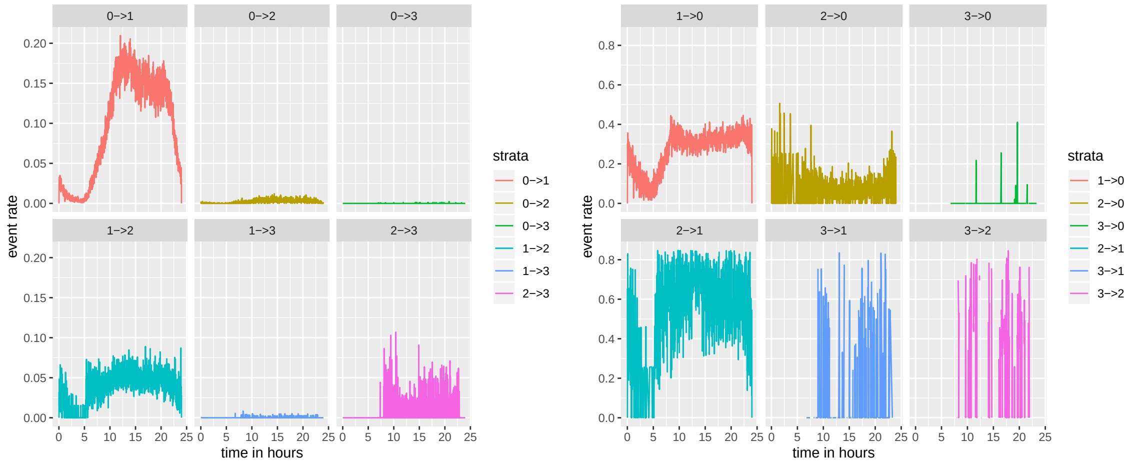


Figure 4.13: Estimated baseline event rates for the increased-activity transitions (left) and the decreased-activity transitions (right) among girls with shared regression parameters across different transitions (see Tables 4.8 and 4.9). The x-axes are labeled by daily clock time of $0 - 24$ hours.

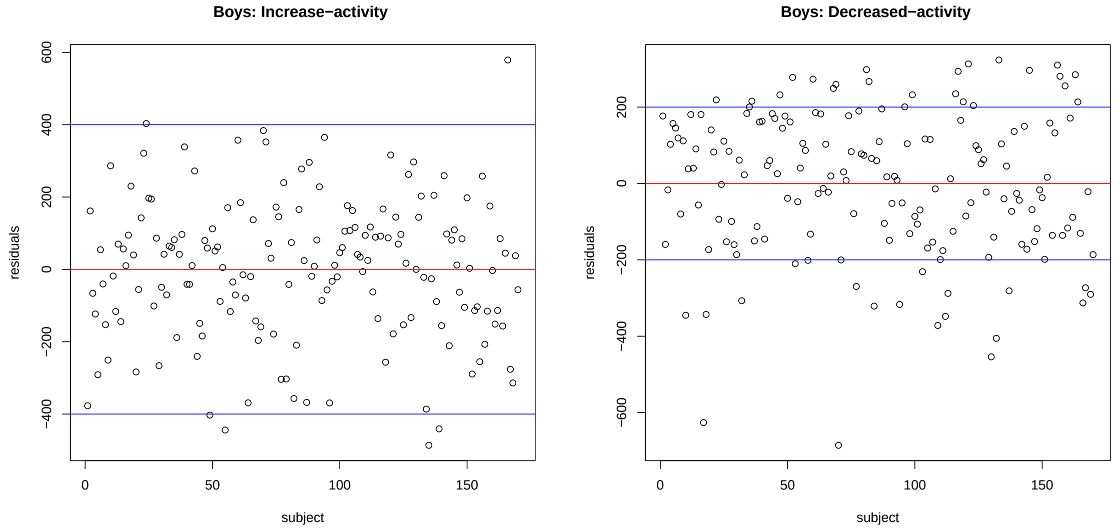


Figure 4.14: Summed residuals within subjects for the increased-activity transitions (left) and the decreased-activity transitions (right) among boys with shared regression parameters across different transitions (see Tables 4.6 and 4.7).

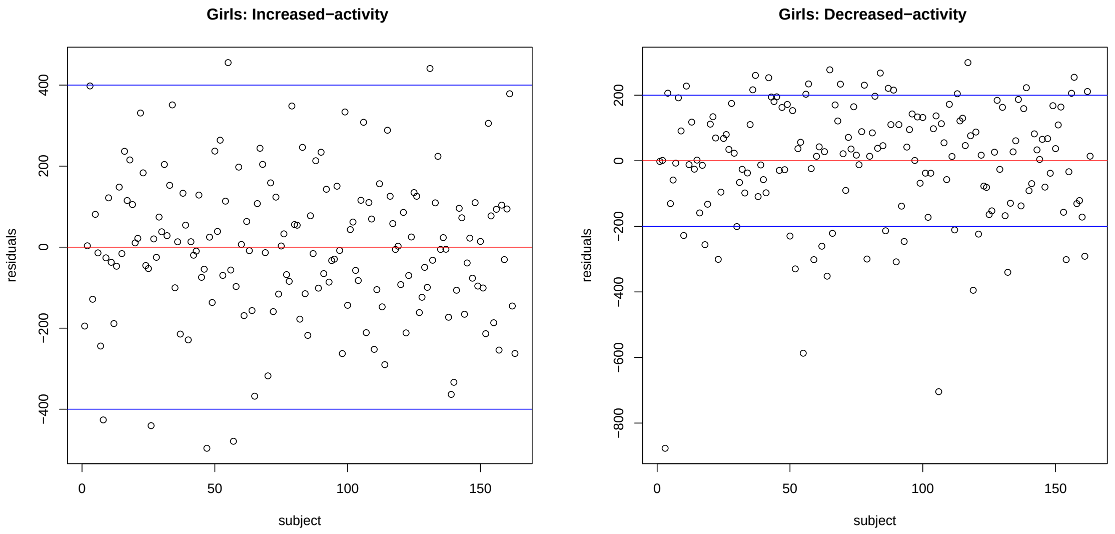


Figure 4.15: Summed residuals within subjects for the increased-activity transitions (left) and the decreased-activity transitions (right) among girls with shared regression parameters across different transitions (see Tables 4.8 and 4.9)

physical activity states were categorized into four types ranging from sedentary to vigorously active using the cut-offs defined by *Chandler et al.* (2016) for the minute-epoch vector magnitudes. We modeled the activity transitions along the line of time-dependent multivariate event data analysis, in that, the proposed family of multistate rate models (4.7) inherits the technical framework of the rate models (*Lin et al.*, 2000) and the conceptual framework of competing risks (*Kalbfleisch and Prentice*, 2002).

Comparing the analysis results obtained by the multistate rate models, respectively, with or without gender stratification, we have shown different patterns of transition rate curves for girls and boys. In particular, for the increased-activity transitions, the girls tended to experience higher frequent transitions in the a low-activity category $0 \rightarrow 1$ than the boys, and lower frequent vigorous-activity transitions like $1 \rightarrow 3$ and $2 \rightarrow 3$. Moreover, the girls are more likely to drop steeply from the vigorously active state to lower activity states ($3 \rightarrow 1$ and $3 \rightarrow 2$) than boys. By adding the gender to achieve a finer stratification, lead exposure Pb was found positively associated with transition $0 \rightarrow 1$ among the girls, but negatively associated with this transition among the boys, as opposed to the non-significant association detected under no gender stratification. Also, Pb is positively associated with the $2 \rightarrow 3$ transition among boys not girls, and an elevated level of blood Pb tended to increase the rates of decreased-activity transitions from the vigorous active state to lower ones (e.g., $3 \rightarrow 0$ and $3 \rightarrow 2$). In addition, allowing to share some similar stratum-specific regression parameters can improve the estimation efficiency. The parsimonious models may be derived by conducting the Wald test using the robust asymptotic variance estimates. The simplified models were found to exhibit comparable performances to the models with no shared effects, but for better interpretations.

Due to the missingness in the explanatory variables, in particular to the Pb concentration, only 333 out of 545 subjects were finally included for the data analysis. The mechanism of missingness seems to be at random since their marginal and conditional transition distribution plots are quite similar for the original data set ($n = 545$, Figures 4.2 and 4.3) and the complete data set ($n = 333$, Appendix Figures C.1 and C.2). A possible remedy to overcome the missing data issue is to invoke the imputation method, which can be done in the future analysis (*Little*, 1995; *Catellier et al.*, 2005). Another issue is the validity of the physical activity states that are determined by the validated cutoff values from one study. It is know that misclassification of the activity states may lead to biased conclusions (*Staudenmayer et al.*, 2012). More extensive physical activity measures on different populations with various normalized metrics that are comparable between devices and platforms (*Bai et al.*, 2014, 2016) would largely improve the classification accuracy. It is hoped that the proposed event rate models can fit in a more general framework of multistate event data analysis beyond accelerometer data.

CHAPTER V

An Epidemiological Forecast Model and Software Assessing Interventions on the COVID-19 Epidemic

5.1 Introduction

The outbreak of the coronavirus disease 2019 or COVID-19, originated in Wuhan, the capital city of Hubei province. From there, it spread quickly through Hubei and then to China and globally to more than 200 countries, causing over 10 million confirmed cases and about 500,000 deaths cumulatively, according to the WHO data available in June, 2020. Back to February 25, 2020, in China this large-scale epidemic had caused a total of 78,195 confirmed infections, 2,718 deaths. Since the outbreak of the epidemic, many clinical papers (*Jung et al.*, 2020; *Chen et al.*, 2020; *Xiang et al.*, 2020; *Xu et al.*, 2020; *Imai et al.*, 2020; *Gralinski and Menachery*, 2020; *Luk et al.*, 2019; *Fan et al.*, 2019; *Hui et al.*, 2020; *Holshue et al.*, 2020; *Guan et al.*, 2020; *Rothe et al.*, 2020; *Huang et al.*, 2020; *Zhu et al.*, 2020; *Wang et al.*, 2020a) have been published to uncover limited but important knowledge of COVID-19, including that (i) COVID-19 is an infectious disease caused by SARS-CoV-2, a virus closely related to the SARS coronavirus (SARS-CoV) (*Luk et al.*, 2019; *Fan et al.*, 2019; *Subissi et al.*, 2014; *World Health Organization*, 2020a); (ii) it can spread from person to person, primarily via droplet transmission (*Hui et al.*, 2020; *Holshue et al.*, 2020); (iii) it has a relatively high person-to-person transmission rate, especially via close contact; (iv) the median incubation time is approximately 5-6 days (*Lauer et al.*, 2020; *Backer et al.*, 2020), which can be as long as 24 days (*Guan et al.*, 2020); and (v) asymptomatic person carrying SARS-CoV-2 is contagious (*Rothe et al.*, 2020). This epidemic has been concerning not only in China but also in the rest of the world given the fast growing number of infected cases in South Korea, Japan, Italy, US, India, etc.

Quarantine or medical isolation is a key non-pharmaceutical intervention approach to stop the spreading of infectious diseases such as SARS (*World Health Organization*, 2020b; *Smith*, 2006; *World Health Organization*, 2003) and plague (*Dennis et al.*, 1999). The basic idea of quarantine and isolation is to separate infected cases from the susceptible population and *vice versa*. Since mid-January 2020, the Chinese government has implemented all kinds of very strict in-home isolation protocols nationwide, which have been elevated day by day through various government enforced quarantine and societally organized inspections to control the spread of COVID-19 in Hubei and other regions in China. In the meantime, the Chinese government has quickly increased the capacity of hospitals or as such that took symptomatic patients to be quarantined and treated by medical doctors and nurses.

The question of the most importance, which draws most attention, concerns when the spread of COVID-19 will end. This question has to be answered by a prediction model using the daily most-updated data from the China CDC. Moreover, the complexity of the impact of human interventions on the spread of COVID-19, including but not limited to in-home quarantine, hospitalization, community enforcement of wearing masks, minimizing outdoor activities, and changed diagnostic criteria by the government makes it difficult for a prediction model to take such features into account in order to provide meaningful analyses and hopefully reasonable predictions. Currently ost existing prediction models do not have the capacity to incorporate changing interventions in reality, and with no such critical component of time-varying intervention in the model, predicted turning points would be untrustworthy. Our new model and analytic toolbox aims to fill in this significant gap.

We develop an R package `eSIR` (*Wang et al.*, 2020b) for R (*R Core Team*, 2018b), that helps accomplish the following specific aims:

AIM 1: Incorporate time-varying quarantine protocols in the model of COVID-19 infection dynamics via an extension of the classical epidemiological SIR model. This dynamic infection system necessitates the forecast of the future trend of COVID-19 spread.

AIM 2: Provide an R software package to health workers who can conveniently perform their own analyses using their substantive knowledge and perhaps better quality data from provinces in China or from other countries.

We hope to provide a data analytic toolbox to people who may have better domain-specific knowledge and experience as well as high quality data to carry out independent predictions.

Our informatics toolbox is built upon a state-space model (*Zhu et al.*, 2012; *Jørgensen et al.*, 1999; *Song*, 2000; *Jøsrgensen and Song*, 2007) shown in Figure 5.1 with an extended Markov SIR model (*Kermack and McKendrick*, 1927), which has the following key features: (i) Our model is specified with the temporally varying prevalence of susceptible, infected and removed (recovered and death) compartments, not directly on time series of respective counts; (ii) estimation and inference are carried out and implemented using Markov Chain Monte Carlo (MCMC); (iii) it outputs various sample draws from the posteriors of the model parameters (e.g. transmission and removal rates) and the underlying prevalence of susceptible, infected and removed compartments, as well as their credible intervals. The latter is of extreme importance to quantify prediction uncertainty. In addition, this toolbox provides predicted turning points, including (i) the date when daily increased number of infections begins to decrease or the time at which the second order derivative of the prevalence of infected compartment is zero (i.e. the turning point of infection acceleration to deceleration); and (ii) the date when daily number of removed cases is larger than that of infected cases, or the time at which the first derivative of the prevalence of infected compartment is zero (i.e. the turning point of zero infection speed). As a byproduct, the method also provides a predicted time when the COVID-19 epidemic ends.

This paper is organized as follows. Section 5.2 presents our new epidemiological forecast model incorporating time-varying quarantine protocols. Section 5.3 concerns the algorithmic implementation via Markov Chain Monte Carlo and a deliverable R software. Section 5.4 is devoted to the analysis of COVID-19 data within and outside Hubei. Section 5.5 gives some concluding remarks, and some technical details are included in the appendices.

5.2 State-space SIR Epidemiological Model

5.2.1 Basic Model of Coronavirus Infection

We begin with a basic epidemiological model in the framework of state-space SIR models with no consideration of quarantine protocols. This framework was proposed by *Osthus et al.* (2017) with only one-dimensional time series of infected proportions. Refer to Chapter 9-12 of *Song* (2007) for an introduction to state-space models. Here we consider two time series of proportions of infected and removed cases, denoted by Y_t^I and Y_t^R at time t, respectively, where the compartment of removed R is a sum of the proportions of recovered cases and deaths at time t. We assume that the 2-dimensional time series of $(Y_t^I, Y_t^R)^\top$ follows a state-space model with the beta distributions at time t:

$$Y_t^I|\boldsymbol{\theta}_t, \lambda^I \sim \text{Beta}(\lambda^I\theta_t^I, \lambda^I(1-\theta_t^I)), \tag{5.1}$$

$$Y_t^R|\boldsymbol{\theta}_t, \lambda^R \sim \text{Beta}(\lambda^R\theta_t^I, \lambda^R(1-\theta_t^I)), \tag{5.2}$$

where θ_t^I and θ_t^R are the respective prevalence of infection and removal at time t, and λ^I and λ^R are the parameters controlling the respective variances of the observed proportions (noting that the superscripts here indicate labels rather than exponents).

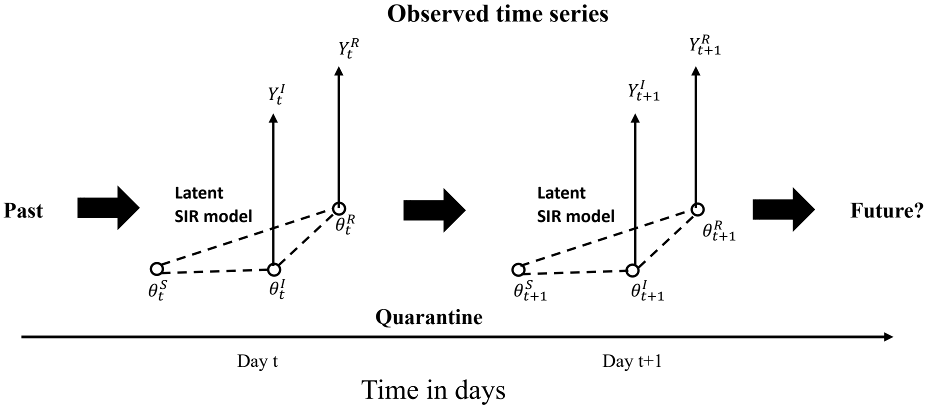


Figure 5.1: A conceptual framework of the proposed epidemiological state-space SIR model.

As shown in Figure 5.1, these observed time series are emitted from the underlying latent dynamics of COVID-19 infection characterized by the latent Markov process $\boldsymbol{\theta}_t$. It is easy to see that the expected proportions in both Equations (5.1) and (5.2) are equal to the prevalence of infection and removal at time t, namely $E(Y_t^I|\boldsymbol{\theta}_t) = \theta_t^I$ and $E(Y_t^R|\boldsymbol{\theta}_t) = \theta_t^R$. See Appendix D.2. Moreover, the latent population prevalence $\boldsymbol{\theta}_t = (\theta_t^S, \theta_t^I, \theta_t^R)^\top$ is a three-dimensional Markov process, in which θ_t^S is the probability of a person being susceptible or at risk at time t, θ_t^I is the probability of a person being infected at time t, and θ_t^R is the probability of a person being removed from the infectious system (either recovered or dead) at time t. Obviously, $\theta_t^S + \theta_t^I + \theta_t^R = 1$.

We assume that this 3-dimensional probability process $\boldsymbol{\theta}_t$ is governed by the following Markov model:

$$\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \boldsymbol{\tau} \sim \text{Dirichlet}(\kappa f(\boldsymbol{\theta}_{t-1}, \beta, \gamma)), \tag{5.3}$$

where parameter κ scales the variance of the Dirichlet distribution and function $f(\cdot)$ is a 3-dimensional vector that determines the mean of the Dirichlet distribution. We have all the relevant parameters be $\boldsymbol{\tau} = (\beta, \gamma, \kappa, \boldsymbol{\theta}_0, \lambda^I, \lambda^R)^\top$, where β and γ denote the transmission and removal rates of the SIR model given in (5.4), and $\boldsymbol{\theta}_0 = (\theta_0^S, \theta_0^I, \theta_0^R)$ are initial prevalence of the three compartments. The function f is the engine of the infection dynamics which operates according to SIR model of the form:

$$\frac{d\theta_t^S}{dt} = -\beta\theta_t^S\theta_t^I, \quad \frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I, \quad \text{and } \frac{d\theta_t^R}{dt} = \gamma\theta_t^I. \tag{5.4}$$

The ratio between the transmission and removal rates is the basic reproduction number $R_0 = \beta/\gamma$ which measures contagiousness or transmissibility of infectious agents. It provides the average secondary cases generated from one infected case when the whole population is susceptible (*Fraser et al.*, 2009; *Delamater et al.*, 2019). Note that the explicit solution to the above system (5.4) of ordinary differential equations is unavailable. Following *Osthus et al.* (2017), we invoke the fourth-order Runge–Kutta (RK4) approximation, resulting in an approximate of $f(\theta_{t-1}, \beta, \gamma)$ as follows:

$$f(\boldsymbol{\theta}_{t-1}, \beta, \gamma) = \begin{pmatrix} \theta_{t-1}^S + 1/6[k_{t-1}^{S_1} + 2k_{t-1}^{S_2} + 2k_{t-1}^{S_3} + k_{t-1}^{S_4}] \\ \theta_{t-1}^I + 1/6[k_{t-1}^{I_1} + 2k_{t-1}^{I_2} + 2k_{t-1}^{I_3} + k_{t-1}^{I_4}] \\ \theta_{t-1}^R + 1/6[k_{t-1}^{R_1} + 2k_{t-1}^{R_2} + 2k_{t-1}^{R_3} + k_{t-1}^{R_4}] \end{pmatrix},$$

where all these k_t terms are given in Appendix D.1. The set of model parameters $\boldsymbol{\tau}$ will be estimated using the MCMC method (*Carlin et al.*, 1992).

5.2.2 Epidemiological Model with Time-varying Transmission Rate

The basic epidemiological model with both constant transmission and removal rates in the SIR model (5.4) does not reflect the reality in China, where various levels of quarantines have been enforced. Many forms of human interventions that are altering the transmission rate over time include (i) individual-level protective measures such as wearing masks and safety glasses, using hygiene, and taking in-home isolation, and (ii) community-level quarantines such as hospitalization for infected cases, city blockade, traffic control and restricted social activities, and so on. In addition, the virus itself may mutate to evolve, which may increase the potential rate of false negative in the disease diagnosis. Thus, the transmission rate β indeed varies over time, which should be accounted in the modeling.

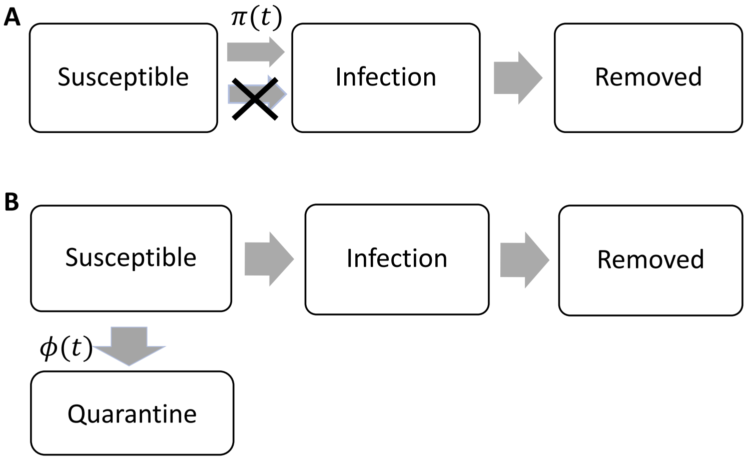


Figure 5.2: Extended SIR models with a time-varying transmission rate modifier $\pi(t)$ (Panel A) or a time-varying quarantine rate $\phi(t)$ (Panel B).

One extension to the above basic epidemiological model is to allow a time-varying probability that a susceptible person meets an infected person or *vice versa*. Suppose at a time t, $q^S(t) \in [0,1]$ is the chance of an at-risk person being in-home isolation, and $q^I(t) \in [0,1]$ is the chance of an infected person being in-hospital quarantine.

Thus, the chance of disease transmission when an at-risk person meets an infected person is modified as:

$$\beta\{1-q^S(t)\}\theta_t^S\{1-q^I(t)\}\theta_t^I := \beta\pi(t)\theta_t^S\theta_t^I,$$

with $\pi(t) := \{1-q^S(t)\}\{1-q^I(t)\}$. In effect, this $\pi(t)$ modifies the chance of a susceptible person meeting with an infected person or *vice versa*, which is termed as a *transmission modifier* due to quarantine in this paper. Obviously, with no quarantine in place, $\pi(t) \equiv 1$ for all time. See Figure 5.2 Panel A. This results in a new SIR model with a time-varying transmission rate modifier:

$$\frac{d\theta_t^S}{dt} = -\beta\boldsymbol{\pi(t)}\theta_t^S\theta_t^I, \quad \frac{d\theta_t^I}{dt} = \beta\boldsymbol{\pi(t)}\theta_t^S\theta_t^I - \gamma\theta_t^I, \quad \text{and } \frac{d\theta_t^R}{dt} = \gamma\theta_t^I, \tag{5.5}$$

where the product term $\beta\pi(t)$ defines an effective transmission rate reflective to the currently enforced quarantine measures of all levels in China. Note that the above extended SIR model assumes primarily that both population-level chance of being susceptible and population-level chance of being infected remain the same, but the chance of a susceptible person meeting with an infected person is reduced via $\pi(t)$.

The transmission rate modifier $\pi(t)$ needs to be specified according to actual quarantine protocols in a given region. A possible choice of $\pi(t)$ may be a step function that reflects government-initiated macro isolation measures in Wuhan, Hubei province:

$$\pi(t) = \begin{cases} \pi_{01}, & \text{if } t \leq \text{Jan } 23, \text{ no concrete quarantine protocols;} \\ \pi_{02}, & \text{if } t \in (\text{Jan } 23, \text{Feb } 4], \text{ city blockade;} \\ \pi_{03}, & \text{if } t \in (\text{Feb } 4, \text{Feb } 8], \text{ enhanced quarantine;} \\ \pi_{04}, & \text{if } t > \text{Feb } 8, \text{ opening of new hospitals.} \end{cases}$$

When $\boldsymbol{\pi}_0 = (\pi_{01}, \pi_{02}, \pi_{03}, \pi_{04})$ are chosen with different values, as shown in Figure 5.3 Panels A-C, we obtain different types of transmission rate modifiers aligned with different quarantine protocols.

Alternatively, the modifier $\pi(t)$ may be specified as a continuous function that reflects steadily increased community-level awareness and responsibility of voluntary quarantine and preventive measures, which may be regarded as a kind of micro isolation measure initiated by individuals or local self-organized inspections. For example, as shown in Figure 5.3 Panels D-F, we may choose the following exponential functions:

$$\pi(t) = \exp(-\lambda_0 t) \text{ or } \pi(t) = \exp\{-(\lambda_0 t)^\nu\}, \lambda_0 > 0, \nu > 0.$$

Technically, the RK's approximate of f function in Appendix D.1 may be easily obtained by replacing β by $\beta\pi(t)$ in the specification of the latent prevalence model (5.3), and moreover in all quantities and steps in the MCMC implementation. See the detailed in Section 5.3.

5.2.3 Epidemiological Model with Quarantine Compartment

An alternative way to incorporate quarantine measures into the basic epidemiological model (5.4) is to add a new quarantine compartment that collects quarantined individuals who would have no chance of meeting any infected individuals in the infection system, as shown in Figure 5.2 Panel B. This model allows to characterize time-varying proportions of susceptible cases due largely to the government-enforced stringent in-home isolation outside of Hubei province. The basic SIR model in equation (5.4) is then extended by adding a quarantine compartment with a time-varying rate of quarantine $\phi(t)$, which is the chance of a susceptible person being willing to take in-home isolation at time t. The extended SIR takes the following 4-dimensional latent process $(\theta_t^S, \theta_t^Q, \theta_t^I, \theta_t^R)^\top$:

$$\begin{aligned} &\frac{d\theta_t^Q}{dt} = \boldsymbol{\phi(t)}\theta_t^S, \ \frac{d\theta_t^S}{dt} = -\beta\theta_t^S\theta_t^I - \boldsymbol{\phi(t)}\theta_t^S, \\ &\frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I, \ \frac{d\theta_t^R}{dt} = \gamma\theta_t^I, \end{aligned} \tag{5.6}$$

where $\theta_t^S + \theta_t^Q + \theta_t^I + \theta_t^R = 1$.

We suppose that the quarantine rate $\phi(t)$ is a Dirac delta function with jumps at times when major macro quarantine measures are enforced. For example, we may specify the $\phi(t)$ function as follows:

$$\phi(t) = \begin{cases} \phi_{01}, & \text{if } t = \text{Jan 23, city blockade;} \\ \phi_{02}, & \text{if } t = \text{Feb 4, enhanced quarantine;} \\ \phi_{03}, & \text{if } t = \text{Feb 8, opening of new hospitals;} \\ 0, & \text{otherwise.} \end{cases}$$

Here we show several examples of multi-point instantaneous quarantine rates in Figure 5.3 Panels G-H with jump sizes equal to $\boldsymbol{\phi}_0 = (\phi_{01}, \phi_{02}, \phi_{03})$ that occur respectively at dates of (Jan 23, Feb 4, Feb 8). In particular, we plot three scenarios, e.g., no intervention ($\boldsymbol{\phi}_0 = (0,0,0)$), multiple moderate jumps ($\boldsymbol{\phi}_0 = (0.1, 0.4, 0.3)$), and only one large jump ($\boldsymbol{\phi}_0 = (0, 0.9, 0)$). Note that at each jump, the respective proportion of the susceptible population would move to the quarantine compartment. For example, with $\boldsymbol{\phi}_0 = (0.1, 0.4, 0.3)$, the quarantine compartment will be enlarged accumulatively over three time points as $0.1\,\theta_{t_1}^S + 0.4\theta_{t_2}^S + 0.3\theta_{t_3}^S$.

The $f(\boldsymbol{\theta}_{t-1}, \beta, \gamma)$ function determined by the above extended SIR model (5.6) can be solved by applying the fourth-order Runge-Kutta approximation, and the resulting solution is given in Appendix D.1. To deal with the Dirac delta function $\phi(t)$, we develop a two-step approximation for model (5.6). In brief, we first solve a continuous function without change points via the differential equations in (5.5), and then we directly move a mass of $\phi(t)\theta_t^S$ out of the susceptible compartment to the quarantine compartment. From our experience, this approach largely improves the approximation accuracy in the presence of discontinuities.

5.3 Implementation: Markov Chain Monte Carlo Algorithm

5.3.1 MCMC Algorithm

We implemented the MCMC algorithm to collect draws from the posterior distributions, and further obtain posterior estimates and credible intervals of the unknown parameters in the above models specified in Section 5.2. Because of the hierarchical structure in the state-space model considered in this paper, the posterior distributions can be obtained straightforwardly. The R package `rjags` (*Plummer*, 2019) can be directly applied to draw samples from the posterior distributions, so we omit the technical details. The latent Markov processes $\boldsymbol{\theta}_t$ are sampled and forecasted by the MCMC sampler, particularly for the prevalence of infection and the probability of removal, θ_t^I and θ_t^R, which enables us to determine the turning points of interest and the reproduction number R_0.

The first turning point of interest is the time when the daily number of new infected cases stops increasing. Mathematically, this corresponds to the time t at which $\ddot{\theta}_t^I = 0$ or the gradient of $\dot{\theta}_t^I$ is zero. As illustrated by Panel A in Figure 5.4, the peak of $\dot{\theta}_t^I$, denoted by the vertical green line, is the first turning point of interest. The second turning point is the time when the cumulative infected cases reaches its maximum, meaning $\dot{\theta}_t^I = 0$. In principle, the second turning point occurs only after the first one, as shown in Panel B in Figure 5.4.

The basic reproduction number R_0 of an infectious disease is estimated by the ratio $R_0 = \beta/\gamma$, where β and γ are both estimated from their posterior distributions. Because our models consider the quarantine compartment, R_0 might change according to the forms of quarantine protocols. We adopt a standard MCMC algorithm to draw samples of the latent process. Let t_0 be the current time up to which we have observed data $(Y^I_{0:t_0}, Y^R_{0:t_0})$. To perform M draws of Y^I_t, Y^R_t for $t \in [t_0+1, T]$, we proceed as follows: for each $m = 1, \dots, M$,

(1) draw $\boldsymbol{\theta}^{(m)}_t$ from the posterior $[\boldsymbol{\theta}_t | \boldsymbol{\theta}^{(m)}_{t-1}, \boldsymbol{\tau}^{(m)}]$ of the prevalence process, at $t = t_0 + 1, \dots, T$;

(2) draw $(Y^{I(m)}_t, Y^{R(m)}_t)$ from $[Y^I_t | \boldsymbol{\theta}^{(m)}_t, \boldsymbol{\tau}^{(m)}]$ and $[Y^R_t | \boldsymbol{\theta}^{(m)}_t, \boldsymbol{\tau}^{(m)}]$ according to the observed process, at $t = t_0 + 1, \dots, T$, respectively;

The prior distributions are specified with some of the hyper-parameters being set according to the SARS data from Hong Kong (*Mkhatshwa and Mummert*, 2010). They are,

$$\boldsymbol{\theta}_0 \sim \text{Dirichlet}(1 - Y^I_1 - Y^R_1, Y^I_1, Y^R_1)$$

$$R_0 \sim \text{LogN}(1.099, 0.096) \text{ with } \text{E}(R_0) = 3.15, \text{SD}(R_0) = 1;$$

$$\gamma \sim \text{LogN}(-2.955, 0.910) \text{ with } \text{E}(\gamma) = 0.0821, \text{SD}(\gamma) = 0.1,\ \beta = R_0\gamma;$$

$$\kappa \sim \text{Gamma}(2, 0.0001),\ \lambda^I \sim \text{Gamma}(2, 0.0001),\ \lambda^R \sim \text{Gamma}(2, 0.0001).$$

Note that LogN and Gamma stand for log-normal and gamma distributions respectively, and E and SD represent mean and standard deviation here. In the default setting the variances of the above prior distributions are set at relatively large values to reflect the fact that limited prior knowledge of these epidemiological model parameters is available. When more information becomes accessible during the course of the epidemic, smaller prior variance values may be used, leading to tighter credible intervals for the model parameters and turning points.

5.3.2 R Software Package

We deliver an R software package that is able to output the MCMC estimation, inference and prediction under the epidemiological model with two proposed extended SIR models that incorporate time-varying quarantine protocols. These new models have been discussed in detail in Sections 5.2.2 and 5.2.3. Our R package, named `eSIR`, uses daily-updated time series of infected and removed proportions as input data. The R package is available at GitHub `lilywang1988/eSIR`, and its user manual is appended as the supplementary material of this paper. The quarantine functions are predefined and will be treated as known functions of protocols/policies in the estimation and prediction steps. We let the transmission rate modifier $\pi(t)$ be either a step function or an exponential function, and let the quarantine rate $\phi(t)$ follow a Dirac delta function with pre-specified points of jump and sizes of jumps. The package provides various plots for users to visualize the MCMC results, including the estimated prevalence of infection and the estimated probability of removal, and predicted turning points of interest. Various summary statistics are listed in the output, including posterior mean estimates of the transmission and removal rates, estimate of the reproduction number, and forecasts of turning points and their 95% credible intervals. Moreover, the package gives multiple options to users who can save the entire MCMC results, including the output tables and summary plots, Gelman-Rubin convergence statistic, traceplots for MCMC quality control, and full MCMC draws for user's own summary analyses. Some illustrations on the use of this software package are given in Section 5.4 with sample codes in Appendix D.3. In addition, we developed an online R Shiny App that can automatically update the results whenever the China CDC updates the daily COVID-19 data (*Kleinsasser et al.*, 2020).

5.4 Analysis of the COVID-19 Data

We applied our proposed models, algorithms and R package `eSIR` to analyze the COVID-19 data collected from the public website *DXY.cn* (2020). The earliest public records for the provincial data are available since Jan 20, 2020. According to an existing R package on GitHub `GuangchuangYu/nCov2019` (*Yu*, 2020), the total counts of confirmed infections and deaths are dated back on Jan 13, 2020. We assumed that before Jan 17 all the reported cases and deaths were from Hubei. We imputed by the linear interpolation the missing cases on Jan 18-19. Therefore, the data used in our analyses starts from Jan 13. The data used in analyses for the other provinces starts on Jan 23, which is the earliest time with non-zero values in the removed compartment. Note that there exist some minor discrepancies between different data sources. This section aims to provide a demonstration of our software to analyze the current public COVID-19 data, through which users may understand the proposed methods. We will also elaborate ways to export and interpret the MCMC results. The R package may be applied to analyze infectious data from other countries.

First, we show the analysis of the Hubei COVID-19 data after introducing in a time-verying transmission rate modifier $\pi(t)$ using our R function `txt.eSIR` in the package `eSIR`. The corresponding results are shown in Figure 5.5: Columns B-C represent estimation and forecasting results of a transmission rate following an exponential rate modifier with rate $\lambda_0 = 0.05$ (Panel B of Figure 5.3) and a step function with $\boldsymbol{\pi}_0 = c(1, 0.9, 0.5, 0.1)$ at change points [Jan 23, Feb 4, Feb 8] (Panel C of Figure 5.3), as opposed to a basic model of $\pi(t) \equiv 1$ in Column A. Running R codes were given as Examples 1-3 in Appendix D.3. The forecast plots for infection and removal compartments are presented in Row 1 and Row 3 respectively, with all the black dots left to the blue vertical line denoting observed proportions by the last observational date. That is, the blue vertical marks time t_0 as defined in Section 5.3. The green and purple vertical lines denote the first and second turning points, respectively.

The salmon color area denotes the 95% credible interval of the predicted proportions $[Y^I_{(t_0+1):T}|Y^I_{1:t_0}, Y^R_{1:t_0}]$ and $[Y^R_{(t_0+1):T}|Y^I_{1:t_0}, Y^R_{1:t_0}]$ after t_0, respectively, while the cyan color area represents either the 95% credible intervals of the prevalence $[\theta^I_{1:t_0}|Y^I_{1:t_0}, Y^R_{1:t_0}]$ or that of the probability of removal $[\theta^R_{1:t_0}|Y^I_{1:t_0}, Y^R_{1:t_0}]$ prior to time t_0. The gray and red curves are the posterior mean and median curves. The black curve in the removal compartment plots from Row 3 denotes the estimated proportion of deaths computed based on a pre-specified ratio (`death_in_R`). Row 2 provide a series of important dynamic features of the infection via a spaghetti plot, in which 20 randomly selected MCMC draws of the first-order derivative of the posterior prevalence of infection, namely $\dot{\theta}^I_t$. The black curve is the posterior mean of the derivative, and the vertical lines mark times of turning points corresponding respectively to those shown in Row 1 and Row 3. Moreover, the 95% credible intervals of these turning points are also highlighted by semi-transparent rectangles in Panel B and summarized in Web Table 1. In Subfigures A-C we displayed the results for time-dependent transmission rate modifiers. One can see that $\pi(t)$ plays an important roles in shortening the key turning points of the epidemic, and its effect on both estimation and prediction of the COVID-19 infection dynamics has been clearly demonstrated. Note that there exists an abrupt jump on Feb 12, which is believed to be mainly caused by the under-testing and under-reporting before that date. This kind of under-reporting data issue can be calibrated using an algorithm, which assumes an exponential increase at the early stage of the epidemic, as proposed in Section 4.1 of *Wang et al.* (2020c).

Next, we analyzed the data from the rest of the Chinese population (i.e. the provinces outside Hubei) starting on Jan 23. We included two change points for the step function $\pi(t)$ at [Feb 4, Feb 8] with $\boldsymbol{\pi}_0 = (0.8, 0.1)$. The exponential function remained the same. It is noted that the spread of COVID-19 outside Hubei has been so far much less severe. Possible reasons for such low proportions of infection and deaths include (i) discontinuing the traffic connections between Hubei and the other provinces, (ii) more timely caution and preventative measures taken, and (iii) a comparatively less dense distribution of infection with respect to the huge population size. When Panel A1 in Figure 5.6 is zoomed in, some of the observed proportions (black dots) are deviated from the posterior mean or median of the fitted prevalence albeit they all fall in the 95% credible intervals, as shown by Panels B1 and C1 in Figure 5.6. Since the latent process follows the SIR differential equations, there may be a lack of fit for the SIR model to accommodate a very large and complex population of 1.3 billion people, in which most of the subjects are not at risk. The proposed models should work much better for individual provinces.

The other epidemiological model with an added quarantine compartment as an absorbing state was fitted via our R function `qh.eSIR` in the package `eSIR`. We applied the proposed model in analyses of the data within and outside Hubei following Dirac delta functions with jumps of $\boldsymbol{\phi}_0 = [0.1, 0.9, 0.5]$ at change points [Jan 23, Feb 4, Feb 8] and $\boldsymbol{\phi}_0 = [0.9, 0, 5]$ at change points [Feb 4, Feb 8] respectively. Their results were summarized in Column D of Figures 5.5 and 5.6. Their running codes were given as Examples 4-5 in Appendix D.3. Our analyses once again clearly indicated that stringent quarantine protocols can largely reduce the spread of COVID-19 both within Hubei and outside Hubei. Yet, it is known that too strict quarantine can backfire; people may lose their trust and patience in their committed system, and consequently may try to reduce compliance or even avoid quarantine. We also present the posterior mean probability of staying quarantine compartment in Figure 5.7 within Hubei and outside Hubei. Note that Jan 23 was not set as a change point for the cases outside Hubei, leading only to two jumps. It is evident that by Feb 8, more than 90% of the Chinese population have taken in-home isolation or as such, reflective to a very strict quarantine protocol enforced in the entire country.

The reproduction numbers estimated from different models using data within and outside Hubei together with their 95% credible intervals are summarized in Table 5.1.

It is worth pointing out that the estimates of the basic reproduction numbers obtained from the epidemiological models with time-varying transmission or quarantine rates appear larger than those obtained from the basic model with no quarantine. This is not surprising as our new models explicitly incorporate interventions, so that the estimated R_0 is an adjusted number with the influence of interventions be removed. In contrast, the basic model with no use of the quarantine modifier implicitly integrates the effect of interventions into the transmission rate β, and consequently the estimated R_0 is reduced due to the contribution from interventions. Our analyses suggest that reproduction numbers R_0 of COVID-19 without public health interventions would be around 4-6 within Hubei and around 3-3.5 outside Hubei with relatively big credible intervals. As pointed out above, the size of credible interval may be reduced with more accessible data of COVID-19, which permits users to specify smaller variances in the prior distributions given in section 5.3.1.

Table 5.1: The posterior mean and credible intervals of the reproduction number R_0 obtained from different quarantine models and datasets.

	Within Hubei		Outside Hubei	
Model	Mean	95%CI	Mean	95%CI
No quarantine	2.98	[1.90, 4.44]	2.56	[1.50, 4.22]
Exponential	6.34	[2.82, 10.80]	3.16	[1.80, 5.06]
Step-function	4.61	[2.12, 8.16]	2.90	[1.65, 4.76]
Quar. Compart.	4.14	[1.96, 8.08]	3.37	[1.77,5.73]

As pointed out by quite a few reviewers and users of this toolbox that the estimated reproduction numbers R_0 is dependent on the prespecified intervention assumptions, e.g. the function $\pi(t)$. The form of $\pi(t)$ can be specified mainly in two ways. One is to let $\pi(t)$ be a parametric function (e.g., exponential decaying function) and estimate it via regular MCMC, and the other is to estimate the $\pi(t)$ function non-parametrically prior to being passed into the proposed Bayesian state-space model. For the latter, usually the proportion of cumulative infected cases is very small so that $S_t/N \approx 1$, hence one can repeatedly fit the a linear model to estimate the time-dependent function $\pi(t)$ according to *Sun et al.* (2020). The nonparametric estimate of $\pi(t)$ is given by Figure 5.9, and the dashed curve after the last observation date denotes the predicted trend of $\pi(t)$ based on the previously estimated curve (solid line). The estimated transmission rate (posterior mean) is $\widehat{\beta}_0 = 0.123$ (95% CI: [0.0422, 0.256]), the removal rate is $\widehat{\gamma} = 0.0257$ (95% CI: [0.0144, 0.0389]), and thus the basic reproduction number is $\mathcal{R}_0 = 4.71$ (95% CI: [2.20, 8.60]).

Since the turning points in China have been observed by Feb 23, there is an increasing concern about whether and when there would be a second outbreak. We conducted another set of analyses on Hubei calibrated data to forecast the epidemic trends when strict intervention may not last long. We focused on different degrees of relaxation on the intervention. In particular, we added Feb 24 to the step function $\pi(t)$ so that it has change points [Jan 23, Feb 4, Feb 8, Feb 24] with $\boldsymbol{\pi}_0 = (1, 0.9, 0.5, 0.1, \pi_{05})$. Note that in our fitted data, Feb 23 is the last observational date. We considered π_{05} equal to 0.1, 0.3 and 0.5 to describe "strictly continuing", "slightly loosening" or "moderately loosening" the control actions that has made the transmission rate 0.1β since Feb 8. Our results in Figure 5.5 and Web Table 2 indicate that, on average, increasing the transmission rate from 0.1β to 0.5β would end up with a second outbreak with a maximum prevalence 7.5% and totally 16.7% of the population affected by July 20, increasing from 0.1β to 0.3β would end up with a gradual increase in prevalence to 0.6% and about 1.4% of the population being affected. If we continue keeping the transmission rate to be 0.1β, however, the epidemic will eventually vanish in the population with no second outbreak and in total about 0.1% of the population being affected. All these three scenarios are much better than the one without any intervention (Panel A1 in Figure 5.5).

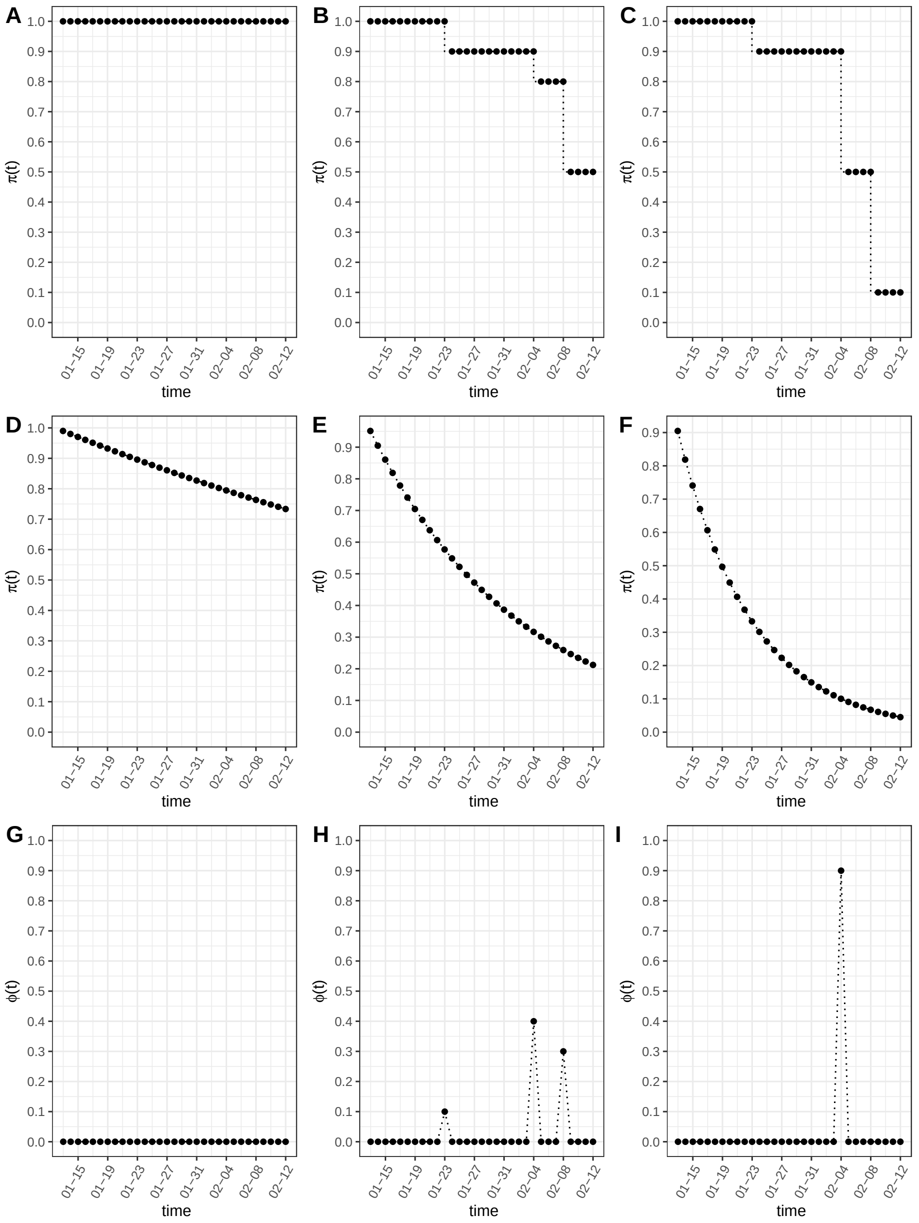
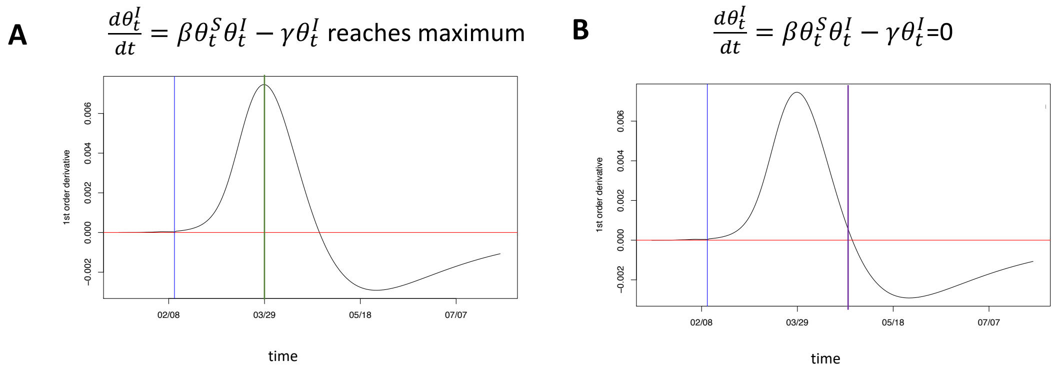


Figure 5.3: The functional forms of the transmission rate modifiers $\pi(t)$ and the quarantine rate $\phi(t)$: 1) Panels A-C depict step functions with $\boldsymbol{\pi}_0 = (\pi_{01}, \pi_{02}, \pi_{03}, \pi_{04})$ equal to $(1, 1, 1, 1)$, $(1, 0.9, 0.8, 0.5)$ and $(1, 0.9, 0.5, 0.1)$ at change points (Jan 23, Feb 4, Feb 8), Panels D-F depict exponential functions under difference micro quarantine measures over time with $\lambda_0 = 0.01$, $\lambda_0 = 0.05$ and $\lambda_0 = 0.1$, and 3) Panels G-I depict multi-point instantaneous quarantine rates with $\boldsymbol{\phi}_0 = (0, 0, 0, 0)$, $\boldsymbol{\phi}_0 = (0.1, 0.4, 0.3)$ and $\boldsymbol{\phi}_0 = (0, 0.9, 0)$ at change points of (Jan 23, Feb 4, Feb 8).



A $\frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I$ reaches maximum

B $\frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I$=0

Figure 5.4: The first turning point in Panel A is marked by a green line, denoting the time when the estimated first-order derivative of the prevalence of infection reaches the maximum. The second turning point in Panel B is marked by a purple line, which is the time when the estimated first-order derivative of the prevalence of infection equals to zero. The vertical blue line labels the last observation day.

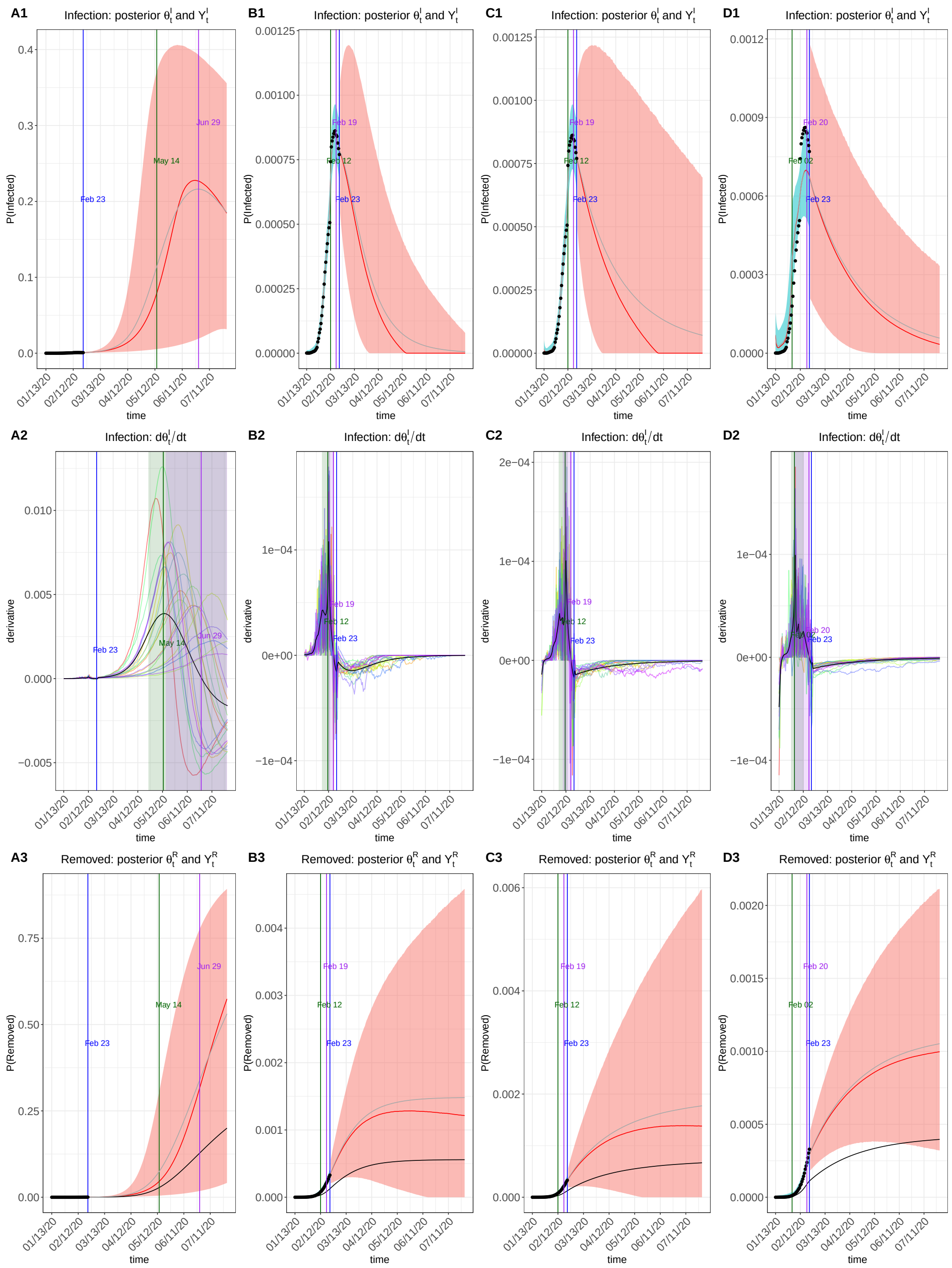


Figure 5.5: Prediction plots of θ_t^I and Y_t^I (Row 1), $\dot{\theta}_t^I$ (Row 2), θ_t^R and Y_t^R (Row 3) for Hubei Province. Subfigures in Column A display the results of basic SIR model with $\pi(t) \equiv 1$ or $\phi(t) \equiv 0$, Subfigures in Column B display results of a continuous transmission modifier $\pi(t) = \exp(-0.05t)$, subfigures in Column C display results of a step-function transmission rate modifier with $\boldsymbol{\pi}_0 = (1, 0.9, 0.5, 0.1)$ at change points [Jan 23, Feb 4, Feb 8], and subfigures in Column D display results of a Dirac delta function quarantine process with $\boldsymbol{\phi}_0 = [0.1, 0.9, 0.5]$ at change points [Jan 23, Feb 4, Feb 8].

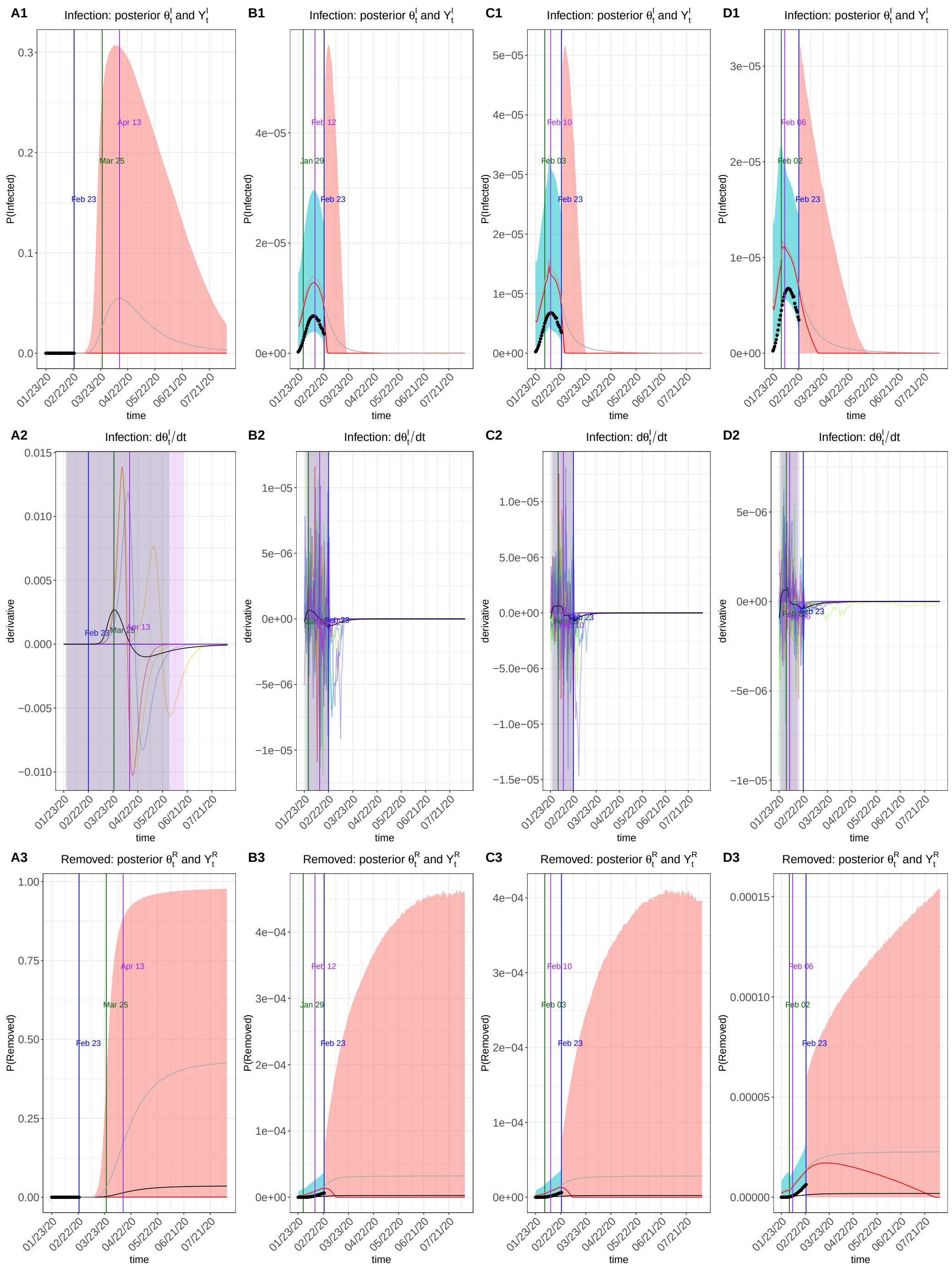


Figure 5.6: Prediction plots of θ_t^I and Y_t^I (Row 1), $\dot{\theta}_t^I$ (Row 2), θ_t^R and Y_t^R (Row 3) for the Chinese population outside Hubei Province. Subfigures in Column A display the results of basic SIR model with $\pi(t) \equiv 1$ or $\phi(t) \equiv 0$, Subfigures in Column B display results of a continuous transmission modifier $\pi(t) = \exp(-0.05t)$, subfigures in Column C display results of a step-function transmission rate modifier with $\boldsymbol{\pi}_0 = (1, 0.9, 0.5, 0.1)$ at change points [Jan 23, Feb 4, Feb 8], and subfigures in Column D display results of a Dirac delta function quarantine process with $\boldsymbol{\phi}_0 = [0.1, 0.9, 0.5]$ at change points [Jan 23, Feb 4, Feb 8].

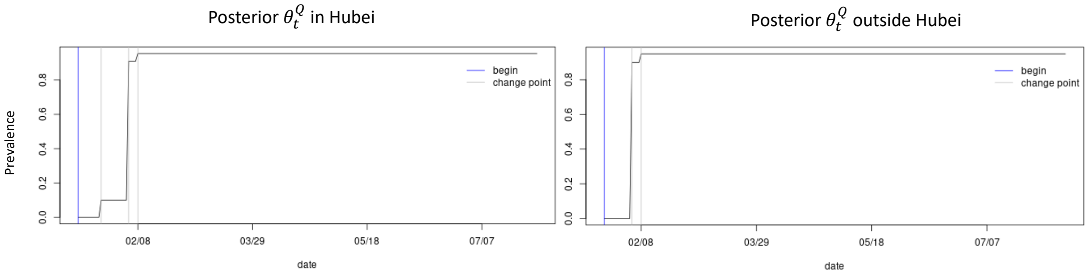


Figure 5.7: The estimated probability of staying in quarantine within and outside Hubei.

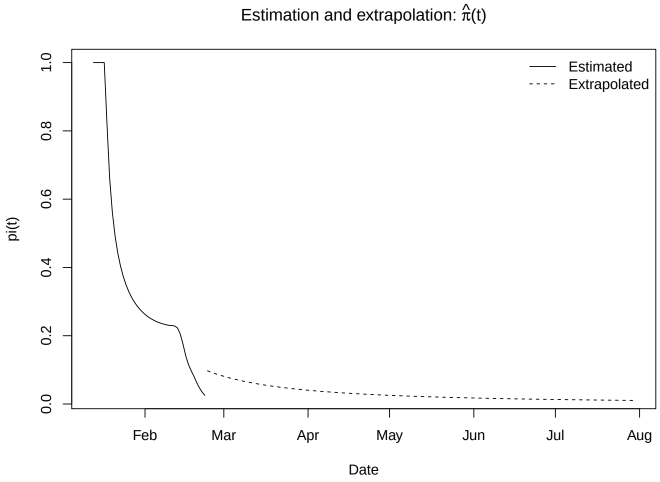


Figure 5.8: The estimated transmission rate modifiers $\widehat{\pi}(t)$. The dashed line is the predicted trend, and the solid line is the estimated curve based on the observed data.

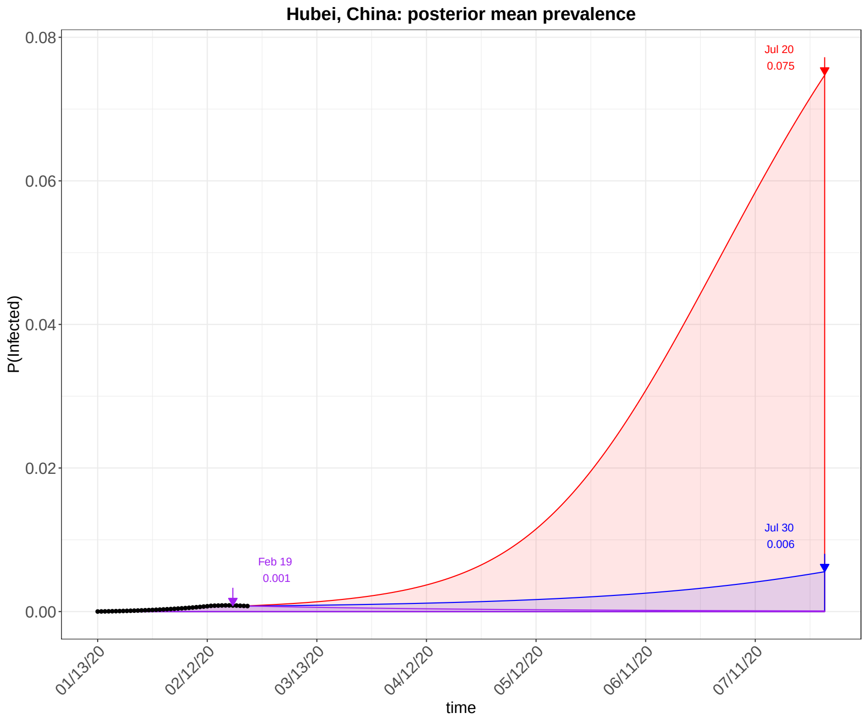


Figure 5.9: Predicted mean prevalence of infection with or without loosening the strict intervention in Hubei. The red semitransparent area denotes the scenario of moderate relaxation of the strict human intervention ($\pi_{05} = 0.5$), the blue area denotes the slight relaxation of intervention ($\pi_{05} = 0.3$), and the purple area denotes the scenario that stringent control is continued ($\pi_{05} = 0.1$). All their corresponding arrows mark the dates of their maximum mean prevalence.

Table 5.2: The summary table of turning points for forecast results in Figures 5.5 and 5.6. Note that TP1 denotes the first turning point with the largest daily increment in the prevalence of infection, and TP2 denotes the second turning point with the maximum prevalence of infection. The last forecast date is July 20 for Hubei and July 30 for other provinces.

		TP1		TP2	
Location	Model	Mean	95%CI	Mean	95%CI
Hubei	No quarantine	May 14	[April 26, July 30]	June 29	[May 17,July 30]
	Exponential	Feb 12	[Feb 5, Feb 15]	Feb 19	[Feb 13, Feb 20]
	Step-function	Feb 12	[Feb 4, Feb 15]	Feb 19	[Feb 9, Feb 20]
	Quar. Compart.	Feb 2	[Jan 29, Feb 13]	Feb 20	[Feb 2, Feb 21]
Others	No quarantine	Mar 25	[Jan 26, May 31]	April 13	[Jan 28, June 18]
	Exponential	Jan 29	[Jan 25, Feb 22]	Feb 12	[Jan 27, Feb 23]
	Step-function	Feb 3	[Jan 25, Feb 22]	Feb 10	[Jan 27, Feb 23]
	Quar. Compart.	Feb 2	[Jan 25, Feb 16]	Feb 6	[Jan 26, Feb 18]

Table 5.3: The summary table for the second outbreak forecast in Hubei with or without relaxation of the human intervention. We used step function $\pi(t)$ as transmission rate modifier with $\boldsymbol{\pi}_0 = (1, 0.9, 0.5, 0.1, \pi_{05})$ at change points [Jan 23, Feb 4, Feb 8, Feb 24]. We considered π_{05} equal 0.1, 0.3 and 0.5 for "strictly continuing", "slightly loosening" and "moderately loosening" the previous control actions, and recorded their maximum prevalence of infection and cumulative infection proportions as well as their 95% credible intervals. The last forecast date is July 20.

	Maximum Prevalence (%)			Cumulative infection (%)	
π_{05}	Date	Mean	95%CI	Mean	95%CI
0.1	Feb 19	0.08	[0.07, 0.10]	0.13	[0, 0.43]
0.3	July 20	0.55	[0, 3.82]	1.44	[0.02, 8.23]
0.5	July 20	7.47	[0.01, 30.12]	16.67	[0.18, 78.68]

5.5 Concluding Remarks

We develop an epidemiological forecast model with an R software package to assess effects of interventions on the COVID-19 epidemic within Hubei and outside Hubei in China. Since our proposed model utilizes the strength of the SIR's dynamic system to capture the primary mechanism of the COVID-19 infectious disease, we are able to generate potential predictions of future episodes of the disease spread patterns over a prespecified window from the last date of data availability. Some turning points of interest are obtained from these forecasting curves as part of the deliverable information, including the predicted time when daily proportion of infected cases becomes smaller than the previous ones and the predicted time when daily proportion of removed cases (i.e. both recovered and dead) becomes larger than that of infected cases, as well as the time when the epidemic ends. Our informatics toolbox provides quantification of uncertainty on the prediction, rather than only point prediction values, which are valuable to see the best versus the worst. The key novel contribution is the incorporation of time-varying quarantine protocols to expand the basic epidemiological model to accommodate changing transmission rates over time in the population. The toolbox can be used by practitioners who have better knowledge of quarantine and better quality data to perform their own analyses. Practitioners can use the toolbox to evaluate different types of quarantine strategies in practice. All summary statistics obtained from the toolbox are of great importance for public health workers and government policy makers to take proper actions on stop spreading of emerging epidemics, such as the COVID-19 epidemic examined here.

We choose the MCMC algorithm to implement the statistical estimation and prediction because of the consideration on the prediction uncertainty. Given the considerable complexity in the COVID-19 virus spread dynamics and potentially inaccurate information of quarantine measures as well as likely under-reported proportions of infected and recovered cases and deaths, it is of critical importance to quantify and report uncertainty in the forecast. Note that the publicly reported data of recovery and death of COVID-19 are mostly collected from hospitals where accessibility to such information is warranted. In contrast, it is very difficult, if not impossible, to collect the data of infected individuals with light symptoms who had in-home isolation and recovered, in spite of serious efforts made by the government for a door-to-door inspection to identify suspected cases.

This toolbox is indeed so general that it may be applicable to analyze and evaluate the COVID-19 epidemic in other countries, as well as the future outbreak of other types of infectious diseases. As noted in the paper, our proposed method does need some existing data of similar infectious disease to set up hyper-parameters in the prior distributions of the model parameters to begin the MCMC. For this, we used the epidemic parameters of the SARS outbreak in Hong Kong given some similarity of COVID-19 to SARS. From this perspective, what we learned from this COVID-19 epidemic in this paper is extremely valuable to form initial conditions in the analysis of any future outbreak of similar infectious disease. In addition, understanding forms and strengths of quarantines for the controlling of disease spread is an inevitable path to making effective preventive policies, which is the key analytic capacity that our toolbox offers.

The proposed approach is extremely useful for policy decision makers to conduct interventions forecast. Our analyses have shown that implementing strict intervention can well control the spread of COVID-19 in China. Moreover, continuing relatively strict intervention can help avoid a second outbreak. Though a slight to moderate relaxation on the intervention will lead to increased infection among the population, an interval of stringent control will still largely delay the progression of pandemic and reduce the maximum prevalence, or "flatten" the infection curves. A flattened infection curve means more preparation time and fewer infectious cases at each critical moment, hence more lives can be saved.

The proposed method has several limitations. First, it ignores the compartment of exposure; it is known that incubation period is relevant to disease transmission, which is particularly true for the COVID-19 as asymptomatic individuals are infectious. Second, the number of removed cases may be inaccurate due to the fact that many of deaths occurring outside of hospitals may not be diagnosed for the COVID-19 infection. Third, it assumes that the recovered cases are automatically immune to the coronavirus, which has not been clinically validated yet.

This analysis also has several limitations. Firstly, this analysis used an underlying SIR model structure, which is fairly simple—there are a number of additional processes that are known to be involved in the natural history of COVID-19 and could potentially be incorporated into the model. For example, the incubation period is known to be approximately a median of 5 days (*Lauer et al.*, 2020), which could be incorporated into the model. Similarly, age structure, potential superspreading events, asymptomatic infections and variation in transmissibility across individuals, and more complex contact patterns (e.g. accounting for spatial structure when examining larger-scale dynamics such as across the whole country) could all play a potentially important role in the epidemic dynamics, altering the predictions of the model. Further, the model does not explicitly account for the underreporting fraction or how it may change over time, which can affect predictions and forecasts (*Gamado et al.*, 2017, 2014; *Eisenberg et al.*, 2015). Future work to account for more complex dynamics and incorporate these features into the package will be useful, both for model comparison and for extending the model to new contexts and diseases.

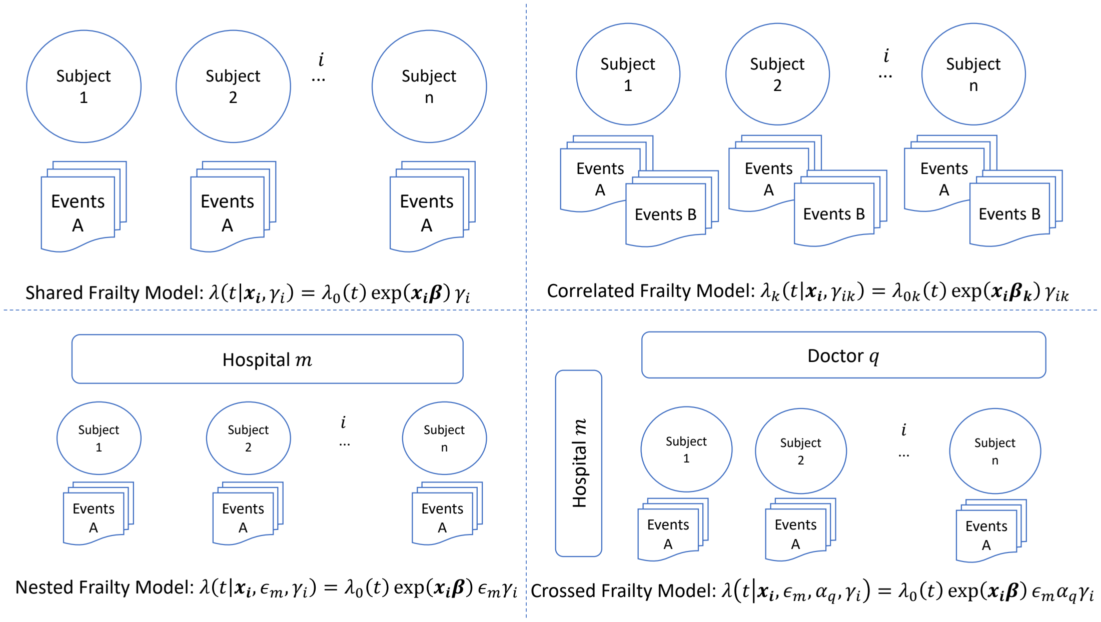
A second important future direction for this work is the validation of the predictions made by the model using subsequent data, such as cross-validating the model using data across different countries given that the COVID-19 has become a global pandemic. To fully evaluate the usefulness of this approach, it will be important to compare the model predictions to the actual trajectory of the epidemic—either for COVID-19 or for other epidemics, e.g. as a hindcasting exercise. This is an important next step for this approach to be used as a forecasting tool in public health practice.

Additionally, the proposed epidemiological models can be further extended to accommodate more data reported by the China CDC, which are worth future exploration. Two types of data that may be used in the future extension are the daily number of suspected cases and the daily number of hospitalized cases. We did not use such data due to the concern of data accuracy. For example, the number of suspected cases is largely dependent on the diagnostic protocols, which have been revised a few times since the outbreak of the disease, and the sensitivity of the viral test. Given such concerns, our strategy in the proposed model was to only use necessary data for analysis, and over the course of improved data quality in the near future, our toolbox may be extended to enjoy greater statistical power and more accurate predictions.

CHAPTER VI

Summary and Future Work

Clustered events and multiple events from the same subject are examples of multivariate failure time data commonly seen in clinical researches. In Chapters II-IV, I studied and explored three different frameworks or methods to analyze various types of associated event data. In Figure 6.1, I summarize four types of frailty/mixed models and their corresponding clustering structures: shared frailty models for a single level of clustering; correlated frailty models for multiple event types; nested frailty; and crossed frailty models for multi-level clustering.



Subject 1
Subject 2
i ...
Subject n
Events A
Shared Frailty Model: $\lambda(t|\boldsymbol{x_i}, \gamma_i) = \lambda_0(t)\exp(\boldsymbol{x_i\beta})\gamma_i$
Events B
Correlated Frailty Model: $\lambda_k(t|\boldsymbol{x_i}, \gamma_{ik}) = \lambda_{0k}(t)\exp(\boldsymbol{x_i\beta_k})\gamma_{ik}$
Hospital m
Nested Frailty Model: $\lambda(t|\boldsymbol{x_i}, \epsilon_m, \gamma_i) = \lambda_0(t)\exp(\boldsymbol{x_i\beta})\epsilon_m\gamma_i$
Doctor q
Crossed Frailty Model: $\lambda(t|\boldsymbol{x_i}, \epsilon_m, \alpha_q, \gamma_i) = \lambda_0(t)\exp(\boldsymbol{x_i\beta})\epsilon_m\alpha_q\gamma_i$

Figure 6.1: Different types of frailty models for various clustering structures.

In Chapter II, I developed methodology applicable to correlated frailty models for two alternating correlated recurrent events. The estimating procedure is a two-step iteration wherein the estimated regression parameters and predicted random effects are obtained via a penalized partial likelihood and the covariance matrix for the correlated frailties from an approximate marginal likelihood. The proposed method can readily accommodate more than two event types without knowing the association directions among the different event types. Moreover, a likelihood ratio test based on the approximate marginal likelihood can evaluate the existence of dependence between the different event types. In the future, I plan to add longitudinal outcomes to jointly model them with other recurrent events of interest, obtaining their covariate effects, dependence structures simultaneously, and thus subject-specific risks. In addition, for small clusters, the possible numerical biases caused by Laplace approximation can also be corrected, as discussed in Section 2.6.

In Chapter III, I developed an estimating equation framework for a flexible class of frailty models based on the moment conditions endowed by a nonstationary Poisson process. The proposed approach can estimate the regression parameters, baseline rates, and the variance components without pre-specifying the distribution of the frailties. The estimation framework can accommodate multiple types of frailty models with a variety of clustering structures, as shown in Figure 1, including the shared frailty model, the correlation frailty model, and the nested frailty model. The proposed framework not only provides unbiased estimation but also requires minimum computational cost. As a possible future extension, I will consider a more complicated scenario that patients are treated at different hospitals and by different doctors by developing a crossed frailty model (Figure 6.1) under the proposed framework. Note that, unlike other nonparametric frailty methods that allow the frailty to follow a discrete distribution or estimate the distribution via kernels or smoothing splines, the proposed framework does not attempt to describe the distribution of the frailty.

Instead, the variance components are obtained through moment-based techniques. In the future, the unbiased estimation of the second moment of the frailties can be further exploited to develop useful diagnostic tools for distribution selection before fitting a parametric frailty model which, on the other hand, tends to provide more efficient estimation.

The objective of Chapter IV is to investigate the association between lead exposure and physical activity performance among children in the ELEMENT study. To incorporate a daily renewal property in the multistate model, I propose a class of multistate rate models with shared baseline rates that are stratified by event type. In particular, within the ELEMENT project, physical activities are transformed into categorical states, and their transitions are treated as multiple events. The proposed multistate rate models borrow the concept of both competing risks and event rate models. The estimation of the baseline rates and regression parameters are grounded upon moment conditions rather than the partial likelihoods. Robust sandwich variance estimators are required to provide valid statistical inference. The model-fitting results reveal that lead exposure is significantly associated with physical activity performance, with unequal effects for boys and girls. It would be of scientific interest to investigate whether attention deficit hyperactivity disorder mediates these effects. Moreover, it would be worthwhile to derive the consistency and asymptotic normality for the proposed multistate rate models.

Chapter V is a special project addressing the global outbreak of the COVID-19 pandemic. In this project, I develop a toolbox for public health practitioners to conduct intervention forecasts for COVID-9 and other epidemics with two extended epidemiological SIR models: one with a time-varying transmission rate and the other with a time-changing quarantine process. Currently, I do not model the exposure compartment; thus, the latency period is not taken into account. Moreover, the current extended models require pre-defining the time-varying functions. In addition, the quarantine process only occurs to the susceptible compartment by unidirectionally moving a proportion of subjects out of that compartment. Future improvements would be adding the quarantine process not only to the susceptible compartment but also the infected compartment, allowing bidirectional changes of the quarantine compartment, estimating the time-varying functions, and adding the exposure compartment. I will also continue maintaining the open-source R package developed as part of this project.

APPENDICES

APPENDIX A

Penalized Survival Models for the Analysis of Alternating Recurrent Event Data

A.1 $\widehat{\boldsymbol{D}}^{\#}$ is Positive-definite

Fixing $\boldsymbol{D}$, a partial log-likelihood (PLL) is assumed to be concave with respect to $\boldsymbol{\gamma}$, or in other words, $-(\partial^2 PLL)/(\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}')$ is positive-definite. Variance matrix $\boldsymbol{\Sigma}$ and thus its inverse $\boldsymbol{\Sigma}^{-1}$ are positive-definite. $\boldsymbol{K}_{PPL2}(\boldsymbol{\gamma}) = -(\partial^2 PPLL)/(\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}')$, sum of $-(\partial^2 PLL)/(\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}')$ and $\boldsymbol{\Sigma}^{-1}$, is positive-definite; so is its inverse $\boldsymbol{K}_{PPL2}(\boldsymbol{\gamma})^{-1}$. In line of the *Remark* blow, $\left[\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})^{-1}\right]_{blk_i}$ are positive-definite. As follows, the sum of quadratic terms $\widehat{\boldsymbol{\gamma}}_i\widehat{\boldsymbol{\gamma}}_i'$ and $\left[\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})^{-1}\right]_{blk_i}$ would produce a positive-definite estimator $\widehat{\boldsymbol{D}}^{\#}$. The *remark* is claimed and proved as below.

Remark. If $\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})^{-1}$ is positive-definite, then $\left[\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})^{-1}\right]_{blk_i}$ are positive-definite.

Proof. Let $\boldsymbol{I}_i = [\boldsymbol{0}_{(1)}, \ldots, \boldsymbol{1}_{(i)}, \ldots, \boldsymbol{0}_{(n)}]'_{2\times 2n}$, where $\boldsymbol{1}_i$ is a 2×2 identity matrix located at the i^{th} horizontal block or occupying columns $2i-1$ and $2i$, leaving other components to be 0. Thus we have $\left[\boldsymbol{K}''_{PPL}(\widehat{\boldsymbol{\gamma}})^{-1}\right]_{blk_i} = \boldsymbol{I}_i'\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})^{-1}\boldsymbol{I}_i$. Since $\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})^{-1}$ is positive-definite, for $\forall \boldsymbol{x} \neq \boldsymbol{0}$, we shall have $\boldsymbol{x}'\left[\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})^{-1}\right]_{blk_i}\boldsymbol{x} = \boldsymbol{x}^{*T}\boldsymbol{K}_{PPL2}(\widehat{\boldsymbol{\gamma}})^{-1}\boldsymbol{x}^{*} > 0$, where $\boldsymbol{x}^{*} = [\boldsymbol{0}_1, \ldots, \boldsymbol{x}, \ldots, \boldsymbol{0}_n]' \neq \boldsymbol{0}$. □

A.2 Appendix Table 1

Table A.1: *Estimating regression coefficients and variance components for varying cluster sizes, based on 500 replicates, with* $n = 100$ *and* $\lambda_{01} = \lambda_{02}$.

	True	Strong $\boldsymbol{D}$				True	Weak $\boldsymbol{D}$			
	Value	Bias	ESD	ASE	CP	Value	Mean	ESD	ASE	CP
λ_{0k}	6	$\widetilde{m}_i = 16$					$\widetilde{m}_i = 18$			
β_1	1	-0.001	0.029	0.028	0.948	1	-0.002	0.030	0.030	0.940
β_2	-1	-0.001	0.028	0.029	0.944	-1	0.001	0.030	0.030	0.944
$\boldsymbol{D}[1,1]$	0.7	-0.031	0.105	0.107	0.900	0.25	-0.006	0.046	0.043	0.894
$\boldsymbol{D}[2,2]$	1.2	-0.047	0.193	0.179	0.874	0.25	-0.001	0.044	0.044	0.926
$\boldsymbol{D}[1,2]$	0.2	-0.038	0.101	0.099	0.916	0.125	-0.006	0.033	0.033	0.938
λ_{0k}	15	$\widetilde{m}_i = 40$					$\widetilde{m}_i = 45$			
β_1	1	-0.001	0.018	0.018	0.952	1	-0.001	0.019	0.019	0.954
β_2	-1	-0.000	0.018	0.018	0.958	-1	-0.000	0.018	0.019	0.964
$\boldsymbol{D}[1,1]$	0.7	-0.014	0.104	0.102	0.908	0.25	-0.005	0.041	0.038	0.902
$\boldsymbol{D}[2,2]$	1.2	-0.016	0.163	0.171	0.920	0.25	-0.005	0.040	0.038	0.930
$\boldsymbol{D}[1,2]$	0.2	-0.018	0.099	0.095	0.932	0.125	-0.005	0.031	0.029	0.922

$\widetilde{m}_i$: median of m_i.

APPENDIX B

An Estimating Equation Framework for a Flexible Class of Semiparametric Frailty Models

B.1 Proof: Second Moment of a Non-stationary Poisson Process

For an arbitary non-stationary Poisson process $N(t)$ whose event rate satisfies $E(dN(t)) = \lambda(t)dt = d\Lambda(t)$ and consequently $E(N(t)) = \Lambda(t)$, one can show that

$$\begin{aligned}
E(dN(t)) =& d\Lambda(t) \simeq d\Lambda(t) + [d\Lambda(t)]^2 \\
=& E([dN(t)]^2) = E([N(t) - N(t^-)]^2) = E(N(t)^2 + N(t^-)^2 - 2N(t)N(t^-)) \\
=& E(N(t)^2) + E(N(t^-)^2) - 2E(N(t)N(t^-)) \\
=& E(N(t)^2) + E(N(t^-)^2) - 2E(dN(t)N(t^-) + N(t^-)^2) \\
=& E(N(t)^2) - E(N(t^-)^2) - 2E(dN(t))E(N(t^-)) \text{ by independent increments} \\
=& E(d[N(t)^2]) - 2d\Lambda(t)\Lambda(t^-) = E(d[N(t)^2]) - 2d\Lambda(t)\Lambda(t) \\
=& E(d[N(t)^2]) - d[\Lambda(t)^2],
\end{aligned}$$

and consequently

$$E(N(t)) = E([N(t)^2]) - [\Lambda(t)^2] = Var(N(t)).$$

B.2 Other Estimating Methods for Model A

B.2.1 Estimation via Generalized Method of Moments

In this subsection, we intend to derive the influence functions based on the two moment conditions. For the first moment given in (3.7), the influence functions for subject i are

$$\boldsymbol{g}_1(\boldsymbol{z}_i, \boldsymbol{\beta}_1) = w_{1i}\bar{\boldsymbol{z}}_{1i}(m_i F^{-1}(c_i) - e^{\boldsymbol{\beta}_1'\bar{\boldsymbol{z}}_{1i}}), \tag{B.1}$$

where the most efficient weight function is $w_{1i} = e^{\widehat{\boldsymbol{\beta}}_1'\bar{\boldsymbol{z}}_{1i}}/\widehat{E}[(M_iF^{-1}(C_i) - e^{\boldsymbol{\beta}_1'\bar{\boldsymbol{Z}}_{1i}})^2]$, though for convenience, we usually set them to be $w_{1i} = 1$. The shape distribution function $F(t)$ is unobserved so we replace it with $\widehat{F}(t)$. The estimator $\widehat{\boldsymbol{\beta}}_1$ solves

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}_1(\boldsymbol{z}_i, \boldsymbol{\beta}_1) = 0, \tag{B.2}$$

which can be solved by Newton-Raphson. The first derivative of (B.2) with respect to $\boldsymbol{\beta}_1$ is $-1/n\sum_{i=1}^{n} w_{1i}\exp(\boldsymbol{\beta}_1'\bar{\boldsymbol{z}}_{1i})\bar{\boldsymbol{z}}_{1i}\bar{\boldsymbol{z}}_{1i}'$. The final estimate $\widehat{\boldsymbol{\beta}}_1 = (\widehat{\beta}_0, \widehat{\boldsymbol{\beta}}')'$, where $\widehat{\Lambda}_0(\tau) = \exp(\widehat{\beta}_0)$ and $\widehat{\boldsymbol{\beta}}$ is the estimated effects from the covariates. As follows, we have the estimates for the baseline (not just the shape distribution $\widehat{F}(t)$) from $\widehat{\Lambda}_0(t) = \widehat{\Lambda}_0(\tau)\widehat{F}(t)$.

Based on the second moment condition (3.9), we obtain the second batch of influence functions

$$\boldsymbol{g}_2(\boldsymbol{z}_i, \boldsymbol{\beta}_2) = w_{2i}\bar{\boldsymbol{z}}_{2i}((m_i^2 - m_i)F^{-2}(c_i) - e^{\boldsymbol{\beta}_2'\bar{\boldsymbol{z}}_{2i}}), \tag{B.3}$$

where $\beta_\sigma = 2\ln(\Lambda_0(\tau)) + \ln(E(\gamma^2))$ and $\bar{z}_{2i} = [2z_i', 1]'$.

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}_1(z_i, \boldsymbol{\beta}_1) = 0, \tag{B.4}$$

which can be solved by Newton-Raphson. The first derivative of (B.4) with respect to $\boldsymbol{\beta}_1$ is $-1/n\sum_{i=1}^{n} w_{1i}\exp(\boldsymbol{\beta}_1'\bar{z}_{1i})\bar{z}_{1i}\bar{z}_{1i}'$. The final estimate $\widehat{\boldsymbol{\beta}}_1 = (\widehat{\beta}_0, \widehat{\boldsymbol{\beta}}')'$, where $\widehat{\Lambda}_0(\tau) = \exp(\widehat{\beta}_0)$ and $\widehat{\boldsymbol{\beta}}$ is the estimated effects from the covariates.

Let the regression parameter $\boldsymbol{\beta}$ have p elements, such that the total number of parameters $\boldsymbol{\theta} = (\beta_0, \boldsymbol{\beta}', \beta_\sigma)'$ under estimation is $p + 2$. Note that if we stack the two influence functions together, $\boldsymbol{g}(z_i; \boldsymbol{\theta}) = [\boldsymbol{g}_1(z_i; \boldsymbol{\beta}_1)', \boldsymbol{g}_2(z_i; \boldsymbol{\beta}_2)']'$, we can formulate a combined group of equations

$$\boldsymbol{g}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}(z_i; \boldsymbol{\theta}) = 0. \tag{B.5}$$

Note that $\boldsymbol{g}_n(\boldsymbol{\theta})$ has dimension $2p + 2$, which is larger than the dimension of the parameters we are estimating. Generally, B.5 cannot be solved exactly.

One may utilize GMM for the parameter estimation (*Hansen*, 1982; *Hansen et al.*, 1996). The objective function for the GMM can be defined as

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \boldsymbol{g}_n(\boldsymbol{\theta})'\boldsymbol{W}_n\boldsymbol{g}_n(\boldsymbol{\theta}), \tag{B.6}$$

where $\boldsymbol{W}_n$ is a positive semi-definite matrix with a well-defined limit. The optimal weight matrix is the inverse of the variance matrix of $\sqrt{n}\boldsymbol{g}_n(\boldsymbol{\theta})$, or its consistent estimator

$$\boldsymbol{W}_n = \left[n\boldsymbol{g}_n(\widehat{\boldsymbol{\theta}})\boldsymbol{g}_n(\widehat{\boldsymbol{\theta}})'\right]^{-1}. \tag{B.7}$$

Since we do not know $F(t)$, we replace it with its consistent estimator $\widehat{F}(t)$. The influence functions are then specified as $\widehat{\boldsymbol{g}}_n(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\widehat{\boldsymbol{g}}(z_i; \boldsymbol{\theta})$. Similar to the estimator from Subsection 3.1, we plug in $\widehat{F}(t)$ from (3.6), such that the objective function becomes

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \widehat{\boldsymbol{g}}_n(\boldsymbol{\theta})' \boldsymbol{W}_n \widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}). \tag{B.8}$$

B.2.2 Estimation via Empirical Likelihood

We also develop a method based on empirical likelihood (EL)(*Smith*, 1997). As noted in *Newey and Smith* (2004) and *Anatolyev* (2005), the second order bias of an EL estimator is generally smaller than the bias of the corresponding GMM estimator. In contrast to GMM, the bias does not increase with the number of moment conditions. Moreover, efficiency improves when the number of conditions increases. With these benefits, we are proposing to obtain the estimates of $\boldsymbol{\theta}$ through empirical likelihood. The estimator is defined as the solution to the following constrained minimization problem:

$$\widehat{\boldsymbol{\theta}} = \arg\max_{p_i, \boldsymbol{\theta}} \prod_{i=1}^{n} p_i, \tag{B.9}$$

subject to

$$p_i > 0, \ \sum_{i=1}^{n} p_i = 1, \text{ and } \sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{z}_i; \boldsymbol{\theta}) p_i = 0 \tag{B.10}$$

B.2.3 Simulation Results

The three proposed methods, EE, GMM and EL-based estimations (the latter two were implemented in the R package gmm (*Chaussé*, 2010)). Each simulation experiment sample size is 500, with 1000 replicates. Bivariate covariates were generated from a $\boldsymbol{Bernoulli}(0.5)$ and a $\boldsymbol{Normal}(0,1)$, denoted as $\boldsymbol{Z}_i = (Z_{1i}, Z_{2i})'$, with regression coefficients $\boldsymbol{\beta} = (0.5, -0.3)$. The distributions for frailties we considered here include Gamma ($\gamma_i \sim \boldsymbol{Gamma}(1,1)$) and log-normal ($\gamma_i \sim \boldsymbol{LogNorm}(-\log(2)/2, \log(2))$), with unit mean and variance. For each subject, events were generated following exponential distribution with event rate $\lambda_i(t) = 0.25 \exp(\boldsymbol{\beta}' \boldsymbol{z}_i)\gamma_i$. Let the ending time of the study is $\boldsymbol{\tau} = 10$ and the censoring time $\boldsymbol{C}_i \sim \boldsymbol{Uniform}(2, 10)$, such that about 40% of the subjects were censored ending up with 0 events observed. Note that we did not

Table B.1: Gamma frailty: n=500

	True	EE		GMM		EL	
	Value	Mean	ESE	Mean	ESE	Mean	ESE
$\Lambda_0(\tau)$	2.5	2.512	0.276	2.475	0.278	2.502	0.272
$\boldsymbol{\beta}[1]$	0.5	0.494	0.117	0.490	0.122	0.493	0.123
$\boldsymbol{\beta}[2]$	-0.3	-0.298	0.059	-0.281	0.061	-0.287	0.065
$Var(\gamma)$	1	0.978	0.169	0.910	0.144	0.948	0.147
time(s)	-	0.069	-	0.163	-	1.88	-

Table B.2: Log-normal frailty: n=500

	True	EE		GMM		EL	
	Value	Mean	ESE	Mean	ESE	Mean	ESE
$\Lambda_0(\tau)$	2.5	2.512	0.287	2.458	0.283	2.491	0.286
$\boldsymbol{\beta}[1]$	0.5	0.500	0.118	0.493	0.114	0.498	0.117
$\boldsymbol{\beta}[2]$	-0.3	-0.295	0.060	-0.282	0.057	-0.286	0.066
$Var(\gamma)$	1	0.969	0.279	0.807	0.193	0.876	0.194
time(s)	-	0.056	-	0.142		1.719	-

include bootstrap for standard error estimation in Tables B.1 and B.2.

In both Gamma and lognormal cases, the proposed EE estimation method worked best: most accurate and fast. We also did extra experiments with $n = 10,000$, the estimation results using GMM were quite accurate and efficient, but appeared no obvious advantages in comparison with EE (data not shown here). In sum, we gave up the idea of using GMM or EL to improve the estimating efficiency. It seems that, when the two moment conditions are highly correlated, adding the second moment condition does not help much, but instead, causes a lot of instability.

B.3 Derivation for the IID Representation of $\sqrt{n}(\widehat{F}(t) - F(t))$

We derive the iid representation of $\sqrt{n}(\widehat{F}(t) - F(t))$ for shared frailty model (Model A) and the correlated frailty model (Model B) following the similar proofs of *Wang et al.* (2001) in its appendix, then we derive the iid representation of $\sqrt{n}(\widehat{F}(t) - F(t))$

of the nested frailty model (Model C).

Let $G(t) = E[\gamma \exp(\boldsymbol{\beta}'\boldsymbol{Z})I(C \geq t)] =\in \tau^{\tau}\gamma \exp(\boldsymbol{\beta}'\boldsymbol{z})I(c \geq t)dW(c,\gamma,\boldsymbol{z})$. Then we use $G(t)$ to define $R(t) = G(t)\Lambda_0(t)$ and $Q(t) =\in \tau^{t}G(u)d\Lambda_0(u)$. Since $F(\tau) = 1$ by its definition, the equality holds that

$$-\ln F(t) = \ln F(\tau) - \ln F(t) = \int_t^{\tau} \frac{dF(u)}{F(u)} = \int_t^{\tau} \frac{dQ(u)}{R(u)}.$$

It can be easily shown that their unbiased estimators are $\widehat{R}(t) = 1/n \sum_{i=1}^{n} \sum_{j=1}^{m_i} I(t_{ij} \leq u \leq c_i)$ and $\widehat{Q}(t) = 1/n \sum_{i=1}^{n} \sum_{j=1}^{m_i} I(t_{ij} \leq u)$, satisfying $E(\widehat{R}(t)) = R(t)$ and $E(\widehat{Q}(t)) = Q(t)$.

According to the definition of $\widehat{F}(t)$ in (3.6),

$$\widehat{F}(t) = \prod_{S_{(l)}>t} \left(1 - \frac{d_{(l)}}{N_{(l)}}\right) = \prod_{t<u\leq\tau} \left(1 - \frac{d\widehat{Q}(u)}{\widehat{R}(u)}\right)$$

Equivalently, we have

$$-\ln \widehat{F}(t) = -\int_t^{\tau} \ln\left(1 - \frac{d\widehat{Q}(u)}{\widehat{R}(u)}\right)$$

Note that if the assumptions A1-2 are satisfied, we have $R(u) > 0$ for $\forall u \in [\tau_0, \tau]$, where $\tau_0 > \inf\{t : \Lambda_0(t) > 0\}$. As $n \to \infty$, both $\widehat{R}(t)$ and $\widehat{Q}(t)$ converge almost surely to $R(t)$ and $Q(t)$ in $u \in [\tau_0, \tau]$. Through approximation method for the product-limit estimators and because of the inequality $0 \leq -\ln(1-v) - v \leq v^2(1-v)$ for $v \in [0,1)$, one can show that for $\forall t \in [\tau_0, \tau]$, we have

$$-\int_t^{\tau} \ln\left(1 - \frac{d\widehat{Q}(u)}{\widehat{R}(u)}\right) - \int_t^{\tau} \frac{d\widehat{Q}(u)}{\widehat{R}(u)} = -\ln \widehat{F}(t) - \int_t^{\tau} \frac{d\widehat{Q}(u)}{\widehat{R}(u)} \xrightarrow{a.s.} 0.$$

As follows, through a continuous mapping, because $\widehat{Q}(u) = Q(u) + Op(n^{-\frac{1}{2}})$ and

$\widehat{R}(u) = R(u) + Op(n^{-\frac{1}{2}})$, then at each $t \in [\tau_0, \tau]$, we have

$$\begin{aligned}\widehat{F}(t) &= \exp\left(-\int_t^\tau \frac{d\widehat{Q}(u)}{\widehat{R}(u)}\right) + op(n^{-\frac{1}{2}}) \\ &= \exp\left(-\int_t^\tau \frac{dQ(u)}{R(u)}\right) + Op(n^{-\frac{1}{2}}) \\ &= F(t) + Op(n^{-\frac{1}{2}}).\end{aligned} \tag{B.11}$$

Thus $\widehat{F}(t)$ is a consistent estimator for $F(t)$, while for its iid representation, we will need to figure out the $Op(n^{-\frac{1}{2}})$ part in the above equations (B.11).

$$\begin{aligned}-\ln \widehat{F}(t) &= \int_t^\tau \frac{d\widehat{Q}(u)}{\widehat{R}(u)} = \int_t^\tau \frac{d\widehat{Q}(u)}{R(u)} + \int_t^\tau d\widehat{Q}(u)\left(\frac{1}{\widehat{R}(u)} - \frac{1}{R(u)}\right) \\ &= \int_t^\tau \frac{dQ(u)}{R(u)} + \int_t^\tau \frac{d\widehat{Q}(u) - dQ(u)}{R(u)} + \int_t^\tau d\widehat{Q}(u)\frac{R(u) - \widehat{R}(u)}{\widehat{R}(u)R(u)} \\ &= -\ln F(t) + \int_t^\tau \frac{d\widehat{Q}(u) - dQ(u)}{R(u)} + \int_t^\tau dQ(u)\frac{R(u) - \widehat{R}(u)}{R(u)^2} \\ &+ \int_t^\tau [d\widehat{Q}(u) - dQ(u)]\frac{R(u) - \widehat{R}(u)}{R(u)^2} + \int_t^\tau d\widehat{Q}(u)\frac{R(u) - \widehat{R}(u)}{R(u)}\left(\frac{1}{\widehat{R}(u)} - \frac{1}{R(u)}\right) \\ &= -\ln F(t) + \int_t^\tau \frac{d\widehat{Q}(u) - dQ(u)}{R(u)} + \int_t^\tau dQ(u)\frac{R(u) - \widehat{R}(u)}{R(u)^2} + op(n^{-\frac{1}{2}}) \\ &= -\ln F(t) + \int_t^\tau \frac{d\widehat{Q}(u)}{R(u)} - \int_t^\tau \frac{dQ(u)\widehat{R}(u)}{R(u)^2} + op(n^{-\frac{1}{2}})\end{aligned} \tag{B.12}$$

Hereafter, we have

$$\begin{aligned}\ln \widehat{F}(t) - \ln F(t) &= \int_t^\tau \frac{dQ(u)\widehat{R}(u)}{R(u)^2} - \int_t^\tau \frac{d\widehat{Q}(u)}{R(u)} + op(n^{-\frac{1}{2}}) \\ &= \frac{1}{n}\sum_{i=1}^n b_i(t) + op(n^{-\frac{1}{2}}),\end{aligned} \tag{B.13}$$

where

$$b_i(t) = \sum_{l=1}^{m_i} \left\{ \int_t^\tau \frac{I(t_{il} \leq u \leq c_i) dQ(u)}{R(u)^2} - \frac{I(t < t_{il} \leq \tau)}{R(t_{il})} \right\}. \tag{B.14}$$

It is natural to see that

$$E\{b_i(t)\} = E\left\{\frac{1}{n}\sum_{i=1}^n b_i(t)\right\} = E\left\{\int_t^\tau \frac{dQ(u)\widehat{R}(u)}{R(u)^2} - \int_t^\tau \frac{d\widehat{Q}(u)}{R(u)}\right\} = 0. \tag{B.15}$$

Hence we have the iid representation for $\sqrt{n}(\ln \widehat{F}(t) - \ln F(t))$

$$\sqrt{n}(\ln \widehat{F}(t) - \ln F(t)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n b_i(t) + op(1). \tag{B.16}$$

By delta method, we obtain the iid representation of $\sqrt{n}(\widehat{F}(t) - F(t))$ is

$$\sqrt{n}(\widehat{F}(t) - F(t)) = \frac{F(t)}{\sqrt{n}}\sum_{i=1}^n b_i(t) + op(1). \tag{B.17}$$

For the correlated frailty model (Model B), we introduce in additional subscript $j \in \{1, 2\}$ for the two different event types. Let $G_j(t) = E[\gamma_j \exp(\boldsymbol{\beta}_j' \boldsymbol{Z}) I(C_j \geq t)]$, and thus we have $R_j(t) = G_j(t)\Lambda_{0j}(t)$ and $Q_j(t) = \in \tau^t G_j(u) d\Lambda_{0j}(u)$. Note that for the shape function of each event type, we can show following almost identical steps in (B.12)-(B.17) that

$$\sqrt{n}(\widehat{F}_j(t) - F_j(t)) = \frac{F_j(t)}{\sqrt{n}}\sum_{i=1}^n b_{ji}(t) + op(1),$$

where $b_{ji}(t)$ is given by

$$b_{ji}(t) = \sum_{l=1}^{m_{ji}} \left\{ \int_t^\tau \frac{I(t_{jil} \leq u \leq c_{ji}) dQ_j(u)}{R_j(u)^2} - \frac{I(t < t_{jil} \leq \tau)}{R_j(t_{jil})} \right\}.$$

For the nested frailty model (Model C), $\sqrt{K}(\widehat{F}(t) - F(t))$ can be written into the similar form of (B.17). Because of the independence between I_k and $(N_{ki}(t), c_{ki}, \epsilon_k, \gamma_{ki}, z_{ki})$, the G function will become

$$\begin{aligned} G_c(t) =& E\left\{ \sum_{i=1}^{I_1} \epsilon_1 \gamma_{1i} \exp(\boldsymbol{\beta}' \boldsymbol{Z}) I(C_{1i} \geq t) \right\} = E(I_1) E\left\{ \epsilon_1 \gamma_{11} \exp(\boldsymbol{\beta}' \boldsymbol{Z}) I(C_{11} \geq t) \right\} \\ =& \nu \int_0^\tau \epsilon \gamma \exp(\boldsymbol{\beta}' \boldsymbol{z}) I(c \geq t) dW(c, \epsilon, \gamma, \boldsymbol{z}), \end{aligned} \tag{B.18}$$

where we denote $\nu = E(I_1) = E(I_k)$.

Then we define $R_c(t) = G_c(t)\Lambda_0(t)$ and $Q_c(t) = \in \tau^t G_c(u) d\Lambda_0(u)$, and their unbiased estimators are $\widehat{R}_c(t) = 1/K \sum_{k=1}^K \sum_{i=1}^{I_k} \sum_{j=1}^{m_{ki}} I(t_{kij} \leq u \leq c_i)$ and $\widehat{Q}_c(t) = 1/K \sum_{k=1}^K \sum_{i=1}^{I_k} \sum_{j=1}^{m_{ki}} I(t_{kij} \leq u)$, satisfying $E(\widehat{R}_c(t)) = R_c(t)$ and $E(\widehat{Q}_c(t)) = Q_c(t)$. Thus, we end up with the iid representation of $\sqrt{K}(\widehat{F}(t) - F(t))$ for Model C:

$$\sqrt{K}(\widehat{F}(t) - F(t)) = \frac{F(t)}{\sqrt{K}} \sum_{i=1}^K b_{ck}(t) + op(1), \tag{B.19}$$

where

$$b_{ck}(t) = \sum_{i=1}^{I_k} \sum_{l=1}^{m_{ki}} \left\{ \int_t^\tau \frac{I(t_{kil} \leq u \leq c_i) dQ_c(u)}{R_c^2(u)} - \frac{I(t < t_{kil} \leq \tau)}{R_c(t_{kil})} \right\}. \tag{B.20}$$

B.4 Asymptotic Properties for $\widehat{\boldsymbol{\theta}}$

Taking the shared frailty model (Model A) as an example, which can be extended to correlated frailty model (Model B) and nested frailty model (Model C) analogously.

$$
\begin{aligned}
&\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\bar{z}_i\left(\frac{m_i}{\widehat{F}(c_i)}-\exp(\boldsymbol{\theta}'\bar{z}_i)\right)\right]\\
&=\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\bar{z}_i\left(\frac{m_i}{\widehat{F}(c_i)}-\frac{m_i}{F(c_i)}\right)+\frac{1}{n}\sum_{i=1}^{n}\bar{z}_i\left(\frac{m_i}{F(c_i)}-\exp(\boldsymbol{\theta}'\bar{z}_i)\right)\right\}\\
&=\sqrt{n}\left\{\int\left(\frac{m\bar{z}[F(c)-\widehat{F}(c)]}{F(c)^2}\right)dV(z,m,c)\right.\\
&\qquad\qquad\qquad\left.+\frac{1}{n}\sum_{i=1}^{n}\bar{z}_i\left(\frac{m_i}{F(c_i)}-\exp(\boldsymbol{\theta}'\bar{z}_i)\right)\right\}+op(1)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{-\int\frac{\bar{z}mb_i(c)}{F(c)}dV(z,m,c)+\bar{z}_i\left(\frac{m_i}{F(c_i)}-\exp(\boldsymbol{\theta}'\bar{z}_i)\right)\right\}+op(1)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{e}_i+op(1),
\end{aligned}
\tag{B.21}
$$

where $\boldsymbol{e}_i=-\int\frac{\bar{\boldsymbol{x}}_1mb_i(c)}{F(c)}dV(z,m,c)+w_{1i}\bar{z}_i\left(\frac{m_i}{F(c_i)}-\exp(\boldsymbol{\theta}'\bar{z}_i)\right)$, and $V(z,m,c)$ is the joint probability measure for (z,m,c). Since $E(b_i(t))=0$ and $E(\frac{m_i}{F(c_i)}-\exp(\boldsymbol{\theta}'\bar{z}_i)=0$, we have $E(\boldsymbol{e}_i)=0$

Let $\boldsymbol{\theta}$ be the unique true value in a compact parameter space, through some standard procedures for Z-estimation, we have $\widehat{\boldsymbol{\theta}}\rightarrow_p\boldsymbol{\theta}$, and the following equation through a Taylor expansion:

$$
\begin{aligned}
&0=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{e}_i+\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\boldsymbol{e}_i}{\partial\boldsymbol{\theta}}\right)\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\\
&\Rightarrow\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})=\left(-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\boldsymbol{e}_i}{\partial\boldsymbol{\theta}}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{e}_i+op(1)=E\left[-\frac{\partial\boldsymbol{e}_i}{\partial\boldsymbol{\theta}}\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{e}_i+op(1)
\end{aligned}
\tag{B.22}
$$

Let $\boldsymbol{\psi} = E\left[-\frac{\partial \boldsymbol{e}_i}{\partial \boldsymbol{\theta}}\right]$ and $\boldsymbol{\Sigma} = E(\boldsymbol{e}_i\boldsymbol{e}_i')$, thus we obtain

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{\boldsymbol{\psi}^{-1}}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{e}_i + op(1), \tag{B.23}$$

which converges weakly to the multivariate normal distribution with mean 0 and variance covariance matrix $\boldsymbol{\psi}^{-1}\boldsymbol{\Sigma}(\boldsymbol{\psi}')^{-1}$

For Model C, we similar derivations that

$$\begin{aligned}
&\sqrt{K}\left[\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{I_k}\bar{\boldsymbol{z}}_{ki}\left(\frac{m_{ki}}{\widehat{F}(c_{ki})} - \exp(\boldsymbol{\theta}'\bar{\boldsymbol{z}}_{ki})\right)\right] \\
&= \sqrt{K}\left\{\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{I_k}\bar{\boldsymbol{z}}_i\left(\frac{m_{ki}}{\widehat{F}(c_{ki})} - \frac{m_{ki}}{F(c_{ki})}\right) + \frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{I_k}\bar{\boldsymbol{z}}_{ki}\left(\frac{m_{ki}}{F(c_{ki})} - \exp(\boldsymbol{\theta}'\bar{\boldsymbol{z}}_{ki})\right)\right\} \\
&= \sqrt{K}\left\{\nu\int\left(\frac{m\bar{\boldsymbol{z}}[F(c) - \widehat{F}(c)]}{F(c)^2}\right)dV(\boldsymbol{z}, m, c)\right. \\
&\qquad\qquad\qquad \left. + \frac{1}{K}\sum_{i=1}^{I_k}\bar{\boldsymbol{z}}_{ki}\left(\frac{m_{ki}}{F(c_{ki})} - \exp(\boldsymbol{\theta}'\bar{\boldsymbol{z}}_{ki})\right)\right\} + op(1) \\
&= \frac{1}{\sqrt{K}}\sum_{k=1}^{K}\left\{-\nu\int\frac{m\bar{\boldsymbol{z}}b_{ck}(c)}{F(c)}dV(\boldsymbol{z}, m, c) + \sum_{i=1}^{I_k}\bar{\boldsymbol{z}}_{ki}\left(\frac{m_{ki}}{F(c_{ki})} - \exp(\boldsymbol{\theta}'\bar{\boldsymbol{z}}_{ki})\right)\right\} + op(1) \\
&= \frac{1}{\sqrt{K}}\sum_{k=1}^{K}\boldsymbol{e}_{ck} + op(1),
\end{aligned} \tag{B.24}$$

where $\boldsymbol{e}_{ck} = -\nu\int\frac{m\bar{\boldsymbol{z}}b_{ck}(c)}{F(c)}dV(\boldsymbol{z}, m, c) + \sum_{i=1}^{I_k}\bar{\boldsymbol{z}}_{ki}\left(\frac{m_{ki}}{F(c_{ki})} - \exp(\boldsymbol{\theta}'\bar{\boldsymbol{z}}_{ki})\right)$. Recall that $\nu = E(I_k)$ and is supposed to be independent of $(\boldsymbol{Z}, \boldsymbol{M}, \boldsymbol{C})$ based on the assumptions.

B.5 Proof of Asymptotic Properties for the variance components

We first derive the iid representation for each components and then combine them together using delta method.

$$\begin{aligned}
&\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}(m_i^2-m_i)\widehat{F}(c_i)^{-2}-\Lambda_0^2(\tau)E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]E(\gamma^2)\right\}\\
&=\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}(m_i^2-m_i)F(c_i)^{-2}-\int\frac{(m^2-m)2F(c)(\widehat{F}(c)-F(c))}{F(c)^4}dH(m,c)\right.\\
&\qquad\qquad\left.-\Lambda_0^2(\tau)E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]E(\gamma^2)\right\}+op(1)\\
&=\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}(m_i^2-m_i)F(c_i)^{-2}-\int\frac{(m^2-m)2b_i(c)}{F(c)^2}dH(m,c)\right.\\
&\qquad\qquad\left.-\Lambda_0^2(\tau)E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]E(\gamma^2)\right\}+op(1)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_i+op(1),
\end{aligned}\tag{B.25}$$

where $g_i=(m_i^2-m_i)F(c_i)^{-2}-\int\frac{(m^2-m)2b_i(c)}{F(c)^2}dH(m,c)-\Lambda_0^2(\tau)E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]E(\gamma^2)$ and $H(m,c)$ is the joint probability measure of (m,c). Note that we can also show that $E(g_i)=0$.

Let $[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1$ and $[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_{-1}$ denote the first entry and the rest entries (without the first one) of $\boldsymbol{\psi}^{-1}\boldsymbol{e}_i$ respectively. In other words, $\boldsymbol{\psi}^{-1}\boldsymbol{e}_i=[[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1,[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]'_{-1}]'$. Then we have

$$\begin{aligned}
\sqrt{n}\left(\ln\widehat{\Lambda}_0(\tau)-\ln\Lambda_0(\tau)\right)&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1+op(1)\\
\Rightarrow\sqrt{n}\left(\ln\widehat{\Lambda}_0^2(\tau)-\ln\Lambda_0^2(\tau)\right)&=\frac{2}{\sqrt{n}}\sum_{i=1}^{n}[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1+op(1)\\
\Rightarrow\sqrt{n}\left(\widehat{\Lambda}_0^2(\tau)-\Lambda_0^2(\tau)\right)&=\frac{2\Lambda_0^2(\tau)}{\sqrt{n}}\sum_{i=1}^{n}[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1+op(1)
\end{aligned}\tag{B.26}$$

One can rewrite the regression parameters from (B.21) into

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\boldsymbol{\psi}^{-1} \boldsymbol{e}_i]_{-1} + op(1), \tag{B.27}$$

which also implies that $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + op(1)$, and thus $\widehat{\boldsymbol{\beta}}'\boldsymbol{z} \rightarrow_p \boldsymbol{\beta}'\boldsymbol{z}$ for any bounded $\boldsymbol{z}$.

As follows, if the covariates $\boldsymbol{Z}$ are bounded in a compact subspace of $\mathbb{R}^q$, one can develop a Taylor expansion that

$$\begin{aligned} \exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{z}) &= \exp(\boldsymbol{\beta}'\boldsymbol{z}) + \exp(\boldsymbol{\beta}'\boldsymbol{z})(\widehat{\boldsymbol{\beta}}'\boldsymbol{z} - \boldsymbol{\beta}'\boldsymbol{z}) + Op(1/n) \\ &= \exp(\boldsymbol{\beta}'\boldsymbol{z}) + \exp(\boldsymbol{\beta}'\boldsymbol{z})\boldsymbol{z}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + op(n^{-\frac{1}{2}}) \end{aligned}$$

As follows, we obtain the iid representation for the exponential covariate part is

$$\begin{aligned} &\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} \exp(2\widehat{\boldsymbol{\beta}}'\boldsymbol{z}_i) - E\left\{\exp(2\boldsymbol{\beta}'\boldsymbol{Z})\right\}\right] \\ &= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} \left\{\exp(2\widehat{\boldsymbol{\beta}}'\boldsymbol{z}_i) - \exp(2\boldsymbol{\beta}'\boldsymbol{z}_i)\right\} + \exp(2\boldsymbol{\beta}'\boldsymbol{z}_i) - E(\exp(2\boldsymbol{\beta}'\boldsymbol{Z}))\right] \\ &= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} \left\{\exp(2\boldsymbol{\beta}'\boldsymbol{z}_i)\boldsymbol{z}_i'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\} + \exp(2\boldsymbol{\beta}'\boldsymbol{z}_i) - E(\exp(2\boldsymbol{\beta}'\boldsymbol{Z}))\right] + op(1) \\ &= \sqrt{n}\left\{\int [2\exp(2\boldsymbol{\beta}'\boldsymbol{z})\boldsymbol{z}']\, dU(\boldsymbol{z})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \frac{1}{n}\sum_{i=1}^{n} \exp(2\boldsymbol{\beta}'\boldsymbol{z}_i) - E(\exp(2\boldsymbol{\beta}'\boldsymbol{Z}))\right\} + op(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left\{2E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})\boldsymbol{Z}'][\psi^{-1}\boldsymbol{e}_i]_{-1} + \exp(2\boldsymbol{\beta}'\boldsymbol{z}_i) - E(\exp(2\boldsymbol{\beta}'\boldsymbol{Z}))\right\} + op(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_i + op(1), \end{aligned} \tag{B.28}$$

where

$$h_i = 2E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})\boldsymbol{Z}'][\psi^{-1}\boldsymbol{e}_i]_{-1} + \exp(2\boldsymbol{\beta}'\boldsymbol{z}_i) - E(\exp(2\boldsymbol{\beta}'\boldsymbol{Z})).$$

Note that since $E(e_i) = 0$, it would be straightforward to show that $E(h_i) = 0$.

We combine all the previously derived iid representations and form the iid representation for $\sqrt{n}\left(\widehat{E}(\gamma^2) - E(\gamma^2)\right)$ using the delta method:

$$\sqrt{n}\left(\widehat{E}(\gamma^2) - E(\gamma^2)\right) = \sqrt{n}\left\{\frac{\sum_{i=1}^{n}(m_i^2 - m_i)\widehat{F}(c_i)^{-2}}{\widehat{\Lambda}_0(\tau)^2 \sum_{i=1}^{n} \exp(2\widehat{\boldsymbol{\beta}}'z_i)} - E(\gamma^2)\right\}$$

$$= \left\{\frac{1}{\Lambda_0^2(\tau)E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]}, -\frac{E(\gamma^2)}{\Lambda_0^2(\tau)}, -\frac{E(\gamma^2)}{E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]}\right\}$$

$$\times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{g_i, 2\Lambda_0^2(\tau)[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1, h_i\right\}' + op(1) \tag{B.29}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{g_i}{\Lambda_0^2(\tau)E\left\{\exp(2\boldsymbol{\beta}'\boldsymbol{Z})\right\}} - 2E(\gamma^2)[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1 - \frac{E(\gamma^2)h_i}{E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]}\right\}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i + op(1)$$

where s_i are

$$s_i = \left\{\frac{g_i}{\Lambda_0^2(\tau)E\left\{\exp(2\boldsymbol{\beta}'\boldsymbol{Z})\right\}} - 2E(\gamma^2)[\boldsymbol{\psi}^{-1}\boldsymbol{e}_i]_1 - \frac{2E(\gamma^2)h_i}{E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]}\right\}. \tag{B.30}$$

Thus via the central limit theorem, the simple estimator converges to a mean-0 normal distribution with variance $E(s_i^2)$. Thus the variance $Var(\gamma) = E(\gamma^2) - 1$ also enjoys the identical asymptotic normality. Note here, we cannot ensure its positive value based on the estimating equation given in (3.10).

The derivation of asymptotic distribution of the variance estimators in Model B follow a similar vein as in Model A. In addition, the covariance $E(\gamma_1\gamma_2)$ in Model B can also be represented in an iid form for its asymptotic distribution derivation. Following the delta method, we first derive the iid representation of $\sqrt{n}\left\{\widehat{F}_1(t_1)\widehat{F}_2(t_2) - F_1(t_1)F_2(t_2)\right\}$:

$$\begin{aligned}&\sqrt{n}\left\{\widehat{F}_1(t_1)\widehat{F}_2(t_2) - F_1(t_1)F_2(t_2)\right\}\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} F_1(t_1)F_2(t_2)\left\{b_{1i}(t_1) + b_{2i}(t_2)\right\} + op(1),\end{aligned} \tag{B.31}$$

where $1/\sqrt{n}\sum_{i=1}^{n} b_{1i}(t_1)$ and $1/\sqrt{n}\sum_{i=1}^{n} b_{2i}(t_2)$ are iid representations of $\sqrt{n}\{\widehat{F}_1(t)-F_1(t)\}$ and $\sqrt{n}\{\widehat{F}_2(t)-F_2(t)\}$ respectively. Thus one can easily see that

$$\begin{aligned}&\sqrt{n}\left\{\widehat{F}_1(t_1)^{-1}\widehat{F}_2(t_2)^{-1}-F_1(t_1)^{-1}F_2(t_2)^{-1}\right\}\\&=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}F_1(t_1)^{-1}F_2(t_2)^{-1}\left\{b_{1i}(t_1)+b_{2i}(t_2)\right\}+op(1).\end{aligned}\tag{B.32}$$

Therefore, the iid representation of the covariance of the two correlated frailties in Model B can be derived as

$$\begin{aligned}&\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}(m_{1i}m_{2i})\widehat{F}(c_{1i})^{-1}\widehat{F}(c_{2i})^{-1}-\Lambda_0^2(\tau)E[\exp\{(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)'\boldsymbol{Z}\}]E(\gamma_1\gamma_2)\right\}\\
&=\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}(m_{1i}m_{2i})\left\{\widehat{F}(c_{1i})^{-1}\widehat{F}(c_{2i})^{-1}-F(c_{1i})^{-1}F(c_{2i})^{-1}\right\}+\right.\\
&\qquad\qquad\left.\left\{m_{1i}m_{2i}F(c_{1i})^{-1}F(c_{2i})^{-1}-\Lambda_0^2(\tau)E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]E(\gamma^2)\right\}\right]\\
&=\sqrt{n}\left[\int m_1m_2\{\widehat{F}(c_{1i})^{-1}\widehat{F}(c_{2i})^{-1}-F(c_{1i})^{-1}F(c_{2i})^{-1}\}dH(m_1,m_2,c_1,c_2)+\right.\\
&\qquad\qquad\left.\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{m_{1i}m_{2i}}{F_1(c_{1i})F_2(c_{2i})}-\Lambda_{01}(\tau)\Lambda_{02}(\tau)E[\exp\{(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)'\boldsymbol{Z}\}]E(\gamma_1\gamma_2)\right\}\right]+op(1)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[-\int\frac{m_1m_2}{F(c_1)F(c_2)}\left\{b_{1i}(c_1)+b_{2i}(c_2)\right\}dH(m_1,m_2,c_1,c_2)+\right.\\
&\qquad\qquad\left.\left\{\frac{m_{1i}m_{2i}}{F_1(c_{1i})F_2(c_{2i})}-\Lambda_{01}(\tau)\Lambda_{02}(\tau)E[\exp\{(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)'\boldsymbol{Z}\}]E(\gamma_1\gamma_2)\right\}\right]\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}l_i,\end{aligned}\tag{B.33}$$

where $E(l_i)=0$ because $E(b_{1i})=E(b_{2i})=0$.

Moreover, following the similar derivation in (B.28), one is able to show that

$$
\begin{aligned}
&\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\exp\{(\widehat{\boldsymbol{\beta}}_1' + \widehat{\boldsymbol{\beta}}_2')\boldsymbol{z}_i\} - E\{\exp((\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2)'\boldsymbol{Z})\}\right] \\
&= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\exp\{(\widehat{\boldsymbol{\beta}}_1' + \widehat{\boldsymbol{\beta}}_2')\boldsymbol{z}_i\} - \exp\{(\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2')\boldsymbol{z}_i\}\right. \\
&\qquad\qquad \left.\exp\{(\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2')\boldsymbol{z}_i\} - E\{\exp((\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2)'\boldsymbol{Z})\}\right] \\
&= \sqrt{n}\left[E\left\{\exp((\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2)'\boldsymbol{Z})\boldsymbol{Z}'\right\}(\widehat{\boldsymbol{\beta}}_1 + \widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) + \right. \\
&\qquad\qquad \left.\frac{1}{n}\sum_{i=1}^{n}\exp\{(\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2')\boldsymbol{z}_i\} - E\{\exp((\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2)'\boldsymbol{Z})\}\right] + op(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} p_i + op(1),
\end{aligned}
\tag{B.34}
$$

where p_i are defined as

$$
\begin{aligned}
p_i =& E\left\{\exp((\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2)'\boldsymbol{Z})\boldsymbol{Z}'\right\}([\boldsymbol{\psi}_1^{-1}\boldsymbol{e}_{1i}]_{-1} + \boldsymbol{\psi}_2^{-1}\boldsymbol{e}_{2i}]_{-1}) + \\
&\qquad + \exp\{(\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2')\boldsymbol{z}_i\} - E\{\exp((\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2)'\boldsymbol{Z})\}.
\end{aligned}
\tag{B.35}
$$

Note that $\boldsymbol{e}_{1i}$ and $\boldsymbol{e}_{2i}$ are the iid representation of $\boldsymbol{\theta}_1 = [\ln\Lambda_{01}(\tau), \boldsymbol{\beta}_1']'$ and $\boldsymbol{\theta}_2 = [\ln\Lambda_{02}(\tau), \boldsymbol{\beta}_2']'$, and we also have $\boldsymbol{\psi}_1 = E\left[-\frac{\partial \boldsymbol{e}_{1i}}{\partial \boldsymbol{\theta}_1}\right]$ and $\boldsymbol{\psi}_2 = E\left[-\frac{\partial \boldsymbol{e}_{2i}}{\partial \boldsymbol{\theta}_2}\right]$.

With the additional two iid representations for the baseline rates for the two event types, i.e. $\sqrt{n}(\widehat{\Lambda}_{01}(\tau) - \Lambda_{01}(\tau)) = 1/\sqrt{n}\sum_{i=1}^{n}\Lambda_{01}(\tau)[\boldsymbol{\psi}_1^{-1}\boldsymbol{e}_{1i}]_1$ and $\sqrt{n}(\widehat{\Lambda}_{02}(\tau) - \Lambda_{02}(\tau)) = 1/\sqrt{n}\sum_{i=1}^{n}\Lambda_{02}(\tau)[\boldsymbol{\psi}_2^{-1}\boldsymbol{e}_{2i}]_1$, we obtain the iid representation for the product of two baseline estimators

$$
\begin{aligned}
&\sqrt{n}(\widehat{\Lambda}_{01}(\tau)\widehat{\Lambda}_{02}(\tau) - \Lambda_{01}(\tau)\Lambda_{02}(\tau)) \\
&= \frac{1}{\sqrt{n}}\Lambda_{01}(\tau)\Lambda_{02}(\tau)\sum_{i=1}^{n}([\psi_1^{-1}\boldsymbol{e}_{1i}]_1 + [\psi_1^{-1}\boldsymbol{e}_{2i}]_1) + op(1).
\end{aligned}
\tag{B.36}
$$

Follow the lines of (B.29) using the delta method, we obtain the iid representation of $\sqrt{n}\left(\widehat{E}(\gamma_1\gamma_2) - E(\gamma_1\gamma_2)\right)$ to be:

$$
\begin{aligned}
&\sqrt{n}\left(\widehat{E}(\gamma_1\gamma_2) - E(\gamma_1\gamma_2)\right)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg[\frac{l_i}{\Lambda_{01}(\tau)\Lambda_{02}(\tau)E\{\exp((\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)'\boldsymbol{Z})\}} - E(\gamma_1\gamma_2)\{[\boldsymbol{\psi}_1^{-1}\boldsymbol{e}_{1i}]_1 + [\boldsymbol{\psi}_2^{-1}\boldsymbol{e}_{2i}]_1\}\\
&\qquad\qquad - \frac{E(\gamma_1\gamma_2)p_i}{E\{\exp((\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)'\boldsymbol{Z})\}}\Bigg] + op(1)\\
&= \sum_{i=1}^{n} q_i + op(1),
\end{aligned}
\tag{B.37}
$$

where q_i forms the iid representation and $E(q_i) = 0$. Thus with assumptions A1-5, we conclude that $\sqrt{n}\left(\widehat{E}(\gamma_1\gamma_2) - E(\gamma_1\gamma_2)\right)$ converges weakly towards a mean-0 normal distribution with variance $E(q_i^2)$. The covariance term $cov(\gamma_1, \gamma_2) = E(\gamma_1\gamma_2) - 1$ is supposed to have the identical asymptotic normality.

Through a delta method, the asymptotic normality of the variance components in Model A; and the covariance and correlation coefficient $(\widehat{\rho})$ in Model B.

$$
\begin{aligned}
\sqrt{n}(\widehat{\rho} - \rho) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg[\frac{q_i}{\sqrt{Var(\gamma_1)Var(\gamma_2))}} -\\
&\qquad \frac{s_{1i}}{2Var(\gamma_1)^{\frac{3}{2}}Var(\gamma_2)^{\frac{1}{2}}} - \frac{s_{2i}}{2Var(\gamma_1)^{\frac{1}{2}}Var(\gamma_2)^{\frac{3}{2}}}\Bigg] + op(1)\\
&= \sum_{i=1}^{n} r_i + op(1),
\end{aligned}
\tag{B.38}
$$

where s_{1i} and s_{2i} are iid representation for the variance of γ_1 and γ_2 in Model B following (B.29). Henceforth, $\sqrt{n}(\widehat{\rho} - \rho)$ converges weakly to a mean-0 normal distribution with variance $E(r_i^2)$.

Now we continue to derive the consistency of the variance estimators in Model C. The asymptotic normality of $\sqrt{K}(\widehat{E}(\epsilon^2\gamma^2) - E(\epsilon^2\gamma^2))$ can be easily derived following the similar derivations in (B.25)-(B.29) by treating the sum of observations in each cluster/hospital as K iid observations and assuming that K is sufficiently large ($K \to \infty$).

$$\sqrt{K}(\widehat{E}(\epsilon^2\gamma^2) - E(\epsilon^2\gamma^2)) = \sum_{k=1}^{K} s_{ck} + op(1), \tag{B.39}$$

where we define

$$s_{ck} = \left\{ \frac{g_{ck}}{\nu\Lambda_0^2(\tau)E\left\{\exp(2\boldsymbol{\beta}'\boldsymbol{Z})\right\}} - 2E(\epsilon^2\gamma^2)[\boldsymbol{\psi}_c^{-1}\boldsymbol{e}_{ck}]_1 - \frac{E(\epsilon^2\gamma^2)h_{ck}}{\nu E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]} \right\}. \tag{B.40}$$

Note that $\boldsymbol{e}_{ck}$ is defined in (B.24), $\boldsymbol{\psi}_c = E\left[-\frac{\partial \boldsymbol{e}_{ck}}{\partial \boldsymbol{\theta}}\right]$, g_{ck} and h_{ck} are defined as below

$$g_{ck} = \sum_{i=1}^{I_k}(m_i^2 - m_i)F(c_i)^{-2} - \nu\int \frac{(m^2-m)2b_{ck}(c)}{F(c)^2}dH(m,c) - \nu\Lambda_0^2(\tau)E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})]E(\epsilon^2\gamma^2); \tag{B.41}$$

$$h_{ck} = \sum_{i=1}^{I_k}\exp(2\boldsymbol{\beta}'\boldsymbol{z}_i) + 2\nu E[\exp(2\boldsymbol{\beta}'\boldsymbol{Z})\boldsymbol{Z}'][\boldsymbol{\psi}^{-1}\boldsymbol{e}_{ck}]_{-1} - \nu E(\exp(2\boldsymbol{\beta}'\boldsymbol{Z})). \tag{B.42}$$

The iid representation in (B.39) and its convergence to a mean-0 normal distribution suggests that $\widehat{E}(\epsilon^2\gamma^2) \to_p E(\epsilon^2\gamma^2)$. Moreover, the iid representations for the following equations

$$\begin{aligned}&\sqrt{K}\left[\frac{1}{K}\sum_{k=1}^{K}\left\{\sum_{i=1}^{I_k}(m_{ki}^2 - m_{ki})\widehat{F}^{-2}(c_{ki})\right\} - \nu\Lambda_0^2(\tau)E\{\exp(2\boldsymbol{\beta}'\boldsymbol{Z})E(\epsilon^2\gamma^2)\}\right]\\ &= \frac{1}{\sqrt{K}}\sum_{k=1}^{K} g_{ck} + op(1)\end{aligned} \tag{B.43}$$

and

$$\sqrt{K}(\widehat{\Lambda}_0^2(\tau) - \Lambda_0^2(\tau)) = \frac{2\Lambda_0^2(\tau)}{\sqrt{K}}\sum_{k=1}^{K}[\boldsymbol{\psi}_c^{-1}\boldsymbol{e}_{ck}]_1 + op(1), \tag{B.44}$$

also imply that $\frac{1}{K}\sum_{k=1}^{K}\left\{\sum_{i=1}^{I_k}(m_{ki}^2 - m_{ki})\widehat{F}^{-2}(c_{ki})\right\} \rightarrow_p \nu\Lambda_0^2(\tau)E\{\exp(2\boldsymbol{\beta}'\boldsymbol{Z})E(\epsilon^2\gamma^2)\}$ and $\widehat{\Lambda}_0^2(\tau) \rightarrow_p \Lambda_0^2(\tau)$. Or equivalently, one may simply refer to the weak law of large numbers and the continuous mapping to obtain these conclusions directly.

In order to identify the variance from the two different sources in Model C, we plug in the borrow-strength estimator $\widehat{\epsilon}_k = \sum_{i=1}^{I_k}[m_{ki}/\widehat{F}(c_{ki})]/\widehat{\Lambda}_0(\tau)/\sum_{i=1}^{I_k} e^{\widehat{\boldsymbol{\beta}}' z_{ki}}$, and it can be shown that within each facility (indexed by k), $\widehat{\epsilon}_k = \epsilon_k + Op(I_k^{-\frac{1}{2}})$. Note that this is with respect to the conditional probability measure $P(\cdot \mid \{\epsilon_k, k = 1, \ldots, K\})$. The conditional convergence can also be shown to satisfy in the marginal probability measure (not conditional on ϵ_k) using the dominant convergence theorem. When I_k are large, it is not hard to show the convergence in the conditional probability $\widehat{E}(\gamma^2) \rightarrow_p E(\gamma^2)$, which also implies $\widehat{E}(\epsilon^2) \rightarrow_p E(\epsilon^2)$. One can imagine that, the accuracy of the estimation for the variance components is largely dependent on the estimation quality of $\widehat{\epsilon}_k$, and thus the facility size I_k.

An alternative method is to estimate $E(\epsilon^2)$ using a U-statistic method and thus obtain the estimation of $E(\gamma^2)$ directly, instead of using the borrow-strength estimator given in (3.14). The estimator of $E(\epsilon^2)$ derived from (3.16) is

$$\widehat{E}(\epsilon^2) = \frac{\sum_{k=1}^{K} I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}} m_{ki}m_{kj}F^{-1}(C_{ki})F^{-1}(C_{kj})}{\sum_{k=1}^{K} I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}} \exp(\widehat{\boldsymbol{\beta}}'(\boldsymbol{z}_{ki} + \boldsymbol{z}_{kj}))\widehat{\Lambda}_0^2(\tau)}. \tag{B.45}$$

To obtain the iid representation of (B.45), we need first derive the iid representations that are similar to those in (B.43) and (B.44).

$$
\begin{aligned}
&\sqrt{K}\left[\frac{1}{K}\sum_{k=1}^{K}\left\{I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}} m_{ki}m_{kj}\widehat{F}^{-1}(c_{ki})\widehat{F}^{-1}(c_{kj})\right\} - \right. \\
&\qquad\qquad\qquad\qquad \left. \omega E\{\exp(\boldsymbol{\beta}'(\boldsymbol{Z}_1+\boldsymbol{Z}_2))\}\Lambda_0^2(\tau)E(\epsilon^2)\right] \\
&= \sqrt{K}\left[\omega\int m_1m_2\left\{\widehat{F}^{-1}(c_1)\widehat{F}^{-1}(c_2) - F^{-1}(c_1)F^{-1}(c_2)\right\}dH(m_1,m_2,c_1,c_2)+\right. \\
&\left.\frac{1}{K}\sum_{k=1}^{K} I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}} m_{ki}m_{kj}F^{-1}(C_{ki})F^{-1}(C_{kj}) - \omega E\{\exp(\boldsymbol{\beta}'(\boldsymbol{Z}_1+\boldsymbol{Z}_2))\}\Lambda_0^2(\tau)E(\epsilon^2)\right] + op(1) \\
&= \frac{1}{\sqrt{K}}\sum_{k=1}^{K}\left[-\omega\int\frac{m_1m_2}{F(c_1)F(c_2)}\{b_k(c_1)+b_k(c_2)\}dH(m_1,m_2,c_1,c_2)+\right. \\
&\left. I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}} m_{ki}m_{kj}F^{-1}(C_{ki})F^{-1}(C_{kj}) - \omega E\{\exp(\boldsymbol{\beta}'(\boldsymbol{Z}_1+\boldsymbol{Z}_2))\}\Lambda_0^2(\tau)E(\epsilon^2)\right] + op(1) \\
&= \frac{1}{\sqrt{K}}\sum_{k=1}^{K} u_k + op(1),
\end{aligned}
\tag{B.46}
$$

where $\omega = E\{I(I_k \geq 2)I_k(I_k-1)/2\}$, $E(u_k) = 0$ and it is defined to be

$$
\begin{aligned}
&u_k = -\omega\int\frac{m_1m_2}{F(c_1)F(c_2)}\{b_k(c_1)+b_k(c_2)\}dH(m_1,m_2,c_1,c_2)+ \\
&I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}} m_{ki}m_{kj}F^{-1}(C_{ki})F^{-1}(C_{kj}) - \omega E\{\exp(\boldsymbol{\beta}'(\boldsymbol{Z}_i+\boldsymbol{Z}_j))\}\Lambda_0^2(\tau)E(\epsilon^2).
\end{aligned}
$$

We follow the similar arguments in (B.28) and (B.34), assuming that K is sufficiently large, i.e. $K \to \infty$. Then we can obtain the following iid representation for the exponential covariate part:

$$\begin{aligned}
&\sqrt{K}\left[\frac{1}{K}\sum_{k=1}^{K}\left\{I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}}\exp(\widehat{\boldsymbol{\beta}}'(\boldsymbol{z}_{ki}+\boldsymbol{z}_{kj}))\right\} - \omega E\left\{\exp(\boldsymbol{\beta}'(\boldsymbol{Z}_1+\boldsymbol{Z}_2))\right\}\right] \\
&= \sqrt{K}\left[\frac{1}{K}\sum_{k=1}^{K} I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}}\left\{\exp(\widehat{\boldsymbol{\beta}}'(\boldsymbol{z}_{ki}+\boldsymbol{z}_{kj})) - \exp(\boldsymbol{\beta}'(\boldsymbol{z}_{ki}+\boldsymbol{z}_{kj}))\right\} + \right. \\
&\qquad \left. \frac{1}{K}\sum_{k=1}^{K} I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}}\exp(\boldsymbol{\beta}'(\boldsymbol{z}_{ki}+\boldsymbol{z}_{kj})) - \omega E\left\{\exp(\boldsymbol{\beta}'(\boldsymbol{Z}_1+\boldsymbol{Z}_2))\right\}\right] \\
&= \sqrt{K}\left[\frac{1}{K}\sum_{k=1}^{K}\left\{I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}}\exp(\boldsymbol{\beta}'(\boldsymbol{z}_{ki}+\boldsymbol{z}_{kj}))(\boldsymbol{z}_{ki}+\boldsymbol{z}_{kj})\right\}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + \right. \\
&\qquad \left. \frac{1}{K}\sum_{k=1}^{K} I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}}\exp(\boldsymbol{\beta}'(\boldsymbol{z}_{ki}+\boldsymbol{z}_{kj})) - \omega E\left\{\exp(\boldsymbol{\beta}'(\boldsymbol{Z}_1+\boldsymbol{Z}_2))\right\}\right] + op(1) \\
&= \sqrt{K}\left[\omega E\left\{\exp(\boldsymbol{\beta}'(\boldsymbol{Z}_1+\boldsymbol{Z}_2))(\boldsymbol{Z}_1+\boldsymbol{Z}_2)'\right\}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + \right. \\
&\qquad \left. \frac{1}{K}\sum_{k=1}^{K} I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}}\exp(\boldsymbol{\beta}'(\boldsymbol{z}_{ki}+\boldsymbol{z}_{kj})) - \omega E\left\{\exp(\boldsymbol{\beta}'(\boldsymbol{Z}_1+\boldsymbol{Z}_2))\right\}\right] + op(1) \\
&= \frac{1}{\sqrt{K}}\sum_{k=1}^{K}\left[\omega E\left\{\boldsymbol{\beta}'(\boldsymbol{Z}_1+\boldsymbol{Z}_2)(\boldsymbol{Z}_1+\boldsymbol{Z}_2)'\right\}[\boldsymbol{\psi}_c^{-1}\boldsymbol{e}_{ck}]_{-1} + \right. \\
&\qquad\qquad \left. I(I_k \geq 2)\sum_{(i\neq j)\in \boldsymbol{C}_{2,I_k}}\exp(\boldsymbol{\beta}'(\boldsymbol{z}_{ki}+\boldsymbol{z}_{kj})) - \omega E\left\{\exp(\boldsymbol{\beta}'(\boldsymbol{Z}_1+\boldsymbol{Z}_2))\right\}\right] + op(1) \\
&= \frac{1}{\sqrt{K}}\sum_{k=1}^{K} v_k + op(1),
\end{aligned} \tag{B.47}$$

where due to the independence between covariates from difference individuals, $W(\boldsymbol{z}_1, \boldsymbol{z}_2) = U(\boldsymbol{z}_1)U(\boldsymbol{z}_2)$, and

$$E\left\{\exp(\boldsymbol{\beta}'(\boldsymbol{Z}_1+\boldsymbol{Z}_2))\right\} = E(\exp(\boldsymbol{\beta}'\boldsymbol{Z}))^2.$$

Moreover, we can easily show that $E(v_k) = 0$ and

$$v_k = \omega E\left\{\boldsymbol{\beta}'(\boldsymbol{Z}_1 + \boldsymbol{Z}_2)(\boldsymbol{Z}_1 + \boldsymbol{Z}_2)'\right\}[\boldsymbol{\psi}_c^{-1}\boldsymbol{e}_{ck}]_{-1} +$$
$$I(I_k \geq 2) \sum_{(i \neq j) \in \boldsymbol{C}_{2,I_k}} \exp(\boldsymbol{\beta}'(\boldsymbol{z}_{ki} + \boldsymbol{z}_{kj})) - \omega E\left\{\exp(\boldsymbol{\beta}'(\boldsymbol{Z}_1 + \boldsymbol{Z}_2))\right\}.$$

Now combining the results in (B.44), (B.46) and (B.47), we obtain the iid representation for the U-statistic estimator of $E(\epsilon^2)$ via the delta method when K is large:

$$\begin{aligned}
&\sqrt{K}(\widehat{E}(\epsilon^2) - E(\epsilon^2)) \\
&= \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \left\{ \frac{u_k}{\omega \Lambda_0^2(\tau) E\left\{\exp(\boldsymbol{\beta}'\boldsymbol{Z})\right\}^2} - 2E(\gamma^2)[\boldsymbol{\psi}_c^{-1}\boldsymbol{e}_{ck}]_1 - \frac{E(\gamma^2)v_k}{\omega E\{\exp(\boldsymbol{\beta}'\boldsymbol{Z})\}^2} \right\} + op(1) \\
&= \frac{1}{\sqrt{K}} \sum_{i=1}^{n} w_k + op(1).
\end{aligned} \tag{B.48}$$

Henceforth, with the iid representation in (B.39), the iid representation for $\widehat{E}(\gamma^2) = \widehat{E}(\epsilon^2\gamma^2)/\widehat{E}(\epsilon^2)$ is given by

$$\begin{aligned}
\sqrt{K}\left(\widehat{E}(\gamma^2) - E(\gamma^2)\right) &= \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \left[\frac{1}{E(\epsilon^2)} s_{ck} - \frac{E(\gamma^2)}{E(\epsilon^2)} w_k \right] + op(1) \\
&= \sum_{k=1}^{K} y_k + op(1),
\end{aligned} \tag{B.49}$$

where

$$y_k = \frac{1}{E(\epsilon^2)} s_{ck} - \frac{E(\gamma^2)}{E(\epsilon^2)} w_k.$$

Since it is straightforward to show that $E(w_k) = E(y_k) = 0$, we conclude that the U-statistic estimators converge weakly towards mean-0 normal distributions as given in (3.4).

B.6 Proof of Asymptotic Properties for $\widehat{\boldsymbol{\theta}}$ from GMM

The objective function (B.6) implies that

$$\boldsymbol{Q}_n(\boldsymbol{\theta}) = \boldsymbol{G}_n(\boldsymbol{\theta})\boldsymbol{W}_n\boldsymbol{g}_n(\boldsymbol{\theta}) = \boldsymbol{0}, \tag{B.50}$$

where

$$\boldsymbol{G}_n(\boldsymbol{\theta}) = \frac{d\boldsymbol{g}_n^T(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \frac{d\widehat{\boldsymbol{g}}_n^T(\boldsymbol{\theta})}{d\boldsymbol{\theta}}. \tag{B.51}$$

The objective function (B.8) after plugging in the shape function estimates $\widehat{F}(t)$ imples that

$$\widehat{\boldsymbol{Q}}_n(\boldsymbol{\theta}) = \boldsymbol{G}_n(\boldsymbol{\theta})\boldsymbol{W}_n\widehat{\boldsymbol{g}}_n(\boldsymbol{\theta}) = \boldsymbol{0}. \tag{B.52}$$

Let

$$\boldsymbol{g}(\boldsymbol{\theta}) = E\left[\boldsymbol{g}_n(\boldsymbol{\theta})\right] = E\left[\boldsymbol{g}(\boldsymbol{X};\boldsymbol{\theta})\right] \tag{B.53}$$

$$\boldsymbol{G}(\boldsymbol{\theta}) = E\left[\boldsymbol{G}_n(\boldsymbol{\theta})\right] = \frac{d\boldsymbol{g}^T(\boldsymbol{\theta})}{d\boldsymbol{\theta}}. \tag{B.54}$$

By the weak law of large numbers, we shall have $\boldsymbol{g}_n(\boldsymbol{\theta}) \rightarrow_p \boldsymbol{g}(\boldsymbol{\theta})$ and $\boldsymbol{G}_n(\boldsymbol{\theta}) \rightarrow_p \boldsymbol{G}(\boldsymbol{\theta})$. In addition, we define $\boldsymbol{Q}(\boldsymbol{\theta}) = \boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{W}\boldsymbol{g}(\boldsymbol{\theta})$.

Let $\boldsymbol{\theta}_0$ be the unique solution satisfying $\boldsymbol{Q}(\boldsymbol{\theta}_0) = 0$, and $\widehat{\boldsymbol{\theta}}$ be the GMM estimator satisfying $\widehat{\boldsymbol{Q}}_n(\widehat{\boldsymbol{\theta}}) = 0$. Note that in Appendix B.3 we show that $\widehat{F}(t) \rightarrow_p F(t)$, thus $\sup_{\boldsymbol{\theta}} |\widehat{\boldsymbol{Q}}_n(\boldsymbol{\theta}) - \boldsymbol{Q}_n(\boldsymbol{\theta})| \rightarrow_p 0$.

$$\begin{aligned} 0 &\le \sup_{\boldsymbol{\theta}} |\widehat{\boldsymbol{Q}}_n(\boldsymbol{\theta}) - \boldsymbol{Q}(\boldsymbol{\theta})| \le \sup_{\boldsymbol{\theta}} |\widehat{\boldsymbol{Q}}_n(\boldsymbol{\theta}) - \boldsymbol{Q}_n(\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}} |\boldsymbol{Q}_n(\boldsymbol{\theta}) - \boldsymbol{Q}(\boldsymbol{\theta})| \\ &\le \sup_{\boldsymbol{\theta}} |\widehat{\boldsymbol{Q}}_n(\boldsymbol{\theta}) - \boldsymbol{Q}_n(\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}} |\boldsymbol{G}_n(\boldsymbol{\theta})\boldsymbol{W}_n\boldsymbol{g}_n(\boldsymbol{\theta}) - \boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{W}_n\boldsymbol{g}_n(\boldsymbol{\theta})| + \\ &\sup_{\boldsymbol{\theta}} |\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{W}_n\boldsymbol{g}_n(\boldsymbol{\theta}) - \boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{W}\boldsymbol{g}_n(\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}} |\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{W}\boldsymbol{g}_n(\boldsymbol{\theta}) - \boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{W}\boldsymbol{g}(\boldsymbol{\theta})| \rightarrow_p 0. \end{aligned} \tag{B.55}$$

Thus we have $\widehat{\boldsymbol{\theta}} \rightarrow_p \boldsymbol{\theta}$.

Let $\widehat{\boldsymbol{\theta}}$ denote the solution for

$$0 = \widehat{\boldsymbol{Q}}_n(\widehat{\boldsymbol{\theta}}) = \boldsymbol{Q}_n(\widehat{\boldsymbol{\theta}}) + \left[\widehat{\boldsymbol{Q}}_n(\widehat{\boldsymbol{\theta}}) - \boldsymbol{Q}_n(\widehat{\boldsymbol{\theta}})\right] \tag{B.56}$$

For the second part of (B.56)

$$\begin{aligned}
&\left[\widehat{\boldsymbol{g}}_n(\widehat{\boldsymbol{\theta}}) - \boldsymbol{g}_n(\widehat{\boldsymbol{\theta}})\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \begin{bmatrix} w_{1i}\bar{\boldsymbol{x}}_{1i}m_i\left(\frac{1}{\widehat{F}(c_i)} - \frac{1}{F(c_i)}\right) \\ w_{2i}\bar{\boldsymbol{x}}_{2i}(m_i^2 - m_i)\left(\frac{1}{\widehat{F}^2(c_i)} - \frac{1}{F^2(c_i)}\right) \end{bmatrix} \\
&= \int \begin{bmatrix} w_1\bar{\boldsymbol{x}}_1 m\left(\frac{1}{\widehat{F}(c)} - \frac{1}{F(c)}\right) \\ w_2\bar{\boldsymbol{x}}_2(m^2 - m)\left(\frac{1}{\widehat{F}^2(c)} - \frac{1}{F^2(c)}\right) \end{bmatrix} dV(w, \boldsymbol{x}, m, y) + op(n^{-\frac{1}{2}}) \\
&= -\int \begin{bmatrix} w_1\bar{\boldsymbol{x}}_1 m\left(\frac{\widehat{F}(c)-F(c)}{F^2(c)}\right) \\ 2w_2\bar{\boldsymbol{x}}_2(m^2 - m)\left(\frac{\widehat{F}(c)-F(c)}{F^3(c)}\right) \end{bmatrix} dV(w, \boldsymbol{x}, m, y) + op(n^{-\frac{1}{2}}) \\
&= -\frac{1}{n}\sum_{i=1}^{n}\int \begin{bmatrix} w_1\bar{\boldsymbol{x}}_1 m\left(\frac{b_i(c)}{F(c)}\right) \\ 2w_2\bar{\boldsymbol{x}}_2(m^2 - m)\left(\frac{b_i(c)}{F^2(c)}\right) \end{bmatrix} dV(w, \boldsymbol{x}, m, y) + op(n^{-\frac{1}{2}}) \\
&= -\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{q}_i + op(n^{-\frac{1}{2}}),
\end{aligned} \tag{B.57}$$

where

$$\boldsymbol{q}_i = \int \begin{bmatrix} w_1\bar{\boldsymbol{x}}_1 m\left(\frac{b_i(c)}{F(c)}\right) \\ 2w_2\bar{\boldsymbol{x}}_2(m^2 - m)\left(\frac{b_i(c)}{F^2(c)}\right) \end{bmatrix} dV(w, \boldsymbol{x}, m, y)$$

and $E(\boldsymbol{q}_i) = \boldsymbol{0}$

Thus we have

$$\begin{aligned}
\left[\widehat{\boldsymbol{Q}}_n(\widehat{\boldsymbol{\theta}}) - \boldsymbol{Q}_n(\widehat{\boldsymbol{\theta}})\right] &= \boldsymbol{G}_n(\widehat{\boldsymbol{\theta}})\boldsymbol{W}_n\left[\widehat{\boldsymbol{g}}_n(\widehat{\boldsymbol{\theta}}) - \boldsymbol{g}_n(\widehat{\boldsymbol{\theta}})\right] \\
&= \boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{W}\left[\widehat{\boldsymbol{g}}_n(\widehat{\boldsymbol{\theta}}) - \boldsymbol{g}_n(\widehat{\boldsymbol{\theta}})\right] + op(n^{-\frac{1}{2}}) \\
&= -\boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{W}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{q}_i + op(n^{-\frac{1}{2}})
\end{aligned} \tag{B.58}$$

For the first part of (B.56), we conduct the normal proof for GMM

$$\begin{aligned}\boldsymbol{Q}_n(\widehat{\boldsymbol{\theta}}) = \boldsymbol{Q}_n(\boldsymbol{\theta}_0)+ \\ \left[\boldsymbol{G}_n(\boldsymbol{\theta}^*)\boldsymbol{W}_n\boldsymbol{G}_n(\boldsymbol{\theta}^*)^T + \frac{\partial G_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}}\boldsymbol{W}_n\boldsymbol{g}_n(\boldsymbol{\theta}^*)\right](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),\end{aligned} \tag{B.59}$$

where $\boldsymbol{\theta}^*$ is between $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$, thus $\boldsymbol{\theta}^* \rightarrow_p \boldsymbol{\theta}_0$. And since $\boldsymbol{g}_n(\boldsymbol{\theta}_0) \rightarrow_p 0$, we obtain

$$\boldsymbol{Q}_n(\widehat{\boldsymbol{\theta}}) = \boldsymbol{Q}_n(\boldsymbol{\theta}_0) + \left[\boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{W}\boldsymbol{G}(\boldsymbol{\theta}_0)^T\right](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + op(n^{-\frac{1}{2}}) \tag{B.60}$$

Combining (B.57) and (B.59), we end up with

$$\begin{aligned}\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 &= \left[\boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{W}\boldsymbol{G}(\boldsymbol{\theta}_0)^T\right]^{-1}\frac{1}{n}\sum_{i=1}^{n}\{\boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{W}\boldsymbol{q}_i - \boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{W}\boldsymbol{g}(\boldsymbol{x}_i;\boldsymbol{\theta}_0)\} \\ &= \left[\boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{W}\boldsymbol{G}(\boldsymbol{\theta}_0)^T\right]^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{W}\left\{\boldsymbol{q}_i - \boldsymbol{g}(\boldsymbol{x}_i;\boldsymbol{\theta}_0)\right\}.\end{aligned} \tag{B.61}$$

Since $E(\boldsymbol{q}_i) = E(\boldsymbol{g}(\boldsymbol{x}_i;\boldsymbol{\theta}_0)) = 0$, $\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$ converges weakly to a mean-0 multivariate normal distribution. Assume that we have p parameters for $\boldsymbol{\beta}$, thus the length for $\boldsymbol{\theta}$ is $p + 2$. Through the delta method, we can also prove that the variance estimator $\widehat{\sigma_z^2} = \exp(\widehat{\boldsymbol{\theta}}[p + 2] - 2\widehat{\boldsymbol{\theta}}[1]) - 1$ also follows some asymptotic normality.

APPENDIX C

Multi-state Rate Models to Assess the Impact of Exposure to Lead on Children Behaviors Using Accelerometer Data

C.1 Additional Summary Figures and Table

Figures C.1 and C.2 represent corresponding plots in Figures 4.2 and 4.3 but use only the complete 333 subjects with no missingness in their variables. The almost identical plots using the two batches of data suggest a random missing data mechanism.

In Table C.1, the Subject columns list the total counts of subjects that experienced at least one of the corresponding transitions, while the Case columns provide the total counts of respective transitions, with or without stratified by gender. Thus the the proportion plot on the top panel of Figure 4.5 presents the ratios between the Subject column and the sample size n for each gender, and the average counts at the bottom give the ratios between the Case and Subject columns, representing average transition frequencies among those who have experienced at least one time of the corresponding transition.

Figure C.1: Marginal proportions of individual activity states (left) and the distribution of average daily transitions counts for each subject (right) using the 333 subjects with complete explanatory variables. Note that in the right panel, the salmon boxes denote transitions with increased activities (labeled by "+"), while the cyan boxes denote transitions with decreased activities (labeled by "-"). The activity states include sedentary (0), slightly active (1), moderately active (2), and vigorously active (3) statuses.

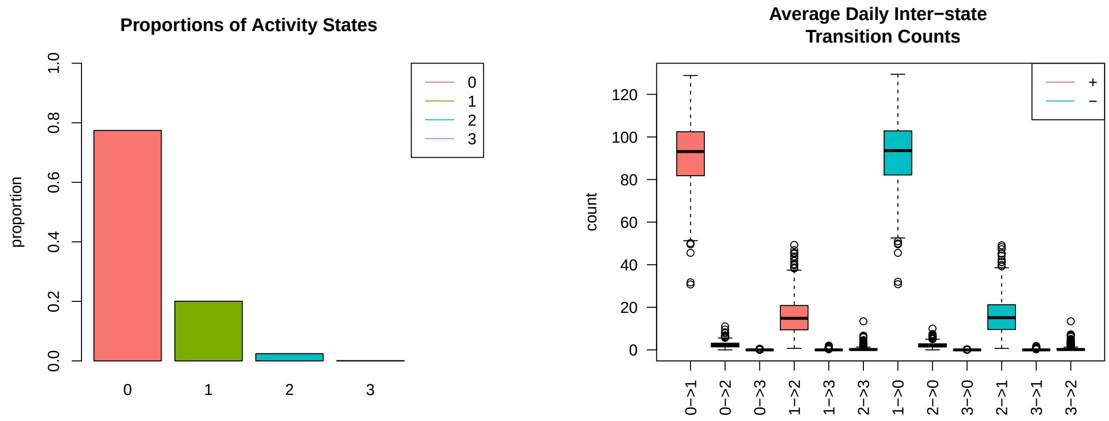


Figure C.2: The left penal shows the conditional proportions of transitions while the right panel describes the conditional proportions of the transition directions using the 333 subjects with complete explanatory variables. Their x-axes correspond to previous states and the y-axes represent the transition proportions. The activity states include sedentary (0), slightly active (1), moderately active (2), and vigorously active (3). The transition directions include increased ("+"), decreased("-"), and remained ("stay").

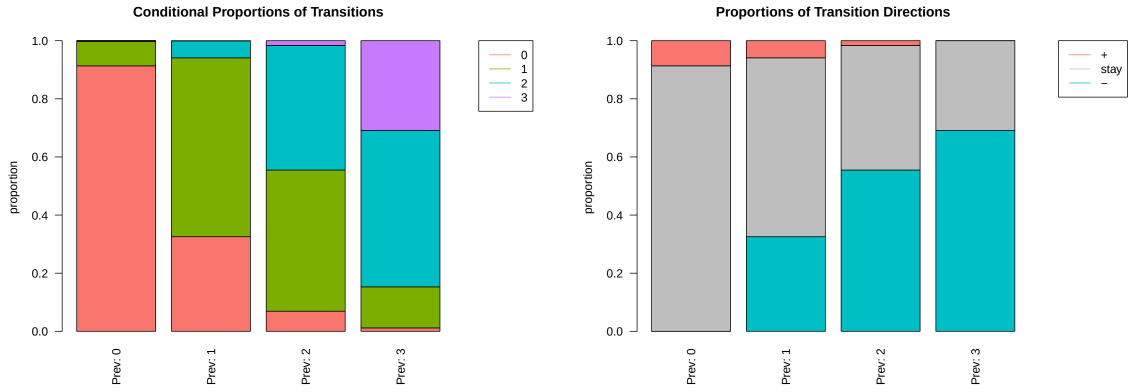


Table C.1: A summary table for the number of subjects with at least one corresponding transitions (Subject) and the total number of transitions (Cases). Their respective counts for each gender are also presented.

	Total (n = 333)		Boys (n = 170)		Girls (n = 163)	
Transition	Subject	Cases	Subject	Cases	Subject	Cases
0 → 1	333	8470	170	107257	163	111371
0 → 2	331	221025	169	3208	162	2865
0 → 3	50	227035	28	37	22	26
1 → 2	333	219753	170	19773	163	19531
1 → 3	165	258682	101	251	64	124
2 → 3	183	45449	107	979	76	349
1 → 0	333	739986	170	107457	163	111614
2 → 0	333	41116	170	3012	163	2649
3 → 0	29	1736	20	21	9	9
2 → 1	333	7010	170	20013	163	19754
3 → 1	156	1405	91	222	65	139
3 → 2	190	391	115	1024	75	351

APPENDIX D

An Epidemiological Forecast Model and Software Assessing Interventions on the COVID-19 Epidemic

D.1 Runga-Kutta Approximation

D.1.1 Approximation in the Basic SIR Model

The forth order Runga-Kutta(RK4) method gives an approximate of $f(\boldsymbol{\theta}_{t-1}, \beta, \gamma)$ in equation (5.4) as follows:

$$f(\boldsymbol{\theta}_{t-1}, \beta, \gamma) = \begin{pmatrix} \theta^S_{t-1} + 1/6[k^{S_1}_{t-1} + 2k^{S_2}_{t-1} + 2k^{S_3}_{t-1} + k^{S_4}_{t-1}] \\ \theta^I_{t-1} + 1/6[k^{I_1}_{t-1} + 2k^{I_2}_{t-1} + 2k^{I_3}_{t-1} + k^{I_4}_{t-1}] \\ \theta^R_{t-1} + 1/6[k^{R_1}_{t-1} + 2k^{R_2}_{t-1} + 2k^{R_3}_{t-1} + k^{R_4}_{t-1}] \end{pmatrix} := \begin{pmatrix} \alpha_{1(t-1)} \\ \alpha_{2(t-1)} \\ \alpha_{3(t-1)} \end{pmatrix},$$

where

$$
\begin{aligned}
k_t^{S_1} &= -\beta\theta_t^S\theta_t^I,\\
k_t^{S_2} &= -\beta[\theta_t^S + 0.5k_t^{S_1}][\theta_t^I + 0.5k_t^{I_1}],\\
k_t^{S_3} &= -\beta[\theta_t^S + 0.5k_t^{S_2}][\theta_t^I + 0.5k_t^{I_2}],\\
k_t^{S_4} &= -\beta[\theta_t^S + k_t^{S_3}][\theta_t^I + k_t^{I_3}];
\end{aligned}
$$

$$
\begin{aligned}
k_t^{I_1} &= \beta\theta_t^S\theta_t^I - \gamma\theta_t^I,\\
k_t^{I_2} &= \beta[\theta_t^S + 0.5k_t^{S_1}][\theta_t^I + 0.5k_t^{I_1}] - \gamma[\theta_t^I + 0.5k_t^{I_1}],\\
k_t^{I_3} &= \beta[\theta_t^S + 0.5k_t^{S_2}][\theta_t^I + 0.5k_t^{I_2}] - \gamma[\theta_t^I + 0.5k_t^{I_2}],\\
k_t^{I_4} &= \beta[\theta_t^S + k_t^{S_3}][\theta_t^I + k_t^{I_3}] - \gamma[\theta_t^I + k_t^{I_3}];
\end{aligned}
$$

and

$$
\begin{aligned}
k_t^{R_1} &= \gamma\theta_t^I,\\
k_t^{R_2} &= \gamma[\theta_t^I + 0.5k_t^{I_1}],\\
k_t^{R_3} &= \gamma[\theta_t^I + 0.5k_t^{I_2}],\\
k_t^{R_4} &= \gamma[\theta_t^I + k_t^{I_3}].
\end{aligned}
$$

D.1.2 Approximation in the Extended SIR Model with Quarantine Compartment

Using the RK4 approximation, $f(\boldsymbol{\theta}_{t-1}, \beta, \gamma)$ in the extended SIR model (5.6) with a quarantine compartment can be approximated following the two iterative steps:

1. Solve the $f(\boldsymbol{\theta}_{t-1}, \beta, \gamma)$ in Appendix D.1 without considering the quarantine with $f(\cdot)$
$$f(\boldsymbol{\theta}_{t-1}, \beta, \gamma) = [\alpha_{1(t-1)}, \alpha_{2(t-1)}, \alpha_{3(t-1)}]^{\mathrm{T}}.$$

2. Due to the quarantine, we deduct the susceptible by $\alpha^*_{1(t-1)} = \alpha_{1(t-1)} - \phi(t)\theta^S_{t-1}$, and let $\theta^Q_t = \theta^Q_{t-1} + \phi(t)\theta^S_{t-1}$ with $\theta^Q_0 = 0$.

Let $\boldsymbol{\alpha}^*_{t-1} = [\alpha^*_{1(t-1)}, \alpha_{2(t-1)}, \alpha_{3(t-1)}]^{\mathrm{T}}$, and it is easy to show that the sum $\sum_{k=1}^{3} \alpha^*_{k(t-1)} = 1 - \theta^Q_t$. Thus we can regenerate the next day's $\boldsymbol{\theta}_t$ following a Dirichlet distribution adjusted by the prevalence of the quarantine compartment $\boldsymbol{\alpha}^*_t \sim \text{Dirichlet}(\kappa\boldsymbol{\alpha}^*_{t-1}/(1 - \theta^Q_t))$. The estimated prevalence values become $\boldsymbol{\theta}_t = (1 - \theta^Q_t)\boldsymbol{\alpha}^*_t$. We follow above two steps and finish the complete prevalence processes. Note that the deduction of susceptible compartments might cause $\theta^S_t \leq 0$, we will bound such prevalence value to be consistently 0, which is equivalent to terminating transmission among susceptible subjects.

D.2 Moment Properties of Beta and Dirichlet Distributions

For the sake of being self-contained, we list the moments of both Beta and Dirichlet distributions. The mean and variance of Beta distribution Beta(α, β) are respectively:

$$\text{Mean} = \frac{\alpha}{\alpha + \beta}, \text{Var} = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

While to Dirchlet distribution $\mathrm{Dir}(\kappa\boldsymbol{\alpha})$, we have

$$\text{Mean} = \boldsymbol{\alpha}, \text{Var} = \frac{1}{\kappa+1}\begin{pmatrix} \alpha_1(1-\alpha_1) & -\alpha_1\alpha_2 & -\alpha_1\alpha_3 & -\alpha_1\alpha_4 \\ -\alpha_1\alpha_2 & \alpha_2(1-\alpha_2) & -\alpha_2\alpha_3 & -\alpha_2\alpha_4 \\ -\alpha_1\alpha_3 & -\alpha_2\alpha_3 & \alpha_3(1-\alpha_3) & -\alpha_3\alpha_4 \\ -\alpha_1\alpha_4 & -\alpha_2\alpha_4 & -\alpha_3\alpha_4 & \alpha_4(1-\alpha_4) \end{pmatrix},$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^{\mathrm{T}}$ with $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$.

D.3 R Codes

First we conducted analysis of the Hubei COVID-19 data using the transmission rate modifier with function `txt.eSIR` from package `eSIR`. Note that option `dic=TRUE` enables to calculate the deviance information criterion (DIC) for model selection, while options, `save_files=TRUE` and `save_mcmc`, allow the storage of MCMC output tables, plots, summary statistics and even full MCMC draws, which may be saved via the path of `file_add`, or otherwise via the current working directory. The major results returned from the package include predicted cumulative proportions, predicted turning points of interest, two ggplot2 (*Wickham*, 2016) objects of the summary plots related to both infection and removed compartments, a summary output table containing all the posterior means, median and credible intervals of the model parameters, and DIC if opted. The trace-plots and other useful diagnostic plots are also provided and automatically saved if `save_files=TRUE` is opted. In the package, function `tvt.eSIR()` works on the epidemiological model with time-varying transmission rate in Section 5.2.2, and `qh.eSIR()` for the other epidemiological model with a quarantine compartment in Section 5.2.3. Note that in function `tvt.eSIR()`, with a choice of `exponential=FALSE`, a step function is run in the MCMC when both `change_time` and `pi0` are specified. To fit the model with a continuous transmission rate modifier function, user may set `exponential=TRUE` and specify a value of `lambda0`. The default is to run the basic epidemiological model with no quarantine or $\pi(t) \equiv 1$ in Section 5.2.1. `death_in_R` is usually set to be the average ratio of death and removed proportions at each observation time point, which is used to estimate the death curve in the forecast plot of the removed compartment. Below are the R scripts used in the analysis.

```
### Example 1: Step function pi(t)
### Y and R are observed proportions of infected and removed compartments
change_time <- c("01/23/2020", "02/04/2020", "02/08/2020")
pi0 <- c(1.0, 0.9, 0.5, 0.1)
res.step <- tvt.eSIR (Y, R, begin_str = "01/13/2020", death_in_R = 0.4,
T_fin = 200, pi0 = pi0, change_time = change_time, dic = TRUE,
casename = "Hubei_step", save_files = TRUE,
save_mcmc = FALSE, M = 5e2, nburnin = 2e2)
res.step$plot_infection
res.step$plot_removed
res.step$dic_val


### Example 2: continuous exponential function pi(t)
res.exp <- tvt.eSIR (Y, R, begin_str = "01/13/2020", death_in_R = 0.4,
T_fin = 200, exponential = TRUE, dic = FALSE, lambda0 = 0.05,
casename = "Hubei_exp", save_files = FALSE, save_mcmc = FALSE,
M = 5e2, nburnin = 2e2)
res.exp$plot_infection
#res.exp$plot_removed


### Example 3: the basic state-space SIR model, pi(t)=1
```

```
res.nopi <- tvt.eSIR (Y, R, begin_str = "01/13/2020", death_in_R = 0.4,
T_fin = 200, casename = "Hubei_nopi", save_files = FALSE,
M=5e2, nburnin = 2e2)
res.nopi$plot_infection
#res.nopi$plot_removed
```

The other epidemiological model with an added quarantine compartment as an absorbing state was fitted via our R function `qh.eSIR` in the package `eSIR`. The arguments used in `qh.eSIR()` are almost identical to those in `tvt.eSIR()`. Note that if the quarantine rate function is set at constant 0, this model will be reduced to a basic epidemiological SIR model.

```
### Example 4: Dirac delta function of the quarantine process
change_time <- c("01/23/2020", "02/04/2020", "02/08/2020")
phi <- c(0.1, 0.4, 0.4)
res.q <- qh.eSIR (Y, R, begin_str = "01/13/2020",death_in_R = 0.4,
phi0 = phi0, change_time = change_time, casename = "Hubei_q",
save_files = TRUE, save_mcmc = FALSE, M = 5e2, nburnin = 2e2)
res.q$plot_infection
#res.q$plot_removed


### Example 5: basic state-space SIR model
res.noq <- qh.eSIR (Y, R, begin_str = "01/13/2020", death_in_R = 0.4,
T_fin = 200, casename = "Hubei_noq", M = 5e2, nburnin = 2e2)
res.noq$plot_infection
```

In the above R coding scripts, only very short MCMC chains are specified for the consideration of running time. In practice, we set `M=5e5` and `nburnin=2e5` to achieve desirable burn-ins and yield stable MCMC draws.

BIBLIOGRAPHY

BIBLIOGRAPHY

Anatolyev, S. (2005), Gmm, gel, serial correlation, and asymptotic bias, *Econometrica*, *73*(3), 983–1002.

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

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1992), *Statistical models based on counting processes*, Springer Science & Business Media.

Backer, J. A., D. Klinkenberg, and J. Wallinga (2020), Incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from wuhan, china, 20–28 january 2020, *Eurosurveillance*, *25*(5), 2000,062.

Bai, J., B. He, H. Shou, V. Zipunnikov, T. A. Glass, and C. M. Crainiceanu (2014), Normalization and extraction of interpretable metrics from raw accelerometry data, *Biostatistics*, *15*(1), 102–116.

Bai, J., C. Di, L. Xiao, K. R. Evenson, A. Z. LaCroix, C. M. Crainiceanu, and D. M. Buchner (2016), An activity index for raw accelerometry data and its comparison with other activity metrics, *PloS one*, *11*(8), e0160,644.

Bai, J., Y. Sun, J. A. Schrack, C. M. Crainiceanu, and M.-C. Wang (2018), A two-stage model for wearable device data, *Biometrics*, *74*(2), 744–752.

Bashash, M., et al. (2017), Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6–12 years of age in mexico, *Environmental health perspectives*, *125*(9), 097,017.

Bashash, M., et al. (2018), Prenatal fluoride exposure and attention deficit hyperactivity disorder (ADHD) symptoms in children at 6–12 years of age in mexico city, *Environment international*, *121*, 658–666.

Bellinger, D., A. Leviton, C. Waternaux, H. Needleman, and M. Rabinowitz (1987), Longitudinal analyses of prenatal and postnatal lead exposure and early cognitive development, *New England journal of medicine*, *316*(17), 1037–1043.

Bickel, P., F. Götze, and W. van Zwet (1997), Resampling fewer than n observations: Gains, losses, and remedies for losses, *Statistica Sinica*, *7*, 1–31.

Bickel, P. J., and A. Sakov (2008), On the choice of m in the m out of n bootstrap and confidence bounds for extrema, *Statistica Sinica*, *18*(3), 967–985.

Boucher, O., S. W. Jacobson, P. Plusquellec, É. Dewailly, P. Ayotte, N. Forget-Dubois, J. L. Jacobson, and G. Muckle (2012), Prenatal methylmercury, postnatal lead exposure, and evidence of attention deficit/hyperactivity disorder among inuit children in arctic quebec, *Environmental health perspectives*, *120*(10), 1456–1461.

Braun, J. M., R. S. Kahn, T. Froehlich, P. Auinger, and B. P. Lanphear (2006), Exposures to environmental toxicants and attention deficit hyperactivity disorder in us children, *Environmental health perspectives*, *114*(12), 1904–1909.

Breslow, N. E., and X. Lin (1995), Bias correction in generalised linear mixed models with a single component of dispersion, *Biometrika*, *82*(1), 81–91.

Carlin, B. P., N. G. Polson, and D. S. Stoffer (1992), A monte carlo approach to nonnormal and nonlinear state-space modeling, *Journal of the american Statistical association*, *87*(418), 493–500.

Caroni, C., M. Crowder, and A. Kimber (2010), Proportional hazards models with discrete frailty, *Lifetime data analysis*, *16*(3), 374–384.

Catellier, D. J., P. J. Hannan, D. M. Murray, C. L. Addy, T. L. Conway, S. Yang, and J. C. Rice (2005), Imputation of missing data when measuring physical activity by accelerometry, *Medicine and science in sports and exercise*, *37*(11 Suppl), S555.

Chandler, J., K. Brazendale, M. Beets, and B. Mealing (2016), Classification of physical activity intensities using a wrist-worn accelerometer in 8–12-year-old children, *Pediatric obesity*, *11*(2), 120–127.

Chaussé, P. (2010), Computing generalized method of moments and generalized empirical likelihood with R, *Journal of Statistical Software*, *34*(11), 1–35.

Chen, H., et al. (2020), Clinical characteristics and intrauterine vertical transmission potential of covid-19 infection in nine pregnant women: a retrospective review of medical records, *The Lancet*.

Colley, R. C., D. Garriguet, I. Janssen, C. L. Craig, J. Clarke, and M. S. Tremblay (2011), Physical activity of canadian adults: accelerometer results from the 2007 to 2009 canadian health measures survey, *Health reports*, *22*(1), 7.

Conners, C. K., M. Staff, V. Connelly, S. Campbell, M. MacLean, and J. Barnes (2000), Conners' continuous performance test ii (cpt ii v. 5), *Multi-Health Syst Inc*, *29*, 175–96.

Cook, R. J., and J. Lawless (2007), *The statistical analysis of recurrent events*, Springer Science & Business Media.

Cox, D. R. (1972), Regression models and life-tables, *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.

Cox, D. R. (1975), Partial likelihood, *Biometrika*, *62*(2), 269–276.

Cummins, S. K., and L. R. Goldman (1992), Even advantaged children show cognitive deficits from low-level lead toxicity, *Pediatrics*, *90*(6), 995–997.

Delamater, P. L., E. J. Street, T. F. Leslie, Y. T. Yang, and K. H. Jacobsen (2019), Complexity of the basic reproduction number (r0), *Emerging infectious diseases*, *25*(1), 1.

Demidenko, E. (2004), *Mixed Models: Theory and Applications*, Wiley-Interscience.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), Maximum likelihood from incomplete data via the em algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.

Dennis, D. T., K. L. Gage, N. G. Gratz, J. D. Poland, E. Tikhomirov, W. H. Organization, et al. (1999), Plague manual: epidemiology, distribution, surveillance and control, *Tech. rep.*, Geneva: World Health Organization.

Dharmarajan, S. H., D. E. Schaubel, and R. Saran (2018), Evaluating center performance in the competing risks setting: Application to outcomes of wait-listed end-stage renal disease patients, *Biometrics*, *74*(1), 289–299.

Dvorkin, D. (2012), *lcmix: Layered and chained mixture models*, r package version 0.3/r5.

DXY.cn (2020), DX Doctor COVID-19 Global Pandemic Real-time Report.

Eddelbuettel, D., and C. Sanderson (2014), Rcpparmadillo: Accelerating r with high-performance c++ linear algebra, *Computational Statistics and Data Analysis*, *71*, 1054–1063.

Eisenberg, M. C., J. N. Eisenberg, J. P. D'Silva, E. V. Wells, S. Cherng, Y.-H. Kao, and R. Meza (2015), Forecasting and uncertainty in modeling the 2014-2015 ebola epidemic in west africa, *arXiv preprint arXiv:1501.05555*.

Evenson, K. R., et al. (2015), Calibrating physical activity intensity for hip-worn accelerometry in women age 60 to 91 years: The women's health initiative opach calibration study, *Preventive medicine reports*, *2*, 750–756.

Fan, Y., K. Zhao, Z.-L. Shi, and P. Zhou (2019), Bat coronaviruses in china, *Viruses*, *11*(3), 210.

Fitzmaurice, G. M., N. M. Laird, and J. H. Ware (2012), *Applied longitudinal analysis*, vol. 998, John Wiley & Sons.

Fleming, T. R., and D. P. Harrington (2011), *Counting processes and survival analysis*, vol. 169, John Wiley & Sons.

Fraser, C., et al. (2009), Pandemic potential of a strain of influenza a (h1n1): early findings, *science*, *324*(5934), 1557–1561.

Freedson, P., E. Melanson, and J. Sirard (1998), Calibration of the computer science and applications, inc. accelerometer, *Medicine & science in sports & exercise*, *30*(5), 777–781.

Freedson, P., D. Pober, and K. F. Janz (2005), Calibration of accelerometer output for children, *Medicine & Science in Sports & Exercise*, *37*(11), S523–S530.

Gamado, K., G. Streftaris, and S. Zachary (2017), Estimation of under-reporting in epidemics using approximations, *Journal of mathematical biology*, *74*(7), 1683–1707.

Gamado, K. M., G. Streftaris, and S. Zachary (2014), Modelling under-reporting in epidemics, *Journal of mathematical biology*, *69*(3), 737–765.

Gasperoni, F., F. Ieva, A. M. Paganoni, C. H. Jackson, and L. Sharples (2018), Non-parametric frailty cox models for hierarchical time-to-event data, *Biostatistics*.

Goldsmith, J., V. Zipunnikov, and J. Schrack (2015), Generalized multilevel function-on-scalar regression and principal component analysis, *Biometrics*, *71*(2), 344–353.

Goldsmith, J., X. Liu, J. Jacobson, and A. Rundle (2016), New insights into activity patterns in children, found using functional data analyses, *Medicine and science in sports and exercise*, *48*(9), 1723.

Gralinski, L. E., and V. D. Menachery (2020), Return of the coronavirus: 2019-ncov, *Viruses*, *12*(2), 135.

Gray, R. J. (1992), Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis, *Journal of the American Statistical Association*, *87*(420), 942–951.

Guan, W.-j., et al. (2020), Clinical characteristics of 2019 novel coronavirus infection in china, *medRxiv*.

Guo, G., and G. Rodriguez (1992), Estimating a multivariate proportional hazards model for clustered data using the em algorithm, with an application to child survival in guatemala, *Journal of the American Statistical Association*, *87*(420), 969–976.

Hansen, L. P. (1982), Large sample properties of generalized method of moments estimators, *Econometrica: Journal of the Econometric Society*, pp. 1029–1054.

Hansen, L. P., J. Heaton, and A. Yaron (1996), Finite-sample properties of some alternative gmm estimators, *Journal of Business & Economic Statistics*, *14*(3), 262–280.

Harrell, F. E., R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati (1982), Evaluating the yield of medical tests, *Jama*, *247*(18), 2543–2546.

Harrington, D. M., G. J. Welk, and A. E. Donnelly (2011), Validation of met estimates and step measurement using the activpal physical activity logger, *Journal of sports sciences*, *29*(6), 627–633.

Heckman, J., and B. Singer (1984), A method for minimizing the impact of distributional assumptions in econometric models for duration data, *Econometrica: Journal of the Econometric Society*, pp. 271–320.

Heckman, J. J., and B. Singer (1982), Population heterogeneity in demographic models.

Hildebrand, M., V. H. VAN, B. H. Hansen, and U. Ekelund (2014), Age group comparability of raw accelerometer output from wrist-and hip-worn monitors., *Medicine and science in sports and exercise*, *46*(9), 1816–1824.

Holshue, M. L., et al. (2020), First case of 2019 novel coronavirus in the united states, *New England Journal of Medicine*.

Hong, S.-B., et al. (2015), Environmental lead exposure and attention deficit/hyperactivity disorder symptom domains in a community sample of south korean school-age children, *Environmental health perspectives*, *123*(3), 271–276.

Huang, C., et al. (2020), Clinical features of patients infected with 2019 novel coronavirus in wuhan, china, *The Lancet*.

Huang, C.-Y., and M.-C. Wang (2004), Joint modeling and estimation for recurrent event processes and failure time data, *Journal of the American Statistical Association*, *99*(468), 1153–1165.

Huang, C.-Y., and M.-C. Wang (2005), Nonparametric estimation of the bivariate recurrence time distribution, *Biometrics*, *61*(2), 392–402.

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

Huang, Y. (2000), Two-sample multistate accelerated sojourn times model, *Journal of the American Statistical Association*, *95*(450), 619–627.

Huang, Y. (2002), Censored regression with the multistate accelerated sojourn times model, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(1), 17–29.

Hui, D. S., et al. (2020), The continuing 2019-ncov epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in wuhan, china, *International Journal of Infectious Diseases*, *91*, 264–266.

Imai, N., I. Dorigatti, A. Cori, S. Riley, and N. M. Ferguson (2020), Estimating the potential total number of novel coronavirus cases in wuhan city, china.

Jørgensen, B., S. Lundbye-Christensen, P.-K. Song, and L. Sun (1999), A state space model for multivariate longitudinal count data, *Biometrika*, *86*(1), 169–181.

Jøsrgensen, B., and P. X.-K. Song (2007), Stationary state space models for longitudinal data, *Canadian Journal of Statistics*, *35*(4), 461–483.

Jung, S.-m., A. R. Akhmetzhanov, K. Hayashi, N. M. Linton, Y. Yang, B. Yuan, T. Kobayashi, R. Kinoshita, and H. Nishiura (2020), Real-time estimation of the risk of death from novel coronavirus (covid-19) infection: Inference using exported cases, *Journal of Clinical Medicine*, *9*(2), 523.

Kalbfleisch, J. D., and R. L. Prentice (2002), *The Statistical Analysis of Failure Time Data*, Wiley.

Kalbfleisch, J. D., D. E. Schaubel, Y. Ye, and Q. Gong (2013), An estimating function approach to the analysis of recurrent and terminal events, *Biometrics*, *69*(2), 366–374.

Kermack, W. O., and A. G. McKendrick (1927), A contribution to the mathematical theory of epidemics, *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, *115*(772), 700–721.

Klein, J. P. (1992), Semiparametric estimation of random effects using the cox model based on the em algorithm, *Biometrics*, pp. 795–806.

Kleinsasser, M., D. Barker, and L. Wang (2020), Explore analysis and forecast results for china.

Laird, N. (1978), Nonparametric maximum likelihood estimation of a mixing distribution, *Journal of the American Statistical Association*, *73*(364), 805–811.

Lange, K. (2010), *Numerical analysis for statisticians*, Springer Science & Business Media.

Lauer, S. A., K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler (2020), The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application, *Annals of internal medicine*.

Lawless, J. F., and C. Nadeau (1995), Some simple robust methods for the analysis of recurrent events, *Technometrics*, *37*(2), 158–168.

Lee, C. H., X. Luo, C.-Y. Huang, T. E. DeFor, C. G. Brunstein, and D. J. Weisdorf (2016), Nonparametric methods for analyzing recurrent gap time data with application to infections after hematopoietic cell transplant, *Biometrics*, *72*(2), 535–545.

Lee, C. H., C.-Y. Huang, G. Xu, and X. Luo (2018), Semiparametric regression analysis for alternating recurrent event data, *Statistics in Medicine*, *37*(6), 996–1008.

Li, H., E. A. Thompson, and E. M. Wijsman (1998), Semiparametric estimation of major gene effects for age of onset, *Genetic Epidemiology*, *15*(3), 279–298.

Li, H., J. Staudenmayer, and R. J. Carroll (2014), Hierarchical functional data with mixed continuous and binary measurements, *Biometrics*, *70*(4), 802–811.

Li, H., Y. Zhang, R. J. Carroll, S. K. Keadle, J. N. Sampson, and C. E. Matthews (2017), A joint modeling and estimation method for multivariate longitudinal data with mixed types of responses to analyze physical activity data generated by accelerometers, *Statistics in medicine*, *36*(25), 4028–4040.

Li, H., J. Staudenmayer, T. Wang, S. K. Keadle, and R. J. Carroll (2018), Three-part joint modeling methods for complex functional data mixed with zero-and-one–inflated proportions and zero-inflated continuous outcomes with skewness, *Statistics in medicine*, *37*(4), 611–626.

Lin, D., W. Sun, and Z. Ying (1999), Nonparametric estimation of the gap time distribution for serial events with censored data, *Biometrika*, *86*(1), 59–70.

Lin, D., L. Wei, I. Yang, and Z. Ying (2000), Semiparametric regression for the mean and rate functions of recurrent events, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(4), 711–730.

Lin, X., and N. E. Breslow (1996), Bias correction in generalized linear mixed models with multiple components of dispersion, *Journal of the American Statistical Association*, *91*(435), 1007–1016.

Little, R. J. (1995), Modeling the drop-out mechanism in repeated-measures studies, *Journal of the american statistical association*, *90*(431), 1112–1121.

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

Liu, Q., and D. A. Pierce (1993), Heterogeneity in mantel-haenszel-type models, *Biometrika*, *80*(3), 543–556.

Luk, H. K., X. Li, J. Fung, S. K. Lau, and P. C. Woo (2019), Molecular epidemiology, evolution and phylogeny of sars coronavirus, *Infection, Genetics and Evolution.*

Ma, R. (1999), An orthodox blup approach to generalized linear mixed models, Ph.D. thesis, University of British Columbia.

Ma, R., D. Krewski, and R. T. Burnett (2003), Random effects cox models: a poisson modelling approach, *Biometrika*, *90*(1), 157–169.

Massonnet, G., T. Burzykowski, and P. Janssen (2006), Resampling plans for frailty models, *Communications in Statistics-Simulation and Computation*, *35*(2), 497–514.

Mkhatshwa, T., and A. Mummert (2010), Modeling super-spreading events for infectious diseases: case study sars, *arXiv preprint arXiv:1007.0908*.

Moreno, E. S. S. (2008), *Nonparametric frailty models for clustered survival data*, Cornell University.

Needleman, H. L., A. Schell, D. Bellinger, A. Leviton, and E. N. Allred (1990), The long-term effects of exposure to low doses of lead in childhood: an 11-year follow-up report, *New England journal of medicine*, *322*(2), 83–88.

Newey, W. K., and R. J. Smith (2004), Higher order properties of gmm and generalized empirical likelihood estimators, *Econometrica*, *72*(1), 219–255.

Nielsen, G. G., R. D. Gill, P. K. Andersen, and T. I. Sørensen (1992), A counting process approach to maximum likelihood estimation in frailty models, *Scandinavian journal of Statistics*, pp. 25–43.

Osthus, D., K. S. Hickmann, P. C. Caragea, D. Higdon, and S. Y. Del Valle (2017), Forecasting seasonal influenza with a state-space sir model, *The annals of applied statistics*, *11*(1), 202.

Pepe, M. S., and J. Cai (1993), Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates, *Journal of the American Statistical Association*, *88*(423), 811–820.

Perng, W., et al. (2019), Early life exposure in mexico to environmental toxicants (element) project, *BMJ open*, *9*(8), e030,427.

Pisoni, R. L., B. W. Gillespie, D. M. Dickinson, K. Chen, M. H. Kutner, and R. A. Wolfe (2004), The dialysis outcomes and practice patterns study: design, data elements, and methodology, *American Journal of Kidney Diseases*, *44*, 7–15.

Plummer, M. (2019), *rjags: Bayesian Graphical Models using MCMC*, r package version 4-10.

Polanczyk, G., M. S. De Lima, B. L. Horta, J. Biederman, and L. A. Rohde (2007), The worldwide prevalence of ADHD: a systematic review and metaregression analysis, *American journal of psychiatry*, *164*(6), 942–948.

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

Puyau, M. R., A. L. Adolph, F. A. Vohra, and N. F. Butte (2002), Validation and calibration of physical activity monitors in children, *Obesity research*, *10*(3), 150–157.

R Core Team (2018a), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

R Core Team (2018b), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

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

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

Robinson, B. M., B. Bieber, R. L. Pisoni, and F. K. Port (2012), Dialysis outcomes and practice patterns study: its strengths, limitations, and role in informing practices and policies, *Clinical Journal of the American Society of Nephrology*, *7*, 1897–1905.

Rothe, C., et al. (2020), Transmission of 2019-ncov infection from an asymptomatic contact in germany, *New England Journal of Medicine*.

Ruli, E., N. Sartori, L. Ventura, et al. (2016), Improved laplace approximation for marginal likelihoods, *Electronic Journal of Statistics*, *10*(2), 3986–4009.

Sasaki, J. E., D. John, and P. S. Freedson (2011), Validation and comparison of actigraph activity monitors, *Journal of science and medicine in sport*, *14*(5), 411–416.

Schaubel, D. E., and J. Cai (2004a), Non-parametric estimation of gap time survival functions for ordered multivariate failure time data, *Statistics in Medicine*, *23*(12), 1885–1900.

Schaubel, D. E., and J. Cai (2004b), Regression methods for gap time hazard functions of sequentially ordered multivariate failure time data, *Biometrika*, *91*(2), 291–303.

Shun, Z., and P. McCullagh (1995), Laplace approximation of high dimensional integrals, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 749–760.

Smith, R. D. (2006), Responding to global infectious disease outbreaks: lessons from sars on the role of risk perception, communication and management, *Social science & medicine*, *63*(12), 3113–3123.

Smith, R. J. (1997), Alternative semi-parametric likelihood approaches to generalised method of moments estimation, *The Economic Journal*, *107*(441), 503–519.

Solomon, P., and D. R. Cox (1992), Nonlinear component of variance models, *Biometrika*, *79*(1), 1–11.

Song, P. X.-K. (2000), Monte carlo kalman filter and smoothing for multivariate discrete state space models, *Canadian Journal of Statistics*, *28*(3), 641–652.

Song, P. X.-K. (2007), *Correlated Data Analysis: Modeling, Analytics, and Applications*, Springer.

Staudenmayer, J., W. Zhu, and D. J. Catellier (2012), Statistical considerations in the analysis of accelerometry-based activity monitor data., *Medicine and science in sports and exercise*, *44*(1 Suppl 1), S61–7.

Subissi, L., C. C. Posthuma, A. Collet, J. C. Zevenhoven-Dobbe, A. E. Gorbalenya, E. Decroly, E. J. Snijder, B. Canard, and I. Imbert (2014), One severe acute respiratory syndrome coronavirus protein complex integrates processive rna polymerase and exonuclease activities, *Proceedings of the National Academy of Sciences*, *111*(37), E3900–E3909.

Sun, H., Y. Qiu, H. Yan, Y. Huang, Y. Zhu, J. Gu, and S. X. Chen (2020), Tracking reproductivity of covid-19 epidemic in china with varying coefficient sir model, *Journal of Data Science*, *to appear*, a earlier version accessible at https://doi.org/10.1101/2020.02.17.20024257.

Sun, J., L. Sun, and D. Liu (2007), Regression analysis of longitudinal data in the presence of informative observation and censoring times, *Journal of the American Statistical Association*, *102*(480), 1397–1406.

Thapar, A., and M. Cooper (2016), Attention deficit hyperactivity disorder, *The Lancet*, *387*(10024), 1240–1250.

Thapar, A., M. Cooper, R. Jefferies, and E. Stergiakouli (2012), What causes attention deficit hyperactivity disorder?, *Archives of disease in childhood*, *97*(3), 260–265.

Thapar, A., M. Cooper, O. Eyre, and K. Langley (2013), Practitioner review: what have we learnt about the causes of ADHD?, *Journal of Child Psychology and Psychiatry*, *54*(1), 3–16.

Therneau, T. M. (2015), *A Package for Survival Analysis in S*, version 2.38.

Therneau, T. M. (2018), *coxme: Mixed Effects Cox Models*, r package version 2.2-10.

Therneau, T. M., and P. M. Grambsch (2000), *Modeling survival data: extending the Cox model*, Springer.

Therneau, T. M., P. M. Grambsch, and V. S. Pankratz (2003), Penalized survival models and frailty, *Journal of Computational and Graphical Statistics*, *12*(1), 156–175.

Troiano, R. P., D. Berrigan, K. W. Dodd, L. C. Masse, T. Tilert, M. McDowell, et al. (2008), Physical activity in the united states measured by accelerometer, *Medicine and science in sports and exercise*, *40*(1), 181.

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

van den Boogaart, K. G., R. Tolosana-Delgado, and M. Bren (2018), *compositions: Compositional Data Analysis*, r package version 1.40-2.

Van Hees, V. T., et al. (2013), Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity, *PloS one*, *8*(4), e61,691.

Verweij, P. J., and H. C. Van Houwelingen (1994), Penalized likelihood in cox regression, *Statistics in Medicine*, *13*(23-24), 2427–2436.

Visser, M. (1996), Nonparametric estimation of the bivariate survival function with an application to vertically transmitted aids, *Biometrika*, *83*(3), 507–518.

Walker, S. G., and B. K. Mallick (1997), Hierarchical generalized linear models and frailty models with bayesian nonparametric mixing, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *59*(4), 845–860.

Wang, C., P. W. Horby, F. G. Hayden, and G. F. Gao (2020a), A novel coronavirus outbreak of global health concern, *The Lancet*.

Wang, H.-L., X.-T. Chen, B. Yang, F.-L. Ma, S. Wang, M.-L. Tang, M.-G. Hao, and D.-Y. Ruan (2008), Case–control study of blood lead levels and attention deficit hyperactivity disorder in chinese children, *Environmental health perspectives*, *116*(10), 1401–1406.

Wang, L., F. Wang, L. Tang, P. Egeler, B. Zhu, Y. Zhou, J. He, and P. X.-K. Song (2020b), *eSIR: Extended state-space SIR models*, R package version 0.2.5.

Wang, L., et al. (2020c), An epidemiological forecast model and software assessing interventions on covid-19 epidemic in china (with discussion), *Journal of Data Science*, *to appear*, an earlier version accessible at https://doi.org/10.1101/2020.02.29.20029421.

Wang, M.-C. (1999), Gap time bias in incident and prevalent cohorts, *Statistica Sinica*, pp. 999–1010.

Wang, M.-C., and S.-H. Chang (1999), Nonparametric estimation of a recurrent survival function, *Journal of the American Statistical Association*, *94*(445), 146–153.

Wang, M.-C., and C.-Y. Huang (2014), Statistical inference methods for recurrent event processes with shape and size parameters, *Biometrika*, *101*(3), 553–566.

Wang, M.-C., N. P. Jewell, and W.-Y. Tsai (1986), Asymptotic properties of the product limit estimate under random truncation, *the Annals of Statistics*, pp. 1597–1605.

Wang, M.-C., J. Qin, and C.-T. Chiang (2001), Analyzing recurrent event data with informative censoring, *Journal of the American Statistical Association*, *96*(455), 1057–1065.

Wang, W., and M. T. Wells (1998), Nonparametric estimation of successive duration times under dependent censoring, *Biometrika*, *85*(3), 561–572.

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

Welk, G. (2005), Principles of design and analyses for the calibration of accelerometry-based activity monitors, *Medicine & Science in Sports & Exercise*, *37*(11).

Wickham, H. (2016), *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.

Wienke, A. (2010), *Frailty models in survival analysis*, CRC Press.

World Health Organization (2003), Summary of probable sars cases with onset of illness from 1 november 2002 to 31 july 2003, *http://www. who. int/csr/sars/country/table2004_04_21/en/index. html.*

World Health Organization (2020a), Naming the coronavirus disease (covid-19) and the virus that causes it, *https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it.*

World Health Organization (2020b), Emergencies preparedness, response. pneumonia of unknown origin – china., *https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/.*

Xiang, Y.-T., et al. (2020), Timely research papers about covid-19 in china, *The Lancet.*

Xiao, Y., and M. Abrahamowicz (2010), Bootstrap-based methods for estimating standard errors in cox's regression analyses of clustered event times, *Statistics in medicine*, *29*(7-8), 915–923.

Xu, X.-W., et al. (2020), Clinical findings in a group of patients infected with the 2019 novel coronavirus (sars-cov-2) outside of wuhan, china: retrospective case series, *BMJ*, *368.*

Xu, Z. (2011), Statistical design and survival analysis in cluster randomized trials.

Xue, X. (1998), Multivariate survival data under bivariate frailty: an estimating equation approach, *Biometrics*, pp. 1631–1637.

Xue, X., and R. Brookmeyer (1996), Bivariate frailty model for the analysis of multivariate survival time, *Lifetime Data Analysis*, *2*(3), 277–289.

Yan, J., and J. P. Fine (2008), Analysis of episodic data with application to recurrent pulmonary exacerbations in cystic fibrosis patients, *Journal of the American Statistical Association*, *103*(482), 498–510.

Yashin, A. I., J. W. Vaupel, and I. A. Iachine (1995), Correlated individual frailty: an advantageous approach to survival analysis of bivariate data, *Mathematical Population Studies*, *5*(2), 145–159.

Ye, Y., J. D. Kalbfleisch, and D. E. Schaubel (2007), Semiparametric analysis of correlated recurrent and terminal events, *Biometrics*, *63*(1), 78–87.

Young, E. W., D. A. Goodkin, D. L. Mapes, F. K. Port, M. L. Keen, K. Chen, et al. (2000), The dialysis outcomes and practice patterns study: an international hemodialysis study, *Kidney International*, *57*, S74–S81.

Yu, G. (2020), *nCov2019: Stats of the '2019-nCov' Cases*, r package version 0.0.8.

Zamfirescu, T. (1972), Fix point theorems in metric spaces, *Archiv der Mathematik*, *23*(1), 292–298.

Zhang, Y., H. Li, S. K. Keadle, C. E. Matthews, and R. J. Carroll (2019), A review of statistical analyses on physical activity data collected from accelerometers, *Statistics in Biosciences*, *11*(2), 465–476.

Zhu, B., J. M. Taylor, and P. X.-K. Song (2012), Signal extraction and breakpoint identification for array cgh data using robust state space model, *arXiv preprint arXiv:1201.5169*.

Zhu, N., et al. (2020), A novel coronavirus from patients with pneumonia in china, 2019, *New England Journal of Medicine*.