2016

Semiparametric Regression Analysis of Panel Count Data and Interval-Censored Failure Time Data

Bin Yao
University of South Carolina

Semiparametric regression analysis of panel count data and interval-censored failure time data

by

Bin Yao

Bachelor of Science
Beijing Normal University 2011

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Statistics

College of Arts and Sciences

University of South Carolina

2016

Accepted by:

Lianming Wang, Major Professor

Edsel Peña, Committee Member

Timothy Hanson, Committee Member

Alexander C. McLain, Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

Acknowledgments

Over the past four years I have received enormous support and encouragement from a great number of people. I would like to express my deepest appreciation to my advisor, Dr. Lianming Wang, for his constant guidance and enthusiastic encouragement during my Ph.D. study, without which this dissertation work would not have been possible. I would also like to express my sincere thankfulness to my committee, Dr. Timothy Hanson, Dr. Edsel Peña, and Dr. Alexander McLain, for their insightful comments on my dissertation and encouragement. My grateful thanks are also extended to all professors in the Department of Statistics. I learned a lot from their classes. I would also like to thank my friends in USC who make my graduate study life joyful. Finally, I would also like to thank my family: my parents, my elder sister and little brother for always supporting me and encouraging me with their best wishes.

Abstract

This dissertation discusses three important research topics on semiparametric regression analysis of panel count data and interval-censored data. Both types of data arise commonly in real-life studies in many fields such as epidemiology, social science, and medical research. In these studies, subjects are usually examined multiple times at periodical or irregular follow-up examinations. For panel count data, the response variable is the counts of some recurrent events, whose exact occurrence times are usually unknown. For interval-censored data, the response variable is the time to some events of interest, often called survival time or failure time, and the exact response time is never observed but is known to fall within some interval formed by two examination times. The primary goal for both types of data is to study effects of covariates on the response variable and can be completed by regression analysis.

Chapter 1 of this dissertation provides some detailed descriptions about panel count data and interval-censored data with several real-life examples. A literature review is conducted on existing approaches and commonly used semiparametric regression models for analyzing the two types of data. Some preliminary knowledge used in our approaches such as monotone splines and EM algorithm is also presented in this chapter.

In Chapter 2, we propose a gamma frailty non-homogeneous Poisson process model for the regression analysis of panel count data to account for the within-subject correlation. This topic is important because ignoring such within-subject correlation results in biased estimation and may lead to misleading conclusions, and literature is limited on this topic. We propose an efficient estimation approach based on an

EM algorithm. Our approach is robust to initial values, converges fast, and provides variance estimate in closed form. Our approach has shown an excellent performance in estimating both regression parameters and the baseline mean function when there is indeed within-subject correlation and can also be used when such correlation does not exist. An R package `PCDSpline` has been developed and available on CRAN to disseminate our approach.

In Chapter 3, we study regression analysis of case 1 interval-censored data, also referred to as current status data, using the generalized odds-rate hazards (GORH) models. The GORH models are a general class of semiparametric models and have been widely used for analyzing right-censored data. However, their use for current status data is not found in the literature. We propose an efficient estimation approach with fixed ρ in the GORH models based on a novel EM algorithm. The proposed method is robust to initial values, fast to converge and provides variance estimates in closed form. A working model approach is proposed when true value of ρ is known but does not require to fit the GORH models with different ρ values. The proposed approach and working model strategy are evaluated and show good performance in an extensive simulation study. They are illustrated by a large real-life data set.

In Chapter 4, we study the joint modeling of panel count data and interval-censored failure time data motivated by a real-life data set about sexually transmitted infections (STI). The failure time of interest is the time to get a new STI since the enrollment, which has an interval-censored data structure. The other response variable is the number of unprotected sex over time, which has a panel count data structure. The proposed joint analysis based on an EM algorithm is more efficient than the univariate analysis of panel count data and interval-censored data separately. The proposed joint model and approach are applied to the STI data.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Panel count data

The study that deals with repeated occurrences of the same event of interest on study subjects are usually referred to as event history study. The resulting data are often called event history data. In general an event history study can be classified into two types. The first type of study monitors study subjects continuously, and consequently all the occurrence times are observed producing recurrent event data. The second type of study produces panel count data when subjects are monitored or examined only at periodic observation times, and only the number of events that occur between consecutive observation times is known. Panel count data arise naturally in many fields, such as in demographical studies, epidemiological studies, medical follow-up studies, oncology clinical trials, and reliability studies.

One famous panel count data example in the literature comes from the National Cooperative Gallstone Study (NCGS). The NCGS was a 10-year, double-blinded, placebo-controlled clinical trial of the use of the natural bile acid chenodeoxycholic acid for the dissolution of cholesterol gallstones (Thall and Lachin, 1988; Sun and Zhao, 2013). A total of 916 patients were randomized into three different groups, placebo, low dose, and high dose and were treated up to two years. Among the various types of symptoms occurred to patients in the study, nausea was commonly associated with gallstone disease. Thall and Lachin (1988) focused on the incidence of nausea during the first year of follow-up in the NCGS. The patients were scheduled

to return for clinic observations at 1, 2, 3, 6, 9, and 12 months during the first year follow-up. However, the actual visit times differ from patient to patient due to early dropout or withdrawal etc. At each clinic visit, the patients were asked to report the total number of each type of symptom that had occurred between consecutive visits such as the number of the incidences of nausea. The observed data include actual visit times and the number of occurrences of nausea between the consecutive visits, therefore we have panel count data on the occurrences of nausea. One research interest was to conduct regression analysis of these panel count data for treatment comparison and estimation of some covariate effects.

When there are no covariates, many parametric approaches have been proposed. Kalbfleisch and Lawless (1985) discussed the fitting of a finite state Markov model to panel count data. Liang and Zeger (1986) and Thall and Vail (1990) presented quasi-likelihood regression models with a generalized estimating equation (GEE) approach by treating panel count data as longitudinal count data. Many approaches have been proposed to analyze panel count data based on the counting process techniques. When no covariates are considered, inferences are focused on estimating the mean function of the counting process and comparison of cumulative mean functions. For the purpose of mean function estimation, Sun and Kalbfleisch (1995) constructed an isotonic regression estimator. Wellner and Zhang (2000) proposed two estimators by maximizing the pseudo-likelihood and likelihood functions under the nonhomogeneous Poisson process. Lu et al. (2007) proposed likelihood-based estimators with the mean function being approximated by the monotone splines of Ramsay (1988) and showed that their spline-based estimators are more efficient than those in Wellner and Zhang (2000). Regarding the comparison of the mean functions for different populations, Sun and Fang (2003), Zhang (2006),and Balakrishnan and Zhao (2009) proposed different nonparametric tests for univariate panel count data. Li et al.

(2014) developed nonparametric tests for multivariate panel count data.

When covariates are available, semiparametric regression analysis is widely used to examine the covariate effects on the response as well as the estimation of the mean function. Among others, Sun and Wei (2000) developed estimation procedures with time-dependent covariates on both observational and censoring processes. Zhang (2002) proposed a pseudo-likelihood approach, and Wellner and Zhang (2007) studied both pseudolikelihood and likelihood methods under the non-homogeneous Poisson process model. Hu et al. (2003) proposed two estimation approaches with different assumptions on the observational process. Lu et al. (2009) modeled the baseline mean function with monotone B-splines and established the asymptotic properties of their spline-based estimators. He et al. (2008) considered the regression analysis of multivariate panel count data. There are also many approaches developed for the cases that the recurrent event process and the observational process are dependent. Huang et al. (2006), Sun et al. (2007) and Zhao and Tong (2011) employed joint modeling procedure by using some shared frailty models. Zhao et al. (2013) proposed a general and robust estimation approach by relaxing the Poisson assumption on the observation process. In addition, semiparametric transformation models and dependent terminal events were also considered for analyzing panel count data. For such work, we refer to Sun and Zhao (2013) for a comprehensive review.

1.2 Interval-censored data

Time-to-event data is also referred to as failure time data, where the event of interest can be a failure or a survival event, such as death and the occurrence of some diseases. Time to event data often arise in medical studies, epidemiological studies, sociological studies and reliability experiments. In such studies the failure time of an event is a random variable of interest.

Interval-censored data are a special case of time-to-event data, in which the failure time of interest is not exactly observed but is known to fall within some interval. Interval-censored data arise naturally in epidemiological, financial, medical, and sociological studies. Let T be a nonnegative random variable representing the failure time to an event of a study subject. The random failure time T is interval-censored if T falls into an interval $(L, R]$ where $0 \leq L \leq R$. Interval censoring contains left censoring and right censoring as special cases, $L = 0$ denotes left censoring, $R = \infty$ represents the right censoring. Especially $L = R$ suggests that the exact observation is obtained. A well known example of interval-censored data is breast cancer data came from a retrospective study on early breast cancer patients who were treated at the Joint Center for Radiation Therapy in Boston between 1976 and 1980. One objective of this study was to detect whether chemotherapy changes the rate of deterioration of the cosmetic state. This data set consists of 94 patients who were assigned randomly into two treatments. 46 patients received radiation therapy alone (RT) and 48 patients received radiation therapy plus adjuvant chemotherapy (RCT). In this study, patients were scheduled to visit the clinic every 4-6 months. However, actual visit times differed from patient to patient because of missing visits. At each visit, physicians evaluated the cosmetic appearance of the patients such as breast retraction, a response that has a negative impact on overall cosmetic appearance. The failure time was defined as the time to breast retraction, which was not observed exactly. We have interval-censored data available on time to breast retraction.

For the analysis of interval-censored data, people are often interested in estimating a survival function, comparing survival functions or treatments and assessing covariate effects on failure time. When no covariates are present, estimation of survival functions is the main target in the analysis of failure time data. One possible reason is that one may need to estimate survival functions to estimate some certain

survival probabilities to compare different treatments graphically, or to predict survival probabilities for future patients. Turnbull (1976) proposed the self-consistent estimator based on an EM algorithm which has been widely used as the nonparametric maximum likelihood estimator (NPMLE) of the survival function. Several other algorithms for deriving the NPMLE of a survival function have been discussed by Groeneboom and Wellner (1992), Gentleman and Geyer (1994), Wellner and Zhan (1997), Li et al. (1997), Huang and Wellner (1997). When multiple treatment groups are available the primary objective of analyzing interval-censored data is to compare survival functions, and most nonparametric test procedures can be classified into two categories: rank-based ones and survival-based ones (Huang and Wellner, 1997; Fang et al., 2002; Zhao and Sun, 2004; Sun et al., 2005; Sun, 2006).

When covariates are present, the primary interest for analyzing interval-censored data is to investigate the influence of covariates on the failure time T . The proportional hazards (PH) model is the most popular survival model in analyzing censored data. Finkelstein (1986) was the first to apply the PH model for interval-censored data and developed a Newton-Raphson algorithm. Alternative approaches under the PH model include Satten (1996), Goggins et al. (1998), Cai and Betensky (2003) and Wang et al. (2015). Several other semiparametric models have also been considered for regression analysis of interval-censored data, including the proportional odds (PO) model, the accelerated failure time (AFT) model. Huang and Rossini (1997) and Shen (1998) studied interval-censored data under the PO model by developing sieve estimation procedures. The AFT model has been studied in Rabinowitz et al. (2000) and Betensky et al. (2001) which developed proposing estimating equation approaches for estimating the regression parameters. Sun (2006) and Zhang and Sun (2010) provided comprehensive reviews of analysis of interval-censored data.

Current status data

Current status data, also referred to as case 1 interval-censored data, is one special case of interval-censored data. Current status data occur when each subject is observed only once and the known information is whether the event of interest has occurred no later than the examination time or not. The observed current status data for one subject can be represented by $\{C, \delta = I(T \leq C)\}$ where C denotes the observation time and $I(\cdot)$ is the indicator function. In consequence, the failure time is either left- or right- censored. Current status data arise commonly in many epidemiological, medical or tumorigenicity studies. For example, in tumorigenicity experiments male rats are exposed to two different environments, conventional environment and germ-free environment, to determine whether the environment accelerates the time until tumor onset. Tumor onset time is not measured directly, mainly because the lung tumor is non-lethal and tumor status can only be determined at the death time of the animal. The time to tumor onset is only known to be less than or greater than the observed time of death, therefore we have current status data on the time to tumor occurrence.

Similar to the objectives of analyzing general interval-censored data, many estimation and inference methods proposed for interval-censored data can be directly applied to analyze current status data. When no covariates are present, a number of authors, including Peto (1973), Turnbull (1976) and Groeneboom et al. (2010) have introduced algorithms for the NPMLE of the cumulative distribution function of T with current status data. As for the nonparametric comparison of survival functions for current status data, several procedures have been proposed, such as Sun and Kalbfleisch (1993) and Andersen and Ronn (1995). When covariates are present, the PH model and the PO model have also been studied to fit current status data. Under the PH model, among others, Huang (1996) and Huang and Wellner (1997) explored

efficiency issues and established asymptotic results for the maximum likelihood estimator of both the regression parameters and the baseline cumulative hazard function. Huang (1995) and Rossini and Tsiatis (1996) investigated the current status data with the PO model. McMahan et al. (2013) analyzed current status data under both the PH and PO models using EM algorithm. In addition, Lin et al. (1998) and Shiboski (1998) analyzed current status data with additive hazards model and generalized additive models, respectively. In Chapter 2, we study current status data under a more flexible semiparameteric model called generalized odds-rate hazards models.

1.3 Preliminaries

Several popular survival models

The Cox PH model is the most widely used model in the last four decades in the survival analysis. The PH model relates hazard function $\lambda(t|\boldsymbol{x})$ to a $p \times 1$ covariate vector $\boldsymbol{x}$,

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t) \exp(\boldsymbol{x}'\boldsymbol{\beta}),$$

where $\lambda_0(t)$ is the unspecified nondecreasing baseline hazard function and $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression parameters. The PH model implies that the hazard ratio for two covariate sets remains constant as time changes. However, in reality, the PH assumption may be violated. For remedy, one can stratify the data into some subgroups and apply the PH model for each stratum. Another feasible solution is to include the time-varying covariates in the model. Several other models can be considered when the PH assumption does not hold, including the PO model, the GORH models and the linear transformation model. Cox (1972) proposed the partial likelihood approach for making inference about the regression parameters. The partial likelihood is simple and efficient because the partial likelihood function is only a

function of regression parameters $\boldsymbol{\beta}$ so that people do not have to deal with the nonparametric part which is the baselihe hazard function $\lambda_0(t)$.

The PO model (Bennett, 1983) is another popular survival model. The PO model specifies

$$\frac{F(t|\boldsymbol{x})}{1-F(t|\boldsymbol{x})} = \left\{\frac{F_0(t)}{1-F_0(t)}\right\}\exp(\boldsymbol{x}'\boldsymbol{\beta}),$$

where $F(t|\boldsymbol{x})$ and $F_0(t)$ are the cumulative distribution functions of the failure time T for the treatment group with covariates $\boldsymbol{x}$ and the baseline group with $\boldsymbol{x} = 0$, respectively. The PO model assumes that the effect of covariates is multiplicative on the odds of the survival functions. The PO model is an alternative model to capture the non-proportionality. Unlike the PH assumption, the proportional odds assumption is the hazard ratio between two sets of covariate values converges to unity rather than staying constant as time increases.

Frailty models are commonly used to analyze clustered or multivariate survival data. Clayton (1978) first proposed to use gamma frailty in modeling correlated failure times. A frailty model is a random effect for time-to-event data, where the frailty has multiplicative effect on the baseline hazard function. The univariate gamma frailty model extends the PH model such that

$$\lambda(t|\boldsymbol{x}) = \phi\lambda_0(t)\exp(\boldsymbol{x}'\boldsymbol{\beta}),$$

where $\phi \sim \mathcal{G}(\nu,\nu)$ with mean 1 and variance ν^{-1}. The purpose of using common shape and rate parameters in the gamma distribution is to avoid the non-identifiability issue between ϕ and λ_0. This model will be used in Chapter 4 of this dissertation.

Expectation-Maximization algorithm

The Expectation-Maximization (EM) algorithm was originally proposed by Dempster et al. (1977) to overcome the difficulties in obtaining the maximum likelihood

estimates (MLE). For instance, when the observed likelihood is an integral without having an explicit form, taking derivative of the observed likelihood does not provide us closed-form solution. The EM algorithm can usually simplify the maximization problem by augmenting the missing data into complete data with the goal that the log-likelihood of the complete data is relatively easy to compute. EM algorithm is an iterative algorithm that contains both an expectation step (E-step) and a maximization step (M-step) in each iteration. Let $Y_{com} = (Y_{obs}, Y_{mis})$ be the complete data where Y_{obs} represents the observed data and Y_{mis} stands for the missing data. Let $L^c(\boldsymbol{\theta}|Y_{com})$ denote the complete likelihood function. Denote the expected complete log-likelihood by $Q(\boldsymbol{\theta}|Y_{obs}, \boldsymbol{\theta}_t) = \mathrm{E}\{\log L^c(\boldsymbol{\theta}|Y_{com}, \boldsymbol{\theta}_t)\}$ given the current estimate of $\boldsymbol{\theta}_t$ of $\boldsymbol{\theta}$. The E-step computes $Q(\boldsymbol{\theta}|Y_{obs}, \boldsymbol{\theta}_t)$ at the current step estimate, and the M-step finds $\boldsymbol{\theta} = \boldsymbol{\theta}_{t+1}$ to maximize $Q(\boldsymbol{\theta}|Y_{obs}, \boldsymbol{\theta}_t)$:

E-step: Compute

$$Q(\boldsymbol{\theta}|Y_{obs}, \boldsymbol{\theta}_t) = \int L^c(\boldsymbol{\theta}|Y_{com}) f(Y_{mis}|Y_{obs}, \boldsymbol{\theta}_t) dY_{mis},$$

where $f(Y_{mis}|Y_{obs}, \boldsymbol{\theta}_t)$ is the probability density function of the missing data.

M-step: Obtain $\boldsymbol{\theta}_{t+1} = \mathrm{argmax}_{\boldsymbol{\theta}}\, Q(\boldsymbol{\theta}|Y_{obs}, \boldsymbol{\theta}_t)$.

The algorithm is iterated until $\|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t\|$ is sufficiently small. Many algorithms and formulas have been proposed for obtaining standard errors of parameter estimates from the EM algorithm. Tanner (1996) introduced several variance estimation methods, such as obtaining Hessian matrix numerically and Louis's method (Louis, 1982). Due to its easy implementation and stability the EM algorithm has attracted researchers' attention across different disciplines. Firstly, it can be easily implemented because it relies on complete-data computations: the E-step of each iteration only involves taking expectations over complete-data conditional distributions and the M-step of each iteration only requires complete-data maximum likelihood estimation, for which simple closed form expressions are available. Secondly, it is numerically sta-

ble, because each iteration is required to increase the log-likelihood in each iteration, and if the log-likelihood is bounded, the sequence of the log-likelihood converges to a stationary value.

Monotone splines

A spline function is a piecewise polynomial function where the individual polynomials have the same degree and connect smoothly at join points. The abscissas of these joint points are called knots which partition an interval into a number of subintervals. The most common continuity characteristics imposed on the spline request that for adjacent polynomials the derivatives up to order $k-2$ match. A monotone spline function (Ramsay, 1988), also called integrated splines or I-splines because they can be constructed from I-splines by taking a nonnegative linear combination. I-splines are the integrated functions of M-splines. We can write the M-splines in the form of $f = \sum_{i=1}^{n} c_i M_i$. To construct the M-splines one needs to choose the degree d and set up m' interior knots $\xi_1 < \cdots < \xi_{m'}$ within $[a, b]$. Then one use the following recursive formulas for construction: let $s_1 = a, s_2 = \xi_1, \cdots, s_{m'+1} = \xi_{m'}$, and $s_{m'+2} = b$, then for $i = 1, 2, \ldots, m' + 1$

$$M_i(t|1) = \begin{cases} \frac{1}{s_{i+1}-s_i}, & s_i \le t < s_{i+1}, \\ 0, & \text{otherwise;} \end{cases}$$

for $d \ge 2$, let $s_1 = \cdots = s_d = a, s_{d+1} = \xi_1, \cdots, s_{d+m'} = \xi_{m'}$, and $s_{m'+d+1} = \cdots = s_{m'+2d} = b$, then for $d > 1$ and $i = 1, \ldots, m' + d$,

$$M_i(t|d) = \begin{cases} \frac{k[(t-s_i)M_i(t|d-1)+(s_{i+d}-t)M_{l+1}(t|d-1)]}{(k-1)(s_{i+d}-s_i)}, & s_l \le t < s_{l+d}, \\ 0, & \text{otherwise.} \end{cases}$$

Each $M_l(\cdot|d)$ is a piecewise polynomial with nonzero only within $[s_i, s_{i+d})$ for $i = 1, \ldots, m' + d$. Also we have $\int M_i(x)dx = 1$. Because of this localization a change in coefficient c_i will only effect the spline within this interval. Based on M-spline one can construct two types of splines, B-splines and I-splines. The B-splines can be constructed in this way, $B_i = (t_{i+k} - t_i)M_i/d$. One technique for defining monotone splines is to employ a basis consisting of monotone splines. Because M-splines are nonnegative, one obvious approach is to define the I-splines for the sake of brevity,

$$I_i(t|d) = \int_L^x M_i(u|d)du,$$

and this provides a set of splines which, when combined with nonnegative values of the coefficients c_i, yields monotone splines. The integrated basis function can be constructed in the form of

$$I_i(t|d) = \begin{cases} 0, & i > j, \\ \sum_{h=i}^{j} \frac{(s_{h+d+1} - s_h)M_h(t|d+1)}{d+1}, & j - d + 1 \leq i \leq j, \\ 1, & i < j - d + 1, \end{cases}$$

for each $i = 1, \ldots, m' + d$. In general the shape of a spline function is not very sensitive to knot placement.

Non-homogeneous Poisson process

A counting process $\{N(t) : t \geq 0\}$ is a stochastic process with $N(0) = 0$ and $N(t) < \infty$ almost surely such that the path is right continuous with probability one, piecewise constant, and has only jump discontinuities with jumps of size +1. A Poisson process is a simple and widely used counting process for modeling the times at which arrivals enter a system. Non-homogeneous Poisson process (NHPP) models are commonly used to model recurrent events. The counting process $\{N(t) : t \geq 0\}$ is a non-

homogeneous Poisson process with intensity function $\lambda(t), t \geq 0$ if 1) $N(0) = 0$; 2) For each $t > 0$, $N(t)$ has a Poisson distribution with mean $\mathrm{E}\{N(t)\} = \mu(t) = \int_0^t \lambda(s)ds$. 3) For each $0 \leq t_1 < t_2 < \ldots < t_m$, $N(t_1), N(t_2) - N(t_1), \ldots, N(t_m) - N(t_{m-1})$ are independent Poisson random variables. When $\lambda(t)$ is a constant, the non-homogeneous Poisson process reduces to homogeneous Poisson process with intensity λ. In chapter 2 and chapter 4, we adopt a conditional non-homogeneous Poisson process given frailty for modeling the panel count response.

1.4 Outline

The rest of this dissertation contains three parts about *Semiparametric Regression Analysis of Panel Count Data and Interval-Censored Failure Time Data* from Chapter 2 to Chapter 4.

In Chapter 2, we propose an maximum likelihood approach for panel count data under gamma frailty non-homogeneous Poisson process model to allow for within-subject correlation. An EM algorithm is proposed to estimate the baseline mean function and the regression parameters jointly. The approach can be used to deal with no-within subject correlation data as well. To make researchers apply our method conveniently, we develop a companion R package `PCDSpline` which is available on CRAN for public use.

Chapter 3 considers the regression analysis of current status data with the GORH models. We investigate the non-identifiability issues associated with the GORH models and propose a computationally efficient estimation approach based on an EM algorithm when ρ is known. When ρ is unknown, a working model strategy with $\rho = 1$ is proposed based on our approach that provides valid inferences for testing the significance of covariate effects and estimating survival functions of the true GORH model.

Chapter 4 considers joint modeling of panel count data and interval-censored data. A frailty model is proposed and an EM algorithm is derived for parameter estimation. The proposed joint model includes the gamma frailty non-homogeneous Poisson process model for panel count data and the gamma frailty proportional hazards model for interval-censored data. A computationally efficient estimation approach based on EM algorithm is proposed. Simulation studies show that the proposed approach performs well. The proposed model and approach are applied to a sexually transmitted infections data set.

Chapter 2

Semiparametric regression analysis of panel count data allowing for within-subject correlation

2.1 Introduction

Panel count data often arise in longitudinal prospective studies involving recurrent events that are only detected and recorded at periodic observation/assessment times. For each subject, observations are taken at finite discrete time points, and only the number of events that occur between consecutive observation times is known. Furthermore, the set of observation times can vary from subject to subject. Areas that usually produce panel count data include demographical studies, epidemiological studies, medical follow-up studies, oncology clinical trials, and reliability studies (Kalbfleisch and Lawless, 1985; Sun and Kalbfleisch, 1995; Wellner and Zhang, 2000; Sun and Zhao, 2013).

Many approaches have been proposed to analyze panel count data based on the counting process techniques. When no covariates are considered, inferences are focused on estimating the mean function of the counting process. For this purpose, Sun and Kalbfleisch (1995) constructed an isotonic regression estimator. Wellner and Zhang (2000) proposed two estimators by maximizing the pseudolikelihood and likeliood functions under the non-homogeneous Poisson process. Lu et al. (2007) pro-

posed likelihood-based estimators with the mean function being approximated by the monotone splines of Ramsay (1988) and showed that their spline-based estimators are more efficient than those in Wellner and Zhang (2000). Regarding the comparison of the mean functions for different populations, Sun and Fang (2003), Zhang (2006), and Balakrishnan and Zhao (2009) proposed different nonparametric tests for univariate panel count data. Li et al. (2014) developed nonparametric tests for multivariate panel count data.

When covariates are available, semiparametric regression analysis is widely used to examine the covariate effects on the response as well as the estimation of the mean function. Among others, Sun and Wei (2000) developed estimation procedures with time-dependent covariates on both observational and censoring processes. Zhang (2002) proposed a pseudolikelihood approach and Wellner and Zhang (2007) studied both pseudolikelihood and likelihood methods under the non-homogeneous Poisson process model. Hu et al. (2003) proposed two estimation approaches with different assumptions on the observational process. Lu et al. (2009) modeled the baseline mean function with monotone B-splines and established the asymptotic properties of their spline-based estimators. He et al. (2008) considered the regression analysis of multivariate panel count data. There are also many approaches developed for the cases that the recurrent event process and the observational process are dependent. For such work, we refer to Sun and Zhao (2013) for a comprehensive review.

In this chapter, we study semiparametric regression analysis of panel count data while taking into account within-subject correlation. Such within-subject correlation naturally exists because panel counts are repeatedly measured from the same subject. Ignoring such within-subject correlation may lead to serious problems as shown in our simulation studies. Existing work allowing for within-subject correlation in panel count data is limited. Zhang and Jamshidian (2003) proposed an EM algorithm based

on the gamma frailty Poisson model but without incorporating covariates. Recently, Hua and Zhang (2012) developed a spline-based semiparametric projected generalized estimating equation approach and modeled the baseline mean function with monotone cubic B-splines. Their method does not require the Poisson assumption, and their estimating equation can be regarded as the score equation of the marginal likelihood under the gamma frailty Poisson model when the frailty variance parameter is known. Hua et al. (2014) essentially adopted the same computational algorithm as Hua and Zhang (2012) under the gamma frailty Poisson model and established the asymptotic properties of their spline-based estimators. Although their models allow for handling within-subject correlation, none of the above papers derived or estimated such within-subject correlations.

In this chapter, a maximum likelihood approach is proposed for analyzing panel count data under the gamma frailty Poisson model when there is within-subject correlation. The within-subject correlation is quantified by Pearson's correlation coefficient in an explicit form. Monotone splines of Ramsay (1988) is adopted to model the baseline mean function, and all the parameters are estimated jointly through an efficient EM algorithm. The EM algorithm is robust to initial values, converges fast, and provides variance estimates of all parameters in closed form. Our approach has a good performance under the gamma frailty model and can be also applied to panel count data where there is no within-subject correlation. An R package `PCDSpline` has been developed based on our method and is now available on R CRAN for public use. Discussions on the differences between our method and that in Hua et al. (2014) can be found in the later sections.

The remainder of the article is organized as follows. Section 2.2 presents some notations, the model, the observed likelihood, and the modeling of the baseline mean function with monotone splines. Section 2.3 gives details of our proposed approach

including a data augmentation, the derivation of an EM algorithm, the variance estimates, and also a brief discussion of a simplified version adapted for the case where there is no within-subject correlation. Section 2.4 evaluates the performance of our approach through simulations. Section 2.5 provides two real-life applications from a skin cancer study and a bladder tumor study. Section 2.6 concludes with some discussions.

2.2 The proposed model

Notation, model, and likelihood

Consider a study that consists of n independent subjects. We assume that the observational process and the recurrent event process are conditionally independent given covariates. Let $\boldsymbol{x}_i$ denote a vector of $p \times 1$ time-independent covariates and $\{t_{ij}, j = 1, \ldots, K_i\}$ denote the actual observational times for subject i, where K_i is the number of observations and t_{iK_i} is the last observation time. Let $N_i(t)$ denote the counting process for subject i, and this process is observed only at t_{ij}'s. In order to account for the within-subject correlation, we propose a gamma frailty non-homogeneous Poisson process model for the recurrent event process $N_i(t)$. Specifically, conditional on ϕ_i, the frailty associated with subject i, $N_i(t)$ is a non-homogeneous Poisson process with mean function $\mu_0(t)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i$, where $\mu_0(t)$ is an unspecified nondecreasing baseline mean function with $\mu_0(0) = 0$ and ϕ_i's are independently and identically distributed from $\mathcal{G}a(\nu, \nu)$ with mean 1 and variance ν^{-1}. This model implies

$$N_i(t)|\phi_i \sim \mathcal{P}\{\mu_0(t)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i\}$$

for any $t \geq 0$, where $\mathcal{P}(a)$ denotes the Poisson distribution with mean a. In this model the mean constraint of the frailty distribution is made to avoid non-identifiability be-

cause $\mu_0(\cdot)$ is unspecified. Under the proposed gamma frailty Poisson process model, $\mu_0(\cdot)$ is the conditional baseline mean function of the recurrent event process given the frailty but can also be interpreted as the marginal baseline mean function since

$$\mathrm{E}\{N_i(t)\} = \mathrm{E}[\mathrm{E}\{N_i(t)|\phi_i\}] = \mathrm{E}\{\mu_0(t)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i\} = \mu_0(t)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}).$$

Under the proposed model, the common frailty among the panel counts within the same subjects induces within-subject correlation, while the panel counts for different subjects are independent. The ϕ_i's represent the heterogeneity not explained by the covariates among the subjects, and the variance parameter ν attributes to the degree of the within-subject association between the counts of recurrent events within non-overlapping time intervals. To quantify such correlation, consider two non-overlapping intervals $(t_1, t_2]$ and $(t_3, t_4]$, and let Z_1 and Z_2 denote the count of the recurrent events within these two intervals, respectively, from the same subject with covariates $\boldsymbol{x}$. As shown in Appendix A.2, under the gamma frailty Poisson process model, Pearson's correlation coefficient between Z_1 and Z_2 takes the following form,

$$\rho(Z_1, Z_2) = \{(1 + \lambda_1^{-1}\nu)(1 + \lambda_2^{-1}\nu)\}^{-1/2}, \tag{2.1}$$

where $\lambda_1 = \{\mu_0(t_2) - \mu_0(t_1)\}\exp(\boldsymbol{x}'\boldsymbol{\beta})$ and $\lambda_2 = \{\mu_0(t_4) - \mu_0(t_3)\}\exp(\boldsymbol{x}'\boldsymbol{\beta})$ are the mean numbers of the recurrent events occurring within $(t_1, t_2]$ and $(t_3, t_4]$, respectively. It is clear from equation (2.1) that ρ depends not only on ν but also on the mean numbers of recurrent events within the two considered time spans. However, in general, the larger value of ν is (corresponding to a smaller variance of the frailties), the smaller the within-subject association is. It is interesting to note two extreme cases. As $\nu \to \infty$, $\rho \to 0$ corresponds to the independent case where there is no within-subject correlation. As $\nu \to 0$, $\rho \to 1$ indicates a perfect linear correlation between such counts.

Define $Z_{ij} = N_i(t_{ij}) - N_i(t_{ij-1})$, the count of recurrent events within time interval $(t_{ij-1}, t_{ij}]$ for $j = 1, \ldots, K_i$ and $i = 1, \ldots, n$. Define $t_{i0} = 0$ for each i for notation convenience. By the properties of non-homogeneous Poisson process, all Z_{ij}'s are conditionally independent given ϕ_i for all j, and Z_{ij}'s have a Poisson distribution conditional on ϕ_i with

$$Z_{ij}|\phi_i \sim \mathcal{P}[\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i]$$

for each i. Thus, the observed likelihood given the observed data $\mathcal{D} = \{(t_{ij}, N_i(t_{ij}), \boldsymbol{x}_i) : j = 1, \ldots, K_i, i = 1, \ldots, n\}$ can be written as

$$\mathcal{L}_{obs} = \prod_{i=1}^{n} \int \prod_{j=1}^{K_i} P(Z_{ij}|\phi_i) g(\phi_i) d\phi_i,$$

where $P(Z_{ij}|\phi_i)$ is the conditional probability mass function of Z_{ij} given ϕ_i and $g(\phi_i)$ is the gamma density function of ϕ_i with both the shape and rate parameters equal to ν. It is straightforward to derive the following explicit form of the observed likelihood,

$$\mathcal{L}_{obs} = \prod_{i=1}^{n} \frac{\nu^\nu \Gamma(\nu + Z_{i\cdot})}{\Gamma(\nu)\{\nu + \mu_0(t_{iK_i}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^{\nu + Z_{i\cdot}}} \prod_{j=1}^{K_i} \frac{[\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})]^{Z_{ij}}}{Z_{ij}!}, \tag{2.2}$$

where $Z_{i\cdot} = \sum_{j=1}^{K_i} Z_{ij}$ is the total count of events for subject i, $i = 1, \ldots, n$. The unknown parameters of interest in the above likelihood are the regression parameters $\boldsymbol{\beta}$, the baseline mean function $\mu_0(\cdot)$, and the frailty variance parameter ν.

Monotone splines

Estimating the baseline mean function $\mu_0(\cdot)$ is challenging because it is infinitely dimensional. The number of parameters involved in μ_0 is on the order of sample size when the observation times differ from subject to subject. To handle this situation, we propose to approximate the baseline mean function $\mu_0(t)$ with monotone spline of

Ramsay (1988) in the following manner,

$$\mu_0(t) = \sum_{l=1}^{L} \gamma_l b_l(t), \tag{2.3}$$

where b_l's are the integrated spline basis functions, each of which is nondecreasing from 0 to 1, and γ_l's are nonnegative spline coefficients. The monotone spline expression (2.3) is very flexible to approximate nondecreasing functions as seen in Ramsay (1988), and Wang and Dunson (2011) among many others. There are two key components in specifying monotone spline basis functions: knots and degree. Although these two components work together to determine the spline basis functions, the degree mainly controls the smoothness of functions, and the knot placement mainly controls the shape of those functions. Once knot placement and degree are determined, the spline basis functions are deterministic. To construct the spline basis functions, one needs to specify a sequence of increasing m time points as knots and to specify a value d for the degree. The degree d could take on values 1, 2 and 3 for linear, quadratic, and cubic functions, respectively. The total number L of spline basis functions, or the number of spline coefficients, is determined by $L = m + d - 2$. One major advantage of using monotone splines is that it leads to only a finite number of parameters to estimate while maintaining great modeling flexibility.

The knot placement plays an important role in determining the shapes of the monotone splines and may potentially affect the performance of an estimation approach. Ramsay (1988) recommended to use a few knots such as at the median or at the quartiles, while Lin and Wang (2010) recommended to use 10$\sim$30 equally-spaced interior knots in their Bayesian methods to guarantee adequate modeling flexibility. Following Rosenberg (1995) and McMahan et al. (2013), we propose to use different numbers of equally-spaced knots within the time range and to select the "best" model by using some model selection criteria such as Akaike information criterion (AIC) or Bayesian information criterion (BIC). The same strategy can be used for determin-

ing the best value for the degree. However, it is observed that using different values for degree usually does not affect much of the performance of an estimation method when using monotone splines (Lin and Wang, 2010; Wang and Dunson, 2011; Cai et al., 2011). Hence, the degree was set to 2 to ensure adequate smoothing in all our simulation studies and real data applications.

2.3 The proposed estimation approach

A data augmentation

Under the monotone spline expression (2.3), there are only a finite number of unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}', \nu)'$ in the observed likelihood (2.2), where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_L)'$ are the spline coefficients. However, finding the maximum likelihood estimate of $\boldsymbol{\theta}$ by maximizing the observed likelihood (2.2) is challenging due to its complex form. It is our experience that both general statistical routines and the Newton-related algorithms fail to provide converged results because they are very sensitive to the initial values of the spline coefficients. To solve this problem we develop an EM algorithm based on the following data augmentation.

We first consider the following conditional likelihood by taking ϕ_i's as missing values,

$$\mathcal{L}_{con}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{j=1}^{K_i} P(Z_{ij}|\phi_i) g(\phi_i).$$

In order to take advantage of the Poisson likelihood and additive form of spline expression (2.3), we decompose Z_{ij} into the sum of L conditionally independent Poisson latent variables $\{Z_{ijl}\}_{l=1}^{L}$ given ϕ_i, for each i and j, such that

$$Z_{ij} = \sum_{l=1}^{L} Z_{ijl}, \quad Z_{ijl}|\phi_i \sim \mathcal{P}[\gamma_l\{b_l(t_{ij}) - b_l(t_{ij-1})\} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i], \quad \forall\ l = 1, \ldots, L.$$

The augmented data likelihood, given all Z_{ijl}'s and ϕ_i's, has the following form,

$$\mathcal{L}_c(\boldsymbol{\theta}) = \prod_{i=1}^{n} g(\phi_i) \prod_{j=1}^{K_i} \prod_{l=1}^{L} P(Z_{ijl}|\phi_i),$$

where $P(Z_{ijl}|\phi_i)$ is the Poisson probability mass function of Z_{ijl} given ϕ_i. This augmented data likelihood will serve as the complete data likelihood for the derivation of our EM algorithm.

The EM algorithm

To derive the EM algorithm, we first take the expectation of $\log\mathcal{L}_c(\boldsymbol{\theta})$ with respect to the latent variables Z_{ijl}'s and ϕ_i's conditional on the observed data $\mathcal{D}$ and the current parameter value $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)'}, \boldsymbol{\gamma}^{(d)'}, \nu^{(d)})'$. This yields $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)}) + H_2(\nu, \boldsymbol{\theta}^{(d)}) + H_3(\boldsymbol{\theta}^{(d)})$, where

$$\begin{aligned} H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)}) &= \sum_{i=1}^{n}\sum_{j=1}^{K_i} Z_{ij}\boldsymbol{x}_i'\boldsymbol{\beta} - \sum_{i=1}^{n} \mu_0(t_{iK_i}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \mathbf{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) \\ &\quad + \sum_{i=1}^{n}\sum_{j=1}^{K_i}\sum_{l=1}^{L} \mathbf{E}(Z_{ijl}|\mathcal{D}, \boldsymbol{\theta}^{(d)}) \log(\gamma_l), \\ H_2(\nu, \boldsymbol{\theta}^{(d)}) &= \nu \sum_{i=1}^{n} \mathbf{E}\{\log(\phi_i)|\mathcal{D}, \boldsymbol{\theta}^{(d)}\} - \nu \sum_{i=1}^{n} \mathbf{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) \\ &\quad + n\nu\log(\nu) - n\log\Gamma(\nu), \end{aligned}$$

and $H_3(\boldsymbol{\theta}^{(d)})$ is a function of $\boldsymbol{\theta}^{(d)}$ but free of $\boldsymbol{\theta}$. This suggests that one can separate ν from $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ in the Q function, and this will save the computation cost in the maximization step. All the conditional expectations involved in the Q function have closed forms as follows,

$$\begin{aligned} \mathrm{E}(Z_{ijl}|\mathcal{D}, \boldsymbol{\theta}^{(d)}) &= \frac{\gamma_l^{(d)}\{b_l(t_{ij}) - b_l(t_{ij-1})\}}{\mu_0^{(d)}(t_{ij}) - \mu_0^{(d)}(t_{ij-1})} Z_{ij}, \\ \mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) &= \frac{\nu^{(d)} + Z_{i\cdot}}{\nu^{(d)} + \mu_0^{(d)}(t_{iK_i}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}^{(d)})}, \end{aligned}$$

and

$$\mathrm{E}\{\log(\phi_i)|\mathcal{D},\boldsymbol{\theta}^{(d)}\} \;=\; \psi(\nu^{(d)}+Z_{i\cdot}) - \log\{\nu^{(d)}+\mu_0^{(d)}(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\},$$

where $\psi(\cdot)=\Gamma'(\cdot)/\Gamma(\cdot)$ is the digamma function. These conditional expectations can be easily derived by noting that the conditional distribution of ϕ_i given the observed data $\mathcal{D}$ is $\mathcal{G}a\{\nu+Z_{i\cdot},\nu+\mu_0(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}$ and that the conditional distribution of $(Z_{ij1},\cdots,Z_{ijL})$ given the observed data $\mathcal{D}$ is a multinomial distribution for $j=1,\ldots,K_i$ and $i=1,\ldots,n$.

In the M-step one can find $\boldsymbol{\theta}^{(d+1)}=\underset{\theta}{\mathrm{argmax}}Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(d)})$. Since H_2 is the only function involving ν in the Q function and H_2 does not involve $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, one can maximize $H_2(\nu,\boldsymbol{\theta}^{(d)})$ directly to obtain $\nu^{(d+1)}$. To obtain $\boldsymbol{\beta}^{(d+1)}$ and $\boldsymbol{\gamma}^{(d+1)}$, we consider the following first partial derivatives of $H_1(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\beta}$ and γ_l's,

$$\partial H_1/\partial\boldsymbol{\beta} \;=\; \sum_{i=1}^{n}\left\{Z_{i\cdot}-\mu_0(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i|\mathcal{D},\boldsymbol{\theta}^{(d)})\right\}\boldsymbol{x}_i,$$

and

$$\partial H_1/\partial\gamma_l \;=\; -\sum_{i=1}^{n}b_l(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i|\mathcal{D},\boldsymbol{\theta}^{(d)})+\sum_{i=1}^{n}\sum_{j=1}^{K_i}\frac{\mathbf{E}(Z_{ijl}|\mathcal{D},\boldsymbol{\theta}^{(d)})}{\gamma_l},$$

where $l=1,\cdots,L$. Setting $\partial Q/\partial\gamma_l=0$ leads to a closed-form solution for γ_l as a function of $\boldsymbol{\beta}$,

$$\gamma_l(\boldsymbol{\beta})=\frac{\sum_{i=1}^{n}\sum_{j=1}^{K_i}\mathbf{E}(Z_{ijl}|\mathcal{D},\boldsymbol{\theta}^{(d)})}{\sum_{i=1}^{n}b_l(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i|\mathcal{D},\boldsymbol{\theta}^{(d)})},\quad l=1,\cdots,L. \tag{2.4}$$

We can plug the closed-form expressions of γ_l's from (2.4) into the estimating equation $\partial H_1/\partial\boldsymbol{\beta}=0$ and solve it for $\boldsymbol{\beta}^{(d+1)}$. Then we can update $\boldsymbol{\gamma}^{(d)}$ using equation (2.4) with $\boldsymbol{\beta}$ replaced by $\boldsymbol{\beta}^{(d+1)}$.

Now we summarize our EM algorithm as below: initialize $\boldsymbol{\theta}^{(d)}=(\boldsymbol{\beta}^{(d)'},\boldsymbol{\gamma}^{(d)'},\nu^{(d)})'$ with $d=0$ and repeat the following three steps until convergence.

1. Obtain $\boldsymbol{\beta}^{(d+1)}$ by solving the following system of p equations

$$\sum_{i=1}^{n}\left[Z_{i\cdot}-\sum_{l=1}^{L}\left\{\frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i|\mathcal{D},\boldsymbol{\theta}^{(d)})\{\sum_{c=1}^{n}\sum_{j=1}^{K_c}\mathbf{E}(Z_{cjl}|\mathcal{D},\boldsymbol{\theta}^{(d)})\}b_l(t_{iK_i})}{\sum_{c=1}^{n}b_l(t_{cK_c})\exp(\boldsymbol{x}_c'\boldsymbol{\beta})\mathbf{E}(\phi_c|\mathcal{D},\boldsymbol{\theta}^{(d)})}\right\}\right]\boldsymbol{x}_i=\mathbf{0}.$$

2. Calculate $\gamma_l^{(d+1)}$ for $l = 1, \ldots, L$ using

$$\gamma_l^{(d+1)} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{K_i} \mathbf{E}(Z_{ijl}|\mathcal{D}, \boldsymbol{\theta}^{(d)})}{\sum_{i=1}^{n} b_l(t_{iK_i}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}^{(d+1)}) \mathbf{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})}.$$

3. Calculate $\nu^{(d+1)}$ by maximizing $H_2(\nu)$ directly.

At each iteration of the EM algorithm, the first step involves solving a system of equations for the regression parameters, the second updates the spline coefficients in closed form, and the third maximizes a univariate function. In step 1 the estimating equations have a unique solution $\boldsymbol{\beta}^{(d+1)}$ and thus $\boldsymbol{\theta}^{(d+1)}$ is a unique maximizer of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$; see the sketched proof in Appendix A.3. Convergence of the EM algorithm is claimed when the maximum change of all unknown para meters is smaller than a prespecified tolerance value ϵ. It has been our experience through extensive simulation studies that the proposed EM algorithm converges fast and is robust to the initial values. These properties are pertained in a large part due to the unique solution of $\boldsymbol{\beta}^{(d+1)}$ and the closed-form expressions of $\gamma_l^{(d+1)}$'s, which are easy to update and also automatically satisfy the nonnegative constraints. Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\boldsymbol{\gamma}}', \widehat{\nu})'$ denote the converged value of $\boldsymbol{\theta}^{(d)}$ from the EM algorithm, and $\widehat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$.

Asymptotic properties and variance estimation

The asymptotic properties of $\widehat{\boldsymbol{\theta}}$ could be studied under two different assumptions: (1) the number and position of the knots are known a priori and do not depend on the sample size; or (2) the cardinality of the knot set grows with the sample size (e.g. as in Hua et al. (2014)). Proceeding under the first assumption implies that the baseline mean function $\mu_0(\cdot)$ can be expressed as a linear combination of monotone spline functions (2.3), while the second assumption allows for a consistent estimate of $\mu_0(\cdot)$ under less stringent assumptions. For the purpose of this work the asymptotic properties of the proposed estimator will be presented under the

former setting. Consequently, the general theory of maximum likelihood estimation guarantees, under the standard regularity conditions,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to N\{0, \mathcal{I}^{-1}(\boldsymbol{\theta})\}, \quad \text{as} \quad n \to \infty,$$

where $\mathcal{I}(\boldsymbol{\theta})$ is the usual Fisher information matrix.

To obtain the variance estimate of $\widehat{\boldsymbol{\theta}}$, we adopt Louis's method (Louis, 1982). Specifically the variance estimate of $\widehat{\boldsymbol{\theta}}$ is taken to be $I^{-1}(\widehat{\boldsymbol{\theta}})$, where $I(\boldsymbol{\theta})$ is the observed information matrix and takes the form

$$I(\boldsymbol{\theta}) = -\frac{\partial^2 Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \text{var}\left\{\frac{\partial \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}.$$

All the quantities involved in $\partial^2 Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ and $\text{var}\{\partial \log \mathcal{L}_c(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\}$ have closed-form expressions and can be evaluated easily from the output of our EM algorithm. The details of the formula for these quantities are presented in the Appendix A.1.

The case of no within-subject correlation

The proposed gamma frailty non-homogeneous Poisson model reduces to the regular non-homogeneous Poisson model studied by Wellner and Zhang (2007) and Lu et al. (2009) when all ϕ_i's are taken to be 1. This is the limiting case of the proposed model when $\nu \to \infty$, where there is no within-subject correlation. It is worth noting that the idea of our approach can be applied to this special case with much simplification. The resulting EM algorithm associated with this reduced model involves only solving a system of low-dimensional equations for the regression parameters and updating the spline coefficients in closed form at each iteration. There are fewer terms in the conditional expectations, covariances, and variances in this case because there are no frailty terms in the corresponding complete likelihood.

To distinguish the two methods, we call the proposed approach under the gamma frailty non-homogeneous Poisson model with monotone splines `GFNPMS` and name

the simplified approach under the non-homogeneous Poisson model `NPMS`. These two approaches are evaluated and compared on the simulated panel count data with and without the within-subject correlation. We remark that `NPMS` is computationally competitive to the approaches of Wellner and Zhang (2007) and Lu et al. (2009) although it is not the main focus of this article.

2.4 Simulation evidences

Extensive simulation studies were carried out to evaluate the performance of the proposed approach. Three different simulation scenarios were considered: (1) the true cases where the data were generated from gamma frailty Poisson models; (2) the independent case where the data were generated from the non-homogeneous Poisson model without frailty; and (3) the misspecified cases where the frailty distribution was misspecified in the gamma frailty Poisson model. We provided a general setting for these simulation scenarios as below. To generate the observational process for subject i, we first generated K_i from Poisson(6)+1 to ensure that there was at least one observational time, and then generated K_i gap times independently from an exponential distribution with a rate parameter 2. The counting process associated with subject i was generated from the following model,

$$N_i(t_{ij}) - N_i(t_{ij-1})|\phi_i \sim \mathcal{P}[\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\} \exp(x_{i1}\beta_1 + x_{i2}\beta_2)\phi_i],$$

where $\mu_0(t) = \log(1+t) + t^2$, $x_{i1} \sim \mathcal{N}(0, 0.5^2)$, $x_{i2} \sim$ Bernoulli(0.5), and the true values of (β_1, β_2) took (-1, 1) or (1, -1). The distribution of ϕ_i was taken to be different gamma distributions, a degenerated distribution at 1, and some misspecified distributions in Scenarios 1, 2, and 3, respectively. In each simulation setup, we generated 500 data sets, each with $n = 100$ subjects. The tolerance value ϵ for claiming convergence was taken to be 10^{-4} for all simulations.

Scenario 1 corresponded to the true model cases, where the ϕ_i's were generated from $\mathcal{G}a(\nu, \nu)$ with ν taking 0.5, 1, 4, or 16. To apply `GFNPMS` and `NPMS`, we used 6 equally-spaced interior knots within the data range and took a degree of 2 for the monotone spline specification. From our simulations, it took 142 seconds for `GFNPMS` to converge per data set, while it only took 12 seconds for `NPMS` per data set. The significant difference in the running times of two methods results from the fact that `GFNPMS` involves an additional parameter ν than `NPMS`. It turns out that maximizing $H_2(\nu)$ in `GFNPMS` is time-consuming because $H_2(\nu)$ has a flat region at the maximizer.

Table 2.1 summarized the simulation results on the estimation of (β_1, β_2, ν) from the two methods in terms of relative bias, the difference between the average of 500 point estimates and the true value divided by the true value; SSD, the sample standard deviation of 500 point estimates; ESE, the empirical standard error obtained using the Louis's method; and CP95, the 95% coverage probability based on Wald's confidence intervals.

As seen in Table 2.1, `GFNPMS` has an excellent performance in all parameter configurations. The relative biases are all close to zero, indicating that our proposed estimators are unbiased; the ESEs are close to the SSDs, indicating that the variance estimates using Louis's method are accurate; and the CP95 are close to 0.95, indicating that the asymptotical normality is valid. In contrast, although `NPMS` yields a small bias on average for each parameter configuration, the variation of those point estimates is larger than that from `GFNPMS` by comparing the SSDs between the two methods. More importantly, `NPMS` yields seriously underestimated variances, which consequently leads to much lower coverage probabilities than the nominal level 0.95. These results suggest that ignoring the within-subject correlation may lead to seriously inaccurate estimation and further misleading conclusions. The larger the within-subject correlation, the more severe such problems are.

It is interesting to observe in Table 2.1 that the variances of the regression parameter estimates from `GFNPMS` become smaller as the true value of ν increases. This is not surprising because as ν increases, the frailty variance decreases, and this leads to a decreased variation in the observed data, which further leads to the decreased variances of the regression estimates.

Both `GFNPMS` and `NPMS` provide smooth estimates of the baseline mean function. For comparison, we consider the adjusted mean squared error (AMSE) at each t

$$\text{AMSE}(\widehat{\mu}_0(t)) = \frac{1}{500}\sum_{j=1}^{500}\frac{\{\widehat{\mu}_0^{(j)}(t) - \mu_0(t)\}^2}{\{\mu_0(t)\}^2},$$

where $\widehat{\mu}_0^{(j)}$ is the estimate of μ_0 from the jth data set, $j = 1, \cdots, 500$. This definition differs from the usual MSE in that it adjusts for the scale of $\mu_0(\cdot)$ at different t's. Figure 2.1 plots the true baseline mean function and the average of the baseline mean function estimates from both methods (left panel) as well as the AMSE curves from both methods (right panel) when $(\beta_1, \beta_2, \nu) = (1, -1, 0.5)$. As seen in the left panel of Figure 2.1, the averaged baseline mean estimates from `GFNPMS` essentially overlaps with the true curve, while the one from `NPMS` shows some difference from the true curve in the later time period. The right panel of Figure 2.1 indicates that the AMSE from `GFNPMS` is smaller than that from `NPMS` for all t in the data range. These results suggest that `GFNPMS` has an excellent performance in estimating the baseline mean function while ignoring the within-subject correlation (i.e., `NPMS`) leads to inaccurate estimates.

Table 2.1 Simulation results from non-homogeneous Poisson process models with gamma frailty (GFNPMS) and without frailty (NPMS) in scenario 1 where the data were generated from the gamma frailty Poisson models. Summarized results include the relative bias (RBias), the sample standard deviation of the point estimates (SSD), the average of estimated standard errors (ESE), and the 95% coverage probability (CP95).

			GFNPMS				NPMS			
(β_1, β_2)	ν	Est	RBias	SSD	ESE	CP95	RBias	SSD	ESE	CP95
$(1, -1)$	0.5	$\widehat{\beta}_1$	-0.008	0.291	0.303	0.956	-0.045	0.423	0.031	0.114
		$\widehat{\beta}_2$	-0.007	0.297	0.291	0.952	-0.020	0.404	0.034	0.126
		$\widehat{\nu}$	0.044	0.077	0.074	0.948	—	—	—	—
	1	$\widehat{\beta}_1$	-0.008	0.228	0.213	0.926	-0.022	0.342	0.031	0.168
		$\widehat{\beta}_2$	0.005	0.220	0.209	0.948	0.011	0.305	0.034	0.164
		$\widehat{\nu}$	0.051	0.163	0.154	0.958	—	—	—	—
	4	$\widehat{\beta}_1$	0.014	0.119	0.115	0.944	0.007	0.172	0.031	0.32
		$\widehat{\beta}_2$	0.000	0.117	0.113	0.942	0.004	0.160	0.034	0.326
		$\widehat{\nu}$	0.040	0.605	0.716	0.972	—	—	—	—
	16	$\widehat{\beta}_1$	-0.004	0.068	0.069	0.956	-0.004	0.092	0.031	0.518
		$\widehat{\beta}_2$	0.000	0.067	0.067	0.942	0.001	0.080	0.033	0.602
		$\widehat{\nu}$	0.068	2.625	3.915	0.974	—	—	—	—
$(-1, 1)$	0.5	$\widehat{\beta}_1$	0.014	0.306	0.294	0.942	0.003	0.468	0.019	0.064
		$\widehat{\beta}_2$	-0.008	0.285	0.285	0.952	0.004	0.412	0.021	0.090
		$\widehat{\nu}$	0.046	0.072	0.069	0.936	—	—	—	—
	1	$\widehat{\beta}_1$	0.011	0.199	0.209	0.948	0.003	0.326	0.019	0.096
		$\widehat{\beta}_2$	0.001	0.194	0.203	0.962	0.013	0.289	0.021	0.114
		$\widehat{\nu}$	0.046	0.150	0.142	0.946	—	—	—	—
	4	$\widehat{\beta}_1$	0.010	0.107	0.107	0.956	0.017	0.165	0.019	0.184
		$\widehat{\beta}_2$	-0.004	0.105	0.105	0.956	-0.005	0.141	0.020	0.216
		$\widehat{\nu}$	0.049	0.550	0.643	0.982	—	—	—	—
	16	$\widehat{\beta}_1$	0.001	0.061	0.059	0.956	0.002	0.086	0.019	0.336
		$\widehat{\beta}_2$	-0.001	0.059	0.059	0.944	-0.004	0.073	0.020	0.436
		$\widehat{\nu}$	0.040	2.520	3.050	0.968	—	—	—	—

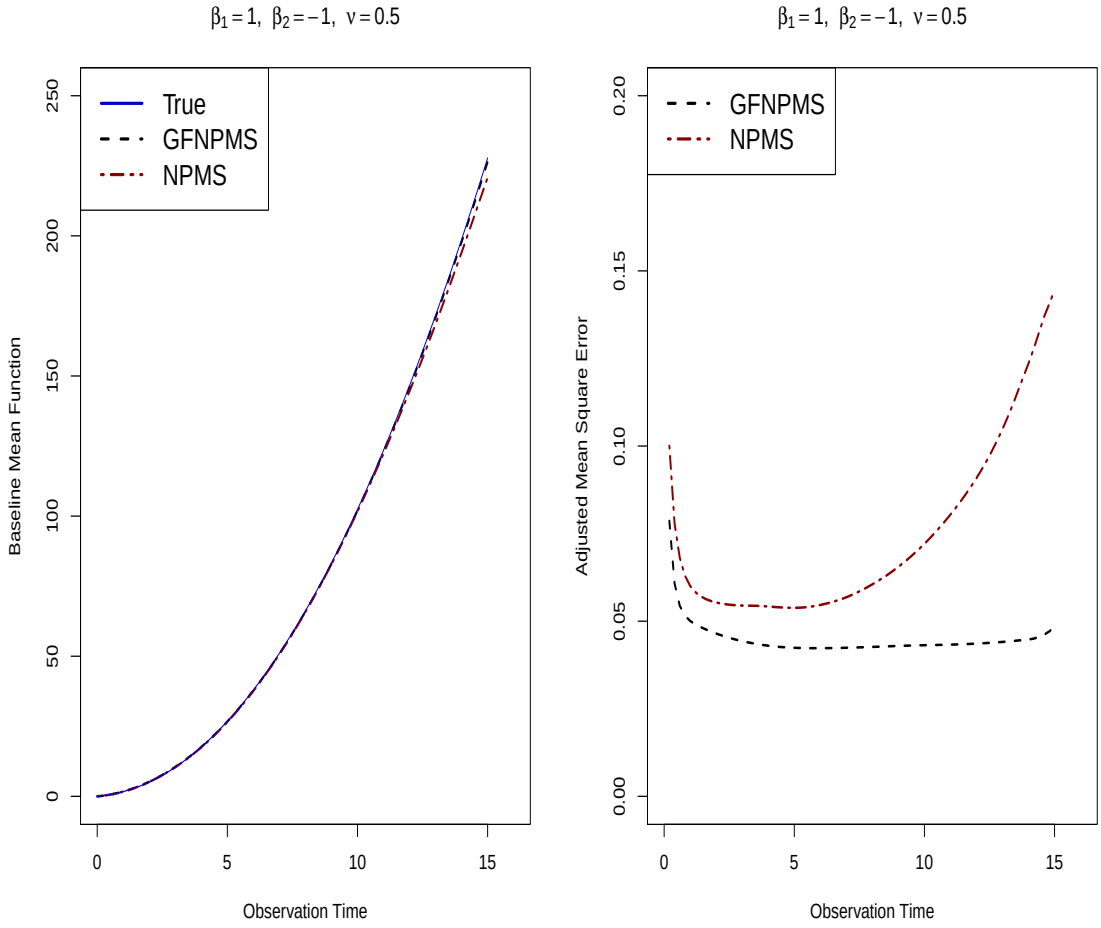

Figure 2.1 Left panel: the true baseline mean function and the average of the estimated baseline mean curves under the non-homogeneous Poisson models with gamma frailty (GFNPMS) and without frailty (NPMS). Right panel: the adjusted mean squared error as a function of t from GFNPMS and NPMS.

Scenario 2 corresponded to the independent case, where the ϕ_i's were all taken to be 1. The true model is the non-homogeneous Poisson model without frailty, under which `NPMS` is specifically developed. When applying `GFNPMS`, we specified a upper bound M for ν in step 3 of the EM algorithm in Section 3.3, where M was a large value taking 1,000 or 10,000 in the simulation. It was found that $\widehat{\nu}$ always reached the upper bound M. This makes sense because the corresponding true value of ν is ∞ under the gamma frailty non-homogeneous Poisson model when there is no

within-subject correlation. Table 2.2 presents the simulation results on the regression parameters from `GFNPMS` and `NPMS`. It is clear that both methods perform very well with very small relative biases in the point estimates, SSDs close to ESEs, and CP95 close to 0.95 for all parameter configurations. This suggests that `GFNPMS` works well in estimating the regression parameters even when there is no within-subject correlation.

Table 2.2 Simulation results from GFNPMS and NPMS in scenario 2 when the data were generated from non-homogeneous Poisson models. Summarized results include the relative bias (RBias), the sample standard deviation of the point estimates (SSD), the average of estimated standard errors, and the 95% coverage probability.

		NPMS				GFNPMS			
(β_1, β_2)	Est	RBias	SSD	ESE	CP95	RBias	SSD	ESE	CP95
$(1, -1)$	$\widehat{\beta}_1$	-0.001	0.032	0.031	0.954	-0.001	0.032	0.033	0.958
	$\widehat{\beta}_2$	0.003	0.035	0.034	0.956	0.003	0.034	0.035	0.968
$(-1, 1)$	$\widehat{\beta}_1$	-0.001	0.019	0.019	0.948	-0.001	0.019	0.021	0.968
	$\widehat{\beta}_2$	0.001	0.020	0.020	0.938	0.001	0.021	0.022	0.952

Scenario 3 corresponded to the misspecified cases of the frailty distribution, where ϕ_i's were generated from a lognormal distribution $\mathcal{LN}(-0.5, 1)$ with a shape parameter -0.5 and a scale parameter 1 or from a mixture distribution $0.4\mathcal{G}a(1, 1) + 0.6\mathcal{LN}(-0.5, 1)$. The results from the proposed method were shown in Table 2.3. From Table 2.3, our method produces underestimated variance estimates (ESEs less than SSDs) and the coverage probabilities below the nominal level in these misspecified cases. Thus, we conclude that the proposed model is not robust to frailty distribution misspecification, i.e., using our method may lead to misleading conclusion when the gamma frailty assumption does not hold. This makes sense because there are multiple counts from the same subjects which may provide adequate information to estimate the frailty distribution accurately. Note that this conclusion is different from that in Hua et al. (2014), which investigated different cases of frailty distribution

misspecification.

Additional simulations were conducted to compare with the approach of Hua et al. (2014) using their simulation settings. The two methods were found to have comparable performance in all the true cases and misspecified cases. Summary results on this comparison can be found in Appendix A.4.

Table 2.3 Simulation results from GFNPMS in scenario 3 when the frailty distribution is misspecified. In this simulation the true frailty distribution is either lognormal distribution $\mathcal{LN}(-0.5, 1)$ or a mixture of Gamma and lognormal distribution $0.4\mathcal{G}a(1,1) + 0.6\mathcal{LN}(-0.5, 1)$.

		Lognormal				Mixture Gamma			
(β_1, β_2)	Est	RBias	SSD	ESE	CP95	RBias	SSD	ESE	CP95
$(1, -1)$	$\widehat{\beta}_1$	-0.008	0.242	0.205	0.900	-0.005	0.183	0.160	0.924
	$\widehat{\beta}_2$	0.012	0.259	0.200	0.868	-0.003	0.185	0.156	0.910
$(-1, 1)$	$\widehat{\beta}_1$	-0.014	0.262	0.200	0.856	-0.005	0.190	0.154	0.882
	$\widehat{\beta}_2$	-0.018	0.260	0.192	0.828	0.005	0.178	0.150	0.910

2.5 Two illustrative real-life applications

The skin cancer study

In this section we apply the proposed method to the data from a skin cancer study conducted by the University of Wisconsin Comprehensive Cancer Center in Madison, Wisconsin (Li et al., 2013). The study was a double-blinded and placebo-controlled randomized Phase III clinical trial. The primary objective of the trial was to evaluate the effectiveness of 0.5 g/m^2/day PO difluoromethylornithine (DFMO) in reducing new skin cancers for patients with a history of non-melanoma skin cancers including basal cell carcinoma and squamous cell carcinoma. The study consisted of 291 patients who were randomized into two treatments for four to five years: placebo group and DFMO group. These patients were scheduled to have examinations every six months,

but the actual observational times varied from subject to subject. The numbers of recurrences of the new skin tumors between the observation times were recorded.

Among the 291 patients, all were white except for one Hispanic. In our data analysis, we excluded the only one Hispanic patient and one patient with missing cancer information. Hence, the final analytic sample consisted of 289 patients, 147 in the placebo group and 142 in the DFMO group. Let $\boldsymbol{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})'$ denote the covariate vector for patient i, where $x_{i1} = 1$ if patient i was in the DFMO group and 0 otherwise, x_{i2} and x_{i3} are the number of prior skin tumors and the age at the enrollment of patient i, and $x_{i4} = 1$ if patient i is male and 0 otherwise. We applied the proposed method to this data set using different numbers of equally-spaced interior knots between 0 and $1,880$ days for the monotone spline specification and obtained robust estimation results. The results can be found in the web-based supplementary materials. The tolerance value ϵ for claiming convergence was taken to be 10^{-8}. The model fit with 3 equally-spaced knots was chosen for our final analysis because it yielded the smallest AIC value. Table 2.4 shows the estimation results of these covariate effects when using 3 equally-spaced knots. As seen in Table 2.4, the DFMO treatment did not have a significant effect on the recurrence of the new skin tumors. Also the age and gender of patients did not seem to be significantly related to the tumor recurrence. However, the number of prior skin tumors had a significant effect on the occurrence of new skin tumors. These conclusions are consistent with those using the approaches of Zhang et al. (2013).

To give a specific illustration of the within-subject correlation, define Z_1 and Z_2 to be the random counts of skin tumors within the first six months and within the next six month for a patient with the median number ($x_2 = 1$) of skin cancer and the median age ($x_3 = 62$) at the enrollment. Using the expression (2.1), the Pearson's correlation coefficient between Z_1 and Z_2 was estimated to be 0.1168 (or 0.0933) if

the patient was male (female) in the DFMO treatment group, and the same measure was 0.1201 (or 0.0959) if the patient was male (female) in the placebo group. This suggests a weak within-subject correlation.

Figure 2.2 shows the estimated mean functions of new skin tumors for male and female patients in the DFMO and placebo groups, respectively, with the number of prior skin tumors being 2 and the age at the enrollment being 62. As seen in Figure 2.2, there seems to be a substantial difference between the estimated mean functions for different gender groups but little difference for the two treatment groups.

Table 2.4 Skin cancer data analysis from the proposed approach (GFNPMS). Summarized results are the point estimates (Point), the standard errors (SE), and the p-values for all the regression parameters and the frailty variance parameter ν.

	Point	SE	95% CI	p-value
$\widehat{\beta}_1$	-0.031	0.143	(-0.311, 0.249)	0.828
$\widehat{\beta}_2$	0.116	0.015	(0.087 ,0.145)	< 0.0001
$\widehat{\beta}_3$	-0.0008	0.0065	(-0.014, 0.012)	0.902
$\widehat{\beta}_4$	0.252	0.145	(-0.032, 0.536)	0.082
$\widehat{\nu}$	1.273	0.205	(0.871, 1.675)	< 0.0001

The bladder tumor study

We also applied the proposed method to the most widely used panel count data example in the literature, which arose from a bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group (Byar et al., 1977). In that study, 118 patients who had superficial bladder tumors were randomized into one of three treatment groups: placebo, thiotepa, and pyridoxine. During the study at each follow-up visit, new tumors since the last visit were counted, measured and then removed transurethrally. The primary objective of the study was to determine if any treatment could reduce the recurrence of bladder tumor. This data set has been analyzed extensively using many different approaches in the literature.

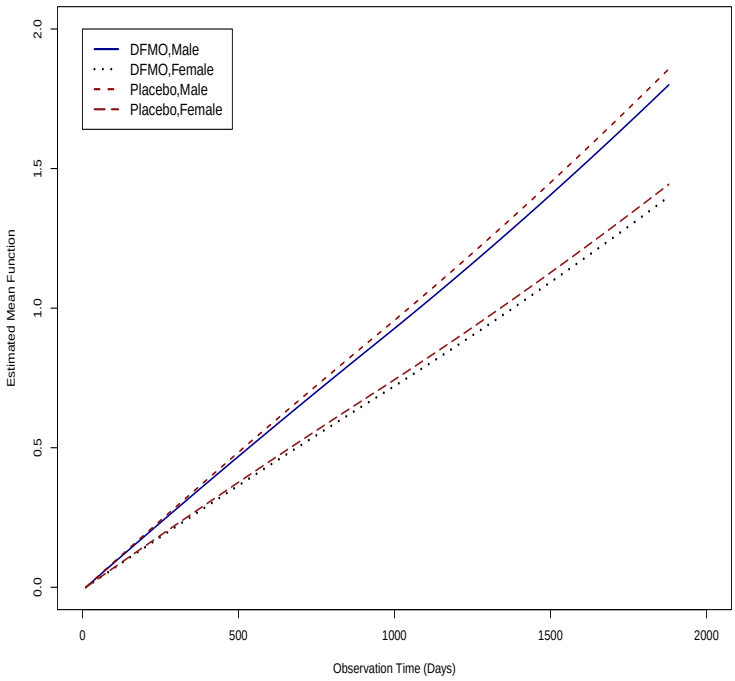

Figure 2.2 The estimated mean functions for different subgroups in the skin cancer data analysis.

Following Wellner and Zhang (2007) and Lu et al. (2009), we focused on 116 patients in the study, who had at least one follow-up observation after the study enrollment. Let $\boldsymbol{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})'$ denote the covariate vector for patient i, where x_{i1} and x_{i2} represent the number of bladder tumors and size of the largest bladder tumors for patient i at the beginning of the trial, and x_{i3} and x_{i4} are the binary indicators whether patient i was assigned to the treatment of pyridoxine pills and thiotepa installation, respectively. When applying the proposed method, we tried different numbers of equally-spaced knots with the data range $0 \sim 64$ months for the monotone spline specification. The tolerance value ϵ was taken to be 10^{-8} for claiming convergence.

Table 2.5 shows the results from our method and from two competitive approaches: Wellner and Zhang (2007) and Lu et al. (2009). The results from these two competi-

tors are directly drawn from their papers. Both of these two competitive approaches were likelihood-based approaches under the non-homogeneous Poisson model without considering the within-subject correlation. As seen in Table 2.5, the results from our method indicates that the number of initial bladder tumors was positively related to the recurrence of the tumor while the size of the largest tumor at the enrollment did not have a significant effect. It is found that the thiotepa instillation treatment significantly reduced the recurrence rate of bladder tumors, while the treatment of pyridoxine pills did not have a significant effect. These conclusions are consistent with those made in Wellner and Zhang (2007) and Lu et al. (2009). However, the proposed method seems to be more efficient than the two competitors in identifying significant covarites based on the reported p-values. This is because the proposed method accounts for the within-subject correlation, which seems not to be ignorable as indicated below. Our estimation results are similar to those from V3 of the GEE approach in Hua and Zhang (2012).

To quantify such within-subject correlation, we define Z_1 and Z_2 to be the random counts of bladder tumors within the first six months and within the next six months for a patient with the median number ($x_1 = 1$) of bladder tumors and the median size of the largest tumors at the enrollment. The Pearson's correlation coefficient between Z_1 and Z_2 was estimated to be 0.6993 if the patient was placebo group and was 0.4325 if the patient was in the thiotepa treatment group. This suggests a medium to large within-subject correlation.

Figure 2.3 plots the estimated mean functions of bladder tumor counts for the control and the two treatment groups. It is clear that the estimated mean functions for the control and the pyridoxine treatment groups are close to each other and they are different from the one for the thiotepa treatment group.

Table 2.5 Bladder tumor data analysis from the proposed approach (GFNPMS), the WZ approach in Wellner and Zhang (2007), and the LZH approach in Lu et al. (2009). Summarized results are the point estimates (Point), the standard errors (SE), and the p-values for all the regression parameters and the frailty variance parameter ν.

	GFNPMS			WZ			LZH		
	Point	SE	p-value	Point	SE	p-value	Point	SE	p-value
$\widehat{\beta}_1$	0.336	0.106	0.002	0.2069	0.0778	0.0078	0.208	0.083	0.012
$\widehat{\beta}_2$	0.012	0.120	0.920	-0.0355	0.0861	0.6801	-0.035	0.085	0.686
$\widehat{\beta}_3$	-0.033	0.409	0.936	0.0664	0.4310	0.8775	0.063	0.414	0.879
$\widehat{\beta}_4$	-1.140	0.435	0.009	-0.7972	0.3603	0.0269	-0.798	0.342	0.019
$\widehat{\nu}$	0.351	0.062	< 0.0001	—	—	—	—	—	—

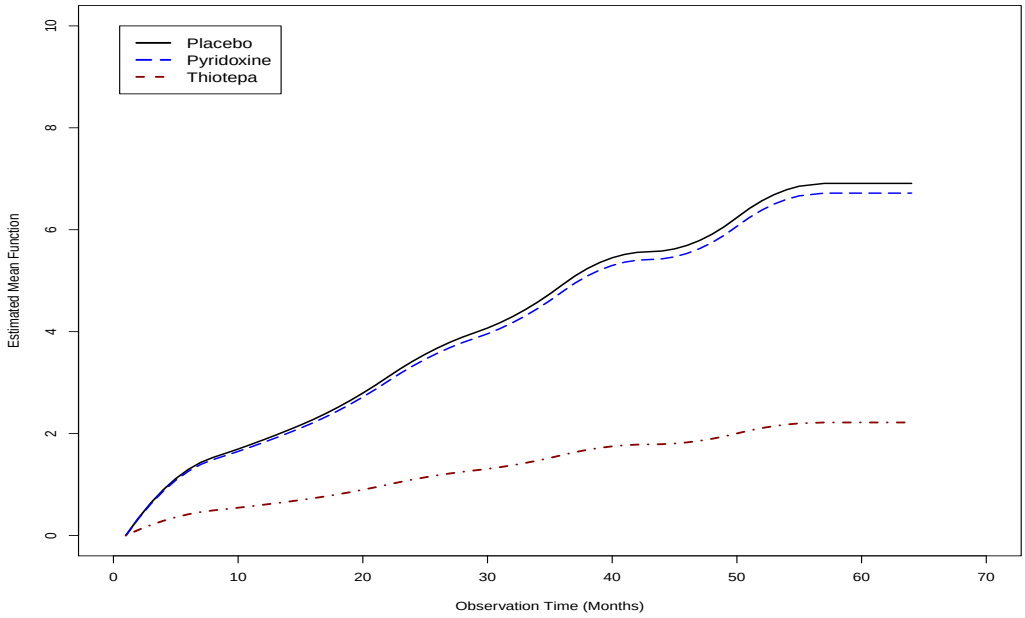

Figure 2.3 The estimated mean functions for different groups for the bladder tumor data analysis.

2.6 Discussions

In this chapter, we proposed a new estimation approach to analyze panel count data accounting for the within-subject correlation under the gamma frailty non-

homogeneous Poisson process model. We derived Pearson's correlation coefficient to quantify the within-subject correlation in closed form for subjects with arbitrary covariates. The modeling of baseline mean function with monotone splines leads to a finite number of parameters to estimate and thus save the computation effort. We developed a computationally efficient EM algorithm based on a Poisson data augmentation to jointly estimate all the unknown parameters. The EM algorithm is robust to initial values, easy to implement, and converges fast from our observations. Also, our approach provides variance estimates in closed form. The proposed method shows an excellent performance of estimating the regression parameters and the baseline mean function in both cases with and without the within-subject correlation in our simulation studies. Our approach is available for public use via an R package `PCDSpline` on CRAN.

Although studying the same topic with the same model, there are several differences between our approach and that in Hua et al. (2014). Firstly, our approach adopts monotone splines of Ramsay (1988) to approximate the baseline mean function, and the spline coefficients are updated in an explicit form in the EM algorithm where the nonnegative constraints of the spline coefficients are automatically satisfied. In contrast, Hua et al. (2014) used monotone B-splines, updated their spline coefficients using Newton-Raphson algorithm, and an additional step of isotonic regression was used to meet the ordering constraints for the spline coefficients. Secondly, we estimate the regression parameters, the spline coefficients, and the frailty variance parameter simultaneously, while Hua et al. (2014) developed a two-stage estimation procedure to estimate the frailty variance parameter in the first stage and other parameters in the second stage. Such two-stage estimation procedure may lead to underestimated variances and coverage probabilities for the regression parameter estimates. Thirdly, unlike Hua et al. (2014), our approach provides an estimate of frailty

variance and allows to estimate the within-subject correlation. Fourthly, our method provides variance estimates of all parameters in closed form using Louis's method. Our variance estimates are easier to calculate than those in Hua et al. (2014). The two approaches have a comparable performance from the simulation results shown in Appendix A.4. Overall our approach has computational advantages over that in Hua et al. (2014) in terms of easy implementation, and we have developed an R package `PCDSpline` to disseminate our approach. On the other hand, Hua et al. (2014) established the asymptotic results of their estimates under a more general assumption that the cardinality of the knot set grows with sample size, while our theoretical results are established under the assumption that the the number and position of the knots are known a priori and do not depend on the sample size.

The gamma frailty assumption plays an important role in the proposed method, and it is observed in the simulation studies that the estimation on the regression parameters may be biased when the gamma frailty assumption does not hold. Our future effort will be devoted to developing a goodness-of-fit test for checking the validity of the gamma frailty assumption based on the proposed method. Also, our approach is applicable for analyzing panel count data where there is a censoring variable that is independent of the recurrent event process, in that case the observed likelihood (2.2) still holds. The proposed approach does not apply directly when the censoring is informative, and this topic will be investigated in our future work.

Chapter 3

Regression analysis of current status data with generalized odds-rate hazards models

3.1 Introduction

Generalized odds-rate hazards (GORH) models represent a general class of semiparametric regression models for analyzing time-to-event data(Banerjee et al., 2007; Scharfstein et al., 1998) . The survival function of this family $S(\cdot|\boldsymbol{x})$ given covariates $\boldsymbol{x}$ is specified as

$$S(t|\boldsymbol{x}) = \{1 + \rho\Lambda_0(t)\exp(\boldsymbol{x}'\boldsymbol{\beta})\}^{-\rho^{-1}},\ t > 0, \tag{3.1}$$

where $\Lambda_0(\cdot)$ is an unspecified increasing function, $\boldsymbol{\beta}$ is $p \times 1$ vector of regression parameters denoting the covariate effects, and ρ is a positive constant. The GORH models have strong connections with other semiparametric regression models in the survival literature. For example, the limiting case of the GORH models when $\rho \to 0$ reduces to the popular proportional hazards (PH) model, while it becomes the proportional odds (PO) model when $\rho = 1$. Also, the accelerated failure time (AFT) model is a special case of the GORH models when $\Lambda_0(t) = t$ for $t > 0$ (Banerjee et al., 2007). Thus, the GORH family is more flexible than the PH model and PO model by introducing an additional parameter ρ. Such modeling flexibility brings desirable properties. For example, the GORH models allow time varying hazard ratios, while the PH model requires constant hazard ratios, an assumption that is often unrealistic in many real-life applications.

In the literature the GORH models are also well recognized as a special class of linear transformation models in the form of

$$g_\rho\{S(t|\boldsymbol{x})\} = \alpha(t) + \boldsymbol{x}'\boldsymbol{\beta},$$

with $g_\rho(s) = \log\{\rho^{-1}(s^{-\rho} - 1)\}$ for $\rho > 0$ and $\alpha(t) = \log\{\Lambda_0(t)\}$ or in the form of

$$\alpha(T) = -\boldsymbol{x}'\boldsymbol{\beta} + \epsilon,$$

where $\exp(\epsilon)$ follows a Pareto distribution with parameter $\rho > 0$. However, it is worth noting that the linear transformation models in the literature do not contain the additional parameter ρ in the transformation function g. The popular gamma frailty PH models, which are commonly used for modeling multivariate or clustered failure times, have the GORH models as the marginal distributions of the failure times.

Many approaches have been proposed for analyzing right-censored data using the GORH models. Among others, Harrington and Fleming (1982) proposed a G^ρ statistic for testing the regression parameters, and Dabrowska and Doksum (1988) studied estimation and testing, both for a two-sample setting. Clayton and Cuzick (1985) proposed a quasi-EM algorithm for the maximum likelihood estimators. Scharfstein et al. (1998) studied efficient estimation problem and proved that their estimates attain the semiparametric variance bound. Xu and Harrington (2001) studied the connections of the estimated treatment effects among the GORH models, the time-varying coefficient PH model, and the gamma frailty PH model under k-sample settings. Banerjee et al. (2007) proposed a Bayesian estimation approach using a piecewise constant baseline hazard function. For right-censored data with a cured subgroup, Zeng et al. (2006) proposed an efficient recursive algorithm for the maximum likelihood estimates (MLEs), and Castro et al. (2014) proposed a Bayesian method for modeling correlated survival events.

Although the GORH models have been widely used to analyze right-censored survival data, no research has been reported using the GORH models for studying current status data or interval-censored data in the literature to our knowledge. In this chapter, we study regression analysis of current status data under the GORH models. Current status data occur naturally in many social science, epidemiological, and biomedical studies. In such studies, subjects are only examined once for the failure event of interest. Thus, only the status of the failure event, whether has occured or not at the examination time, is known instead of observing the exact time of the event. Current status data contain only left- or right-censored observations and no exactly observed observations, which is quite different from the right-censored data in the literature. The structure of current status data brings great challenges in the regression analysis both computationally and theoretically. It was proved that the maximum likelihood estimates of the baseline cumulative hazard function (or the survival function) coverges to the true function at an $n^{-1/3}$ rate under the PH model (Huang, 1996) for current status data, which is much slower than the corresponding rate for right-censored data. The main reason is that current status data do not contain exactly observed failure times and thus have much less information for estimating the unknown parameters.

Most of the approaches in the aforementioned papers for right-censored data have assumed a fixed and known ρ. In particular, Hinkley and Runger (1984) commented "it is unresolved issue as to whether the variability of $\boldsymbol{\beta}$ should be affected by the selection of ρ". Zeng et al. (2006) pointed out the non-identifiabilities of the GORH models and discussed a few cases where ρ may be treated as unknown and be estimated under their cure rate model. Hanson and Yang (2007) fitted the GORH model to a real data application treating ρ as an unknown parameter using a Bayesian method based on polya tree but did not study the performance of their approach under the

GORH model because their paper focused on the PO model. Banerjee et al. (2007) estimated ρ together with other parameters but observed large bias in the estimation in their paper. These literature results naturally raise a question: are the parameters in the GORH models identifiable?

In this chapter, the non-identifiability issues of the GORH models are investigated, and it is found that the GORH models are non-identifiable in the case when there are no covariates but are indeed identifiable in the usual regression settings including the commonly seen two-sample or k-sample settings. However, the great challenges reported in the existing research seem to indicate that right-censored data contain little information about ρ in the GORH models. The same problems were found in our own experience when attempting to estimate ρ for current status data, which contain much less information than right-censored data. Following the conventional idea of fixing ρ when using the GORH models in the literature, we propose a computationally efficient estimation approach based on a novel EM algorithm for the regression analysis of current status data. This approach is shown to be robust to initial values, fast to converge, and provides variance estimates in closed form. We also propose a working model strategy to make valid inferences when the true value of ρ is unknown in the GORH models without losing much efficiency. Consequently, using the working model strategy only requires one to implement our approach with $\rho = 1$ but enjoys the great modeling flexibility of the whole GORH family.

The remainder of this chapter is organized as follows. Section 3.2 discusses the identifiability issues of the GORH models. Section 3.3 introduces some preparation work for the proposed approach including the notations and data likelihood, and the use of monotone splines for modeling the unknown function Λ_0. Section 3.4 gives the details of the proposed approach when ρ is known, including a data augmentation, the details of our EM algorithm, and the variance estimates. Section 3.5 proposes

a working model strategy with the proposed approach for analyzing current status data when ρ is unknown. The proposed approach and the working model strategy are evaluated through extensive simulation studies in Section 3.6 and are illustrated by a large size real-life data set from health care study in Section 3.7. Section 3.8 provides with some discussions.

3.2 Identifiability of GORH models

In this section we discuss the identifiability issues of the GORH models in (3.1). We have the following results for the GORH models without covariates.

Fact 1: The GORH models with $\boldsymbol{x}'\boldsymbol{\beta} = 0$ are non-identifiable.

The case $\boldsymbol{x}'\boldsymbol{\beta} = 0$ occurs when there are no covariates. In this case, the parameters in the GORH models are (ρ, Λ_0). To prove the non-identifiability, one only needs to find two different sets of parameters (ρ_1, Λ_{01}) and (ρ_2, Λ_{02}) such that the two models have the same survival function. To this end, for fixed (ρ_1, Λ_{01}), we take an arbitrary positive value of ρ_2 with $\rho_2 \neq \rho_1$ and define

$$\Lambda_{02} = \rho_2^{-1}[\{1 + \rho_1\Lambda_{01}(t)\}^{\rho_2\rho_1^{-1}} - 1]. \tag{3.2}$$

It is clear to see

$$S(t|\boldsymbol{x}, \rho_1, \Lambda_{01}) = \{1 + \rho_1\Lambda_{01}(t)\}^{-\rho_1^{-1}} = \{1 + \rho_2\Lambda_{02}(t)\}^{-\rho_2^{-1}} = S(t|\boldsymbol{x}, \rho_2, \Lambda_{02}).$$

This suggests that the two GORH models with different parameters (ρ_1, Λ_{01}) and (ρ_2, Λ_{02}) yield the same survival function and thus are not identifiable.

Now consider regression settings. First we consider a simple case of a regression setting: there is only one covariate and it is binary. This is also called two-sample setting in the literature. We then study the general setting when there are a mixture of continuous and binary covariates. Assume that the p covariates are not linearly

correlated, which means not any covariate can be written as a linear combination of the other covariates . We establish the following results.

Theorem 1: The GORH models defined in (3.1) are identifiable in the case of one binary covariate .

Theorem 2: The GORH models defined in (3.1) are identifiable in general regression settings.

The proofs of Theorems 1 and 2 are sketched in the Appendix B.1. These results suggest that the GORH models are identifiable in the usual regression settings.

3.3 Observed likelihood and monotone splines

Data and observed likelihood

Let T denote the failure time of interest, $\boldsymbol{x}$ be a $p \times 1$ vector of time-independent covariates, and C the censoring variable. For current status data, T is never exactly observed, but its status relative to C is known. That is, $\delta = I(T \leq C)$ is observed, where $I(\cdot)$ is an indicator function. Thus, current status data are a mixture of left- and right-censored failure times. Throughout this article it is assumed that the failure time T and the cesnoring time C are conditionally independent given covariates $\boldsymbol{x}$, a common assumption in the literature of current status data.

Let $\mathcal{D} = \{(C_i, \delta_i, \boldsymbol{x}_i), i = 1, \ldots, n\}$ be n independently and identically distributed (i.i.d.) copies of $(C, \Delta, \boldsymbol{x})$, where n is the total number of subjects in the study. Under the conditional independence assumption, the observed likelihood can be written as

$$L_{obs} = \prod_{i=1}^{n} \{1 - S(C_i|\boldsymbol{x}_i)\}^{\delta_i} \{S(C_i|\boldsymbol{x}_i)\}^{1-\delta_i}, \tag{3.3}$$

where $S(\cdot|\boldsymbol{x})$ is the survival function of the failure time T given the covariate vector $\boldsymbol{x}$ taking the form as in (3.1) under the GORH models. The main research interests are to assess the covariate effects and to estimate the survival functions for differ-

ent subgroups for prediction purpose. These require one to estimate the unknown parameters $\boldsymbol{\theta} = (\rho, \boldsymbol{\beta}, \Lambda_0)$.

Estimating ρ under the GORH model has been regarded as a tough issue in the literature for right-censored data as discussed and for current status data based on our own experience as mentioned in the introduction section. Different strategies are proposed for the cases whether or not ρ is treated as known in the estimation procedure.

Modeling Λ_0 with monotone splines

Another complication is to estimate the infinite-dimensional parameter Λ_0 because it is an unspecified increasing function. It is well known that under the PH model the partial likelihood allows one to estimate the regression parameters without the need of estimating the cumulative baseline hazard function Λ_0 for right-censored data (Cox, 1975). However, for current status data there does not seem to exist partial likelihood under the PH model and one has to estimate Λ_0 jointly with the regression parameters. It has been shown that the maximum likelihood estiamtes of Λ_0 has a convergence rate of $n^{-1/3}$ under the PH and PO models for current status data, which is slower than the rate of the corresponding estimates when dealing with right-censored data.

Using splines has been a popular way to model nonparameteric functions in the literature. It leads to a finite number of parameters to estimate while maintaining adaquate modeling flexibility by not assuming a specific shape for the unknown functions. In this chapter, we propose to model $\Lambda_0(\cdot)$ with the mononote splines of Ramsay (1988) following McMahan et al. (2013) and Cai et al. (2011). Specifically, $\Lambda_0(t)$ is approximated as a linear combination of monotone splines in the following

form,

$$\Lambda_0(t) = \sum_{l=1}^{L} \gamma_l b_l(t), \tag{3.4}$$

where b_l's are basis functions or integrated splines (Ramsay, 1988) and γ_l's are non-negative spline coefficients. These spline basis functions are piecewise polynomial functions, taking 0 at first stage, increasing in the second stage, and staying plateau in the third stage. To construct the spline basis functions, one needs to specify the degree d of basis functions and choose an increasing sequence of m points as knots within a time range. The degree controls the overall smoothness of the basis functions; for instance, degree 1, 2, and 3 correspond to linear, quadratic, and cubic basis functions, respectively. The placement of knots controls the overall shape of the basis functions together with the degree. The number of basis functions L is determined by the sum of the degree and the number of interior knots, i.e., $L = d + m - 2$ (Ramsay, 1988). Once the degree and knots are specified, the spline basis functions are fully determined.

3.4 The proposed approach when ρ is known

A data augmentation

Throughout Section 3.4, we assume ρ is fixed and known and try to make inferences on $(\boldsymbol{\beta}, \Lambda_0)$. Although the observed likelihood in (3.3) looks simple, maximizing it directly with respect to $(\boldsymbol{\beta}, \Lambda_0)$ is challenging. Based on our experience the Newton-Raphson or related algorithms often fail to converge when maximizing the observed likelihood (3.3) directly. To tackle this problem, we propose a novel EM algorithm for finding the maximum likelihood estimate of $(\boldsymbol{\beta}, \Lambda_0)$. The following three-stage data augmentation is essential for the derivation of our EM algorithm.

The first stage of our data augmentation utilizes the relationship between the GORH models and the gamma frailty PH models. By introducing gamma frailties ϕ_i's, one obtains an expanded data likelihood

$$\begin{aligned} L_{aug1}(\boldsymbol{\theta}) &= \prod_{i=1}^{n} \Big[1 - \exp\Big\{-\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i\Big\}\Big]^{\delta_i} \\ &\quad \times \exp\Big\{-\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i(1-\delta_i)\Big\} g(\phi_i), \end{aligned} \tag{3.5}$$

where ϕ_i's are i.i.d frailties from $g(\cdot)$, the gamma density function with both the shape and rate parameters equal to ρ^{-1}.

In the second stage, we introduce conditionally independent Poisson latent variables Z_i's with $Z_i|\phi_i \sim \mathcal{P}\{\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i\}$ for $i = 1, \ldots, n$. To connect with the observed data and likelihood, we let $\delta_i = I(Z_i > 0)$ for $i = 1, \ldots, n$. Thus, knowing Z_i's will determine all of the censoring indicators, and the augmented data likelihood based on Z_i's is

$$\begin{aligned} L_{aug2}(\boldsymbol{\theta}) &= \prod_{i=1}^{n} \frac{\{\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i\}^{Z_i}}{Z_i!} \exp\Big\{-\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i\Big\} g(\phi_i) \\ &\quad \times \{\delta_i I(Z_i > 0) + (1-\delta_i) I(Z_i = 0)\}. \end{aligned}$$

It is straightforward to show that integrating out Z_i's in L_{aug2} leads to the augmented likelihood (3.5) in the first stage.

In the third stage, we decompose each Z_i as a sum of conditionally independent Poisson random variables such that $Z_i = \sum_{l=1}^{L} Z_{il}$ given ϕ_i with $Z_{il}|\phi_i \sim \mathcal{P}\{\gamma_l b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i\}$ for $l = 1, \ldots, L$ and $i = 1, \ldots, n$. This augmentation takes advantage of monotone spline representation of Λ_0. The augmented likelihood based on all Z_{il}'s is given by

$$\begin{aligned} L_c &= \prod_{i=1}^{n} g(\phi_i) I\Big(Z_i = \sum_{l=1}^{L} Z_{il}\Big) \Big\{\delta_i I(Z_i > 0) + (1-\delta_i) I(Z_i = 0)\Big\} \\ &\quad \times \prod_{l=1}^{L} \exp\{-\gamma_l b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i\} \frac{\{\gamma_l b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i\}^{Z_{il}}}{Z_{il}!}. \end{aligned}$$

Integrating out Z_{il}'s subject to the constraints leads to the augmented likelihood L_{aug2} in the second stage. The resulting data likelihood is simply a product of Poisson probability mass functions, and it is used as the complete data likelihood for the derivation of our EM algorithm. The nice form of the complete likelihood is the key factor for the promising features of our EM algorithm.

The proposed EM algorithm

Now we describe the derivation of our EM algorithm in detail. The derivation starts with taking the conditional expectation of log-complete likelihood $\log L_c(\boldsymbol{\theta})$ with respect to all the latent variables including ϕ_i's, Z_i's and Z_{il}'s given the observed data $\mathcal{D}$ and the current parameter $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)'}, \boldsymbol{\gamma}^{(d)'})'$ at dth iteration. This yields $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = \mathrm{E}\{\log L_c(\boldsymbol{\theta})|\mathcal{D}, \boldsymbol{\theta}^{(d)}\} = H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)}) + H_2(\boldsymbol{\theta}^{(d)})$, where

$$\begin{aligned} H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)}) =& \sum_{i=1}^{n} \mathrm{E}(Z_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) \boldsymbol{x}_i'\boldsymbol{\beta} - \sum_{i=1}^{n} \Lambda_0(C_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) \\ &+ \sum_{i=1}^{n} \sum_{l=1}^{L} \mathrm{E}(Z_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)}) \log(\gamma_l), \end{aligned}$$

and $H_2(\boldsymbol{\theta}^{(d)})$ is a function of $\boldsymbol{\theta}^{(d)}$ but free of $\boldsymbol{\theta}$. The conditional expectations involved in the Q function all have explicit forms as follows,

$$\begin{aligned} \mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) &= \frac{1}{1 + \rho\Lambda_0^{(d)}(C_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}^{(d)})} I\{\delta_i = 0\} \\ &+ \frac{1 - \left\{1 + \rho\Lambda_0^{(d)}(C_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}^{(d)})\right\}^{-\rho^{-1}-1}}{1 - \left\{1 + \rho\Lambda_0^{(d)}(C_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}^{(d)})\right\}^{-\rho^{-1}}} I\{\delta_i = 1\}, \\ \mathrm{E}(Z_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) &= \frac{\Lambda_0^{(d)}(C_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}^{(d)})\delta_i}{1 - \left\{1 + \rho\Lambda_0^{(d)}(C_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}^{(d)})\right\}^{-\rho^{-1}}}, \\ \mathrm{E}(Z_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)}) &= \frac{\gamma_l^{(d)} b_l(C_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}^{(d)})\delta_i}{1 - \left\{1 + \rho\Lambda_0^{(d)}(C_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}^{(d)})\right\}^{-\rho^{-1}}}, \end{aligned}$$

where $\Lambda_0^{(d)}(C_i) = \sum_{l=1}^{L} \gamma_l^{(d)} b_l(C_i)$. These expressions are derived using the following facts: (i) the conditional distribution of ϕ_i given the observed data is a gamma distribution when $\delta_i = 0$ and is a mixture of gamma distribution when $\delta_i = 1$ based on the augmented data likelihood (3.5) in the first stage of the data augmentation; (ii) the conditional distribution of Z_i given ϕ_i and the observed data is a degenerated distribution at 0 when $\delta_i = 0$ and is a truncated Poisson distribution when $\delta_i = 1$ as seen from the second stage of the data augmentation; and (iii) the conditional distribution of $(Z_{i1}, \cdots, Z_{iL})$ given Z_i and the observed data is a multinomial distribution. The law of iterated expectation is then used for the derivation of $\mathrm{E}(Z_i|\mathcal{D})$ and $\mathrm{E}(Z_{il}|\mathcal{D})$ for each i and l.

The M-step of the EM algorithm requires to maximize $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\theta}$, and it is equivalent to maximize $H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})$. It can be shown that there is a unique global maximizer of $H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})$. To find the maximizer, consider the partial derivatives of $H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\beta}$ and γ_l's, and one obtains the following expressions,

$$\begin{aligned}
\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})/\partial \boldsymbol{\beta} &= \sum_{i=1}^{n} \Big\{ \mathrm{E}(Z_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) - \Lambda_0(C_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) \Big\} \boldsymbol{x}_i, \\
\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})/\partial \gamma_l &= -\sum_{i=1}^{n} b_l(C_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) + \frac{1}{\gamma_l} \sum_{i=1}^{n} \mathrm{E}(Z_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)}),
\end{aligned}$$

for $l = 1, \ldots, L$. Setting these partial derivatives to zero and solving the resulting system of equations for $\boldsymbol{\theta}$ yields $\boldsymbol{\theta}^{(d+1)}$, the maximizer of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$. Note that solving $\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})/\partial \gamma_l = 0$ leads to a closed-form solution for γ_l, as a function of $\boldsymbol{\beta}$,

$$\gamma_l^*(\boldsymbol{\beta}) = \frac{\sum_{i=1}^{n} \mathrm{E}(Z_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)})}{\sum_{i=1}^{n} b_l(C_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})}, \quad l = 1, \ldots, L.$$

Thus, one can replace γ_l with $\gamma_l^*(\boldsymbol{\beta})$ for each l in the equation $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})/\partial \boldsymbol{\beta} = 0$ and solve it for $\boldsymbol{\beta}^{(d+1)}$. Then $\gamma_l^{(d+1)}$ is obtained as $\gamma_l^*(\boldsymbol{\beta}^{(d+1)})$, for $l = 1, \ldots, L$. The

Appendix B.3 presents a proof that $\boldsymbol{\theta}^{(d+1)} = (\boldsymbol{\beta}^{(d+1)}, \boldsymbol{\gamma}^{(d+1)})$ is the unique global maximizer of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$.

Here is a succinct summary of the proposed EM algorithm. First, initialize $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)'}, \boldsymbol{\gamma}^{(d)'})'$ for $d = 0$, and then repeat the following two steps until convergence.

1. Obtain $\boldsymbol{\beta}^{(d+1)}$ by solving the following system of p equations:

$$\sum_{i=1}^{n} \left[\mathrm{E}(Z_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) - \left\{ \sum_{l=1}^{L} \gamma_l^{*(d)}(\boldsymbol{\beta}) b_l(C_i) \right\} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) \right] \boldsymbol{x}_i = \mathbf{0},$$

   where

$$\gamma_l^{*(d)}(\boldsymbol{\beta}) = \frac{\sum_{i=1}^{n} \mathrm{E}(Z_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)})}{\sum_{i=1}^{n} b_l(C_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})}, \quad l = 1, \ldots, L.$$

2. Calculate $\gamma_l^{(d+1)} = \gamma_l^{*(d)}(\boldsymbol{\beta}^{(d+1)})$, for $l = 1, \ldots, L$, and increase d by 1.

The convergence of the EM algorithm is claimed when the maximum change of all elements of $\boldsymbol{\beta}$ and γ_l's between successive iterations is less than a prespecified small value ϵ, say 10^{-4}. The system of equations in step 1 above have a unique solution and can be easily solved using existing root-finding procedures in the literature. Step 2 is a simple updating of the spline coefficients in closed form, which is an appealing point of our algorithm. Moreover, the updated $\gamma_l^{(d+1)}$'s are guaranteed to be nonegative based on the expressions, thus no extra effort is needed to take care of the nonnegative constraints for γ_l's.

Variance estimation

Let $\widehat{\boldsymbol{\theta}}$ denote the converged values of $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)'}, \boldsymbol{\gamma}^{(d)'})'$, and it is known that $\widehat{\boldsymbol{\theta}}$ is a maximum likelihood estimate of $\boldsymbol{\theta}$. Louis's method (Louis, 1982) is adopted to find the variance estimate of $\widehat{\boldsymbol{\theta}}$. Specifically, $\mathrm{var}(\widehat{\boldsymbol{\theta}})$ is taken to be the inverse of the observed information matrix $I(\widehat{\boldsymbol{\theta}})$, i.e., $\mathrm{var}(\widehat{\boldsymbol{\theta}}) = \{I(\widehat{\boldsymbol{\theta}})\}^{-1}$, and $I(\boldsymbol{\theta})$ can be obtained

using the missing information principle as follows,

$$I(\boldsymbol{\theta}) = -\frac{\partial^2 Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \mathrm{var}\Big(\frac{\partial \log\{L_c(\boldsymbol{\theta})\}}{\partial \boldsymbol{\theta}} \Big| \mathcal{D}, \widehat{\boldsymbol{\theta}} \Big). \tag{3.6}$$

All the quantities involved to calculate the last two terms in expression (3.6) are find to have closed from and the details are shown in the Appendix B.2. Those closed-form expressions make the variance estimate easy to compute, which is another appealing point of our approach.

3.5 Strategies when ρ is unknown

In Section 3.4, we propose an estimation approach for regression analysis of current status data under the GORH model by assuming ρ is fixed and known. Although the approach is valid for all possible positive values of ρ, the GORH model essentially loses much of its flexibility and is only comparable to the usual semiparametric regression models such as the PH model and PO model when ρ is known. In this section, we discuss a few strategies in order to take advantage of the great modeling flexibility of the GORH models when ρ is unknown.

The first strategy is to treat ρ as an unknown parameter and develop an estimation approach to estimate ρ together with other parameters $\boldsymbol{\beta}$ and Λ_0. We have attempted this idea and generalized our approach in Section 3.4. The resulting EM algorithm contains only one additional step of maximizing a concave function for ρ and the variance estimates can be derived in closed form using Louis's method. However, the new EM algorithm encountered many numerical problems from our simulation. For example, it fails to converge for many simulated data sets, and it leads to serious biased estimation for both regression parameters $\boldsymbol{\beta}$ and ρ in some setups. This observation is consistent with the existing results in the literature for studying right-censored data when treating ρ as unknown in the GORH models. The

main reason is that there is very little information about parameter ρ from the data. Note that current status data have much less information about the failure time than right-censored data.

A second strategy is to consider a set of candidate values of ρ and apply the proposed approach for each ρ, and then conduct model selection to choose the model with the best fit for the data using some model selection criteria, such as Akaike Information Criterion (AIC). This strategy is widely used for analyzing right-censored data when using the GORH models. The disadvantage is that one needs to implement the esitmation approach many times for all of the candidate values of ρ before finding the best model. Although it is workable and widely accepted, this strategy actually has two limitations: (1) choosing the candidate values of ρ is subjective; (2) the finite number of candidate values of ρ puts a restriction on the best model. Since the parameter space of ρ is the positive real line, the real best model may never be found using this strategy.

Motivated by these limitations and the fact that current status data have little information about ρ, we propose a new strategy of using a working model with $\rho = 1$ when true ρ is unknown. In this strategy, we only need to implement the proposed approach in Section 3.4 with $\rho = 1$ and then use the obtained parameter estimates from the working model to construct estimates of the quantities of interest under the true GORH model. Let ρ_1, $\boldsymbol{\beta}_1$, and Λ_{01} denote the parameter values in the true GORH model, and let $\boldsymbol{\beta}$ and Λ_0 denote the parameters under the working GORH model with $\rho = 1$. Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_0$ denote the maximum likelihood estimates of $\boldsymbol{\beta}$ and Λ_0 under the working GORH model, and $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_0$ can be obtained by applying the proposed approach in Section 3.4. Suppose that one wants to estimate the unknown baseline survival function $S_0(t) = P(T > t | \boldsymbol{x} = 0)$. Since this function takes the following two equivalent forms: $S_0(t) = \{1 + \rho_1 \Lambda_{01}(t)\}^{-\rho_1^{-1}}$ under the true GORH

model and $S_0(t) = \{1+\Lambda_0(t)\}^{-1}$ in the working model, the baseline survival function under the true model can be simply estimated by $\widehat{S}_0(t) = \{1 + \widehat{\Lambda}_0(t)\}^{-1}$ using the estimates from the working GORH model. In general, the survival function $S(t|\boldsymbol{x})$ with specific covariates $\boldsymbol{x}$ under the true GORH model can be approximated by $\widehat{S}(t|\boldsymbol{x}) = \{1 + \widehat{\Lambda}_0(t)\exp(\boldsymbol{x}'\widehat{\boldsymbol{\beta}})\}^{-1}$. These survival estimates allow one to predict the failure times for subjects in specific subgroups.

Note that this working model strategy does not allow one to retrieve an estimate of $\boldsymbol{\beta}_1$ under the true GORH model. However, it allows one to test $H_0 : \boldsymbol{\beta}_{1j} = 0$ vs. $H_a : \boldsymbol{\beta}_{1j} \neq 0$ under the true GORH model by testing $H_0 : \boldsymbol{\beta}_j = 0$ vs. $H_a : \boldsymbol{\beta}_j \neq 0$ under the working model using $\widehat{\boldsymbol{\beta}}_j$ since $\boldsymbol{\beta}_{1j}$ and $\boldsymbol{\beta}_j$ denote the effect of the jth covariate on the failure time under two different GORH models. The powers of these two tests should be close to each other for any covariate. This suggests that even though the working model strategy can not provide estimates of the covaraite effects under the true GORH model directly based on the estimates from the working model, it does provide answers to the important question that study investigators have: which covariates are significant risk factors?

To summarize, the working model strategy only needs to fit the working GORH model with $\rho = 1$ once but allows one to produce valid estimates of survival functions and to identify significant risk factors under other GORH models without the need of fitting those models. The choice of working model with $\rho = 1$ is made because of the simplicity in its survival function.

3.6 Simulation study

An extensive simulation study was conducted to assess the performance of the proposed approach under GORH model. For each subject i the failure time T_i was

generated from

$$F(t|\boldsymbol{x}_i) = 1 - \{1 + \rho\Lambda_0(t)\exp(x_{i1}\beta_1 + x_{i2}\beta_2)\}^{-\rho^{-1}},$$

where the baseline cumulative function $\Lambda_0(t) = \log(1+t) + t^2$, and the covariate vector $\boldsymbol{x}_i = (x_{i1}, x_{i2})'$ in which $x_{i1} \sim N(0, 0.5^2)$, $x_{i2} \sim$ Bernoulli(0.5), $i = 1, \ldots, n$. Both regression parameters β_1 and β_2 take on -1 and 1 and ρ takes the values of 0, 0.5, 1, 1.5 and 2. Note that $\rho = 0$ corresponds to a limiting case of the GORH models where the failure time T_i was generated from the PH model with cumulative distribution function $F(t|\boldsymbol{x}_i) = 1 - \exp\{-\Lambda_0(t)\exp(x_{i1}\beta_1 + x_{i2}\beta_2)\}$. The censoring time C_i was generated from a truncated exponential distribution $\mathcal{E}(1)$ with support $(0, 3)$, and the censoring indicator $\delta_i = I(T_i \leq C_i)$ was obtained by sampling from a Bernoulli distribution with success probability $F(C_i|\boldsymbol{x}_i)$. One advantage of this data generation mechanism is that it avoids generating the failure times T_i's, which are also unobserved in current status data. We generated 500 independent data sets, each with a sample size of $n = 200$, for each parameter configuration.

For each data set, we ran the proposed approach in Section 3.4 under the GORH model with the true value of ρ. In addition, we implemented the same method under the working GORH model with $\rho = 1$ for each simulated data set in order to evaluate the working model strategy in Section 3.5. Throughout the simulation study, we took 6 equally-spaced knots within the minimum and maximum of the observed censoring times for each data set and took degree to be 3 for adequate smoothness for the montone spline specification.

Scenario 1: when ρ is known

Table 3.1 presents the estimation results of the regression parameters from the proposed approach when using the true values of ρ. The summarized results include the

average of 500 point estimates (Point), the sample standard deviations of 500 point estimates (SSD), the average of the 500 estimated standard errors (ESE), and the 95% coverage probability for each regression parameter based on 500 data sets in each configuration. As seen in Table 3.1, all of the point estimates are close to their corresponding true values with small biases; all SSDs are closed to the corresponding ESEs, indicating that the variance estimates based on Louis's method are accurate; all CP95s are close to the nominal value 0.95, indicating that the asymptotic normality of the regression parameter estimates are valid. It is also observed that the estimated variances increases as ρ increases. This is expected because increasing ρ results in less information about the failure time in the observed data. This fact can be seen more clearly if one rewrites the GORH models as the gamma-frailty PH models in which ρ is the frailty variance.

Table 3.2 presents the global mean and maximum squared errors of the estimated survival functions from our approach under the true GORH models. Specifically, we consider a set of time points and define a local mean squared error (MSE) of the estimated survival function $\widehat{S}(t|\boldsymbol{x})$ at time t as

$$\text{MSE}\{\widehat{S}(t|\boldsymbol{x})\} = \frac{1}{500}\sum_{j=1}^{500}\{\widehat{S}^{(j)}(t|\boldsymbol{x}) - S(t|\boldsymbol{x})\}^2, \tag{3.7}$$

where $S(t|\boldsymbol{x})$ is the true survival function under the GORH model and is known, and $\widehat{S}^{(j)}(t|\boldsymbol{x})$ is the estimate of $S(t|\boldsymbol{x})$ from our approach for the jth data set, $j = 1, \ldots, 500$. The global mean (maximum) squared error of $\widehat{S}(t|\boldsymbol{x})$, denoted by meanMSE (maxMSE), is taken as the mean (maximum) of the local MSEs of $S(t|\boldsymbol{x})$ over the set of time points. The smaller these global MSEs are, the better estimation for the survival functions. Table 3.2 shows the global mean and maximum MSEs of the estimated survival functions with different covariate combinations, $\boldsymbol{x} = (0, 0)$, $(0, 1)$, and $(1, 0)$. From Table 3.2, all the global mean and maximum MSEs of survival estimates are very small for all parameter configurations, which suggests that

our method provides accurate estimates of the survival functions.

It is worth noting that the results in Tables 3.1 and 3.2 were obtained from our approach using 6 equally-spaced knots in the monotone spline specification. These results are expected to be improved if a procedure of optimal selection on the number of knots is performed.

Scenario 2: when ρ is unknown

Now we assume ρ is unknown and evaluate our working model strategy proposed in Section 3.5. The same sets of simulated data from the previous simulation were used again, and we implemented our approach under the misspecified working GORH model with $\rho = 1$ on all those data sets. Table 3.3 summarizes the global mean and maximum MSEs of the survival estiamtes under the misspecified working model. Let $\boldsymbol{\theta}^* = (\rho^*, \boldsymbol{\beta}^*, \Lambda_0^*)$ denote the true values of the unknown parameters in a GORH model, from which the failure times were generated. In calculating the local MSEs using (3.7), note that $\widehat{S}^{(j)}(t|\boldsymbol{x})$ and $S(t|\boldsymbol{x})$ have different forms here with

$$S(t|\boldsymbol{x}) = \{1 + \rho^* \Lambda_0^*(t) \exp(\boldsymbol{x}'\boldsymbol{\beta}^*)\}^{-\rho^{*-1}}$$

and

$$\widehat{S}^{(j)}(t|\boldsymbol{x}) = \{1 + \widehat{\Lambda}_0^{(j)}(t) \exp(\boldsymbol{x}'\widehat{\boldsymbol{\beta}}^{(j)})\}^{-1}, \quad j = 1, \ldots, 500,$$

where $\widehat{\boldsymbol{\beta}}^{(j)}$ and $\widehat{\Lambda}_0^{(j)}$ are the MLEs of $\boldsymbol{\theta}$ and Λ_0 under the working GORH model with $\rho = 1$ for the jth data set. As seen from Table 3.3, the global mean and maximum MSEs are small and are very comparable with the corresponding MSEs in Table 3.2 for all parameter configurations. This suggests that the working model strategy provides accurate survival estimates under the true GORH model by fitting the working model.

To assess the ability of identifying significant risk factors using the misspecified working model, we explored the power of the hypothesis testing of $H_0 : \beta_j = 0$ vs.

$H_a : \beta_j \neq 0$ for the jth covariate. For each data set, the proposed method were implemented under two GORH models: the true model and the working model. The null hypothesis was rejected if the resulting p-value based on the Wald test was smaller than the level of significance 0.05 for each data set under each model, and the power was calculated as the proportion of the data sets with rejected null hypotheses for each method since all the true β_j's are non-zeros in our setups. It is observed that the power of the test decreases as ρ increases under both models. This is not surprising because more uncertainty about the failure time is observed as ρ increases. Also, it is clear that from Table 3.4 using the working model leads to comparable powers of tesing the significance of all covariate effects as using the true GORH model. Thus, using the working GORH model lose little power in detecting significant covariates even though it is a misspecified model.

Table 3.1 The estimation results on the regression parameters from the proposed approach under the GORH models using the true values of ρ based on 500 replications. Point denotes the average of 500 point estimates, SSD the sample standard deviations of 500 point estimates, ESE the average of the 500 estimated standard errors, and CP95 the 95% coverage probability for each regression parameter in each parameter configuration.

			$\hat{\beta}_1$				$\hat{\beta}_2$			
ρ	β_1	β_2	Point	SSD	ESE	CP95	Point	SSD	ESE	CP95
0	-1	-1	-1.0725	0.3134	0.3003	0.952	-1.0632	0.2986	0.2875	0.946
	-1	1	-1.0905	0.3271	0.3121	0.938	1.0733	0.3171	0.2991	0.944
	1	-1	1.0783	0.3063	0.3018	0.966	-1.0496	0.3123	0.2908	0.936
	1	1	1.0850	0.3470	0.3110	0.932	1.0720	0.3170	0.2950	0.928
0.5	-1	-1	-1.0873	0.3783	0.3425	0.946	-1.0420	0.3341	0.3319	0.966
	-1	1	-1.0323	0.3592	0.3445	0.938	1.0536	0.3337	0.3279	0.948
	1	-1	1.0452	0.3669	0.3425	0.932	-1.0645	0.3503	0.3341	0.94
	1	1	1.0490	0.3598	0.3419	0.950	1.0464	0.3173	0.3295	0.960
1	-1	-1	-1.0687	0.3988	0.3936	0.944	-1.0522	0.3707	0.3829	0.940
	-1	1	-1.0208	0.3856	0.3814	0.954	1.0157	0.3765	0.3761	0.954
	1	-1	1.0193	0.3802	0.3845	0.970	-1.0125	0.3692	0.3758	0.950
	1	1	1.0304	0.4020	0.3835	0.950	1.0472	0.3814	0.3751	0.952
1.5	-1	-1	-1.0556	0.4401	0.4276	0.964	-1.0496	0.4125	0.4222	0.944
	-1	1	-1.0755	0.4173	0.4250	0.968	1.0571	0.4207	0.4139	0.950
	1	-1	1.0283	0.4435	0.4254	0.946	-1.0422	0.4256	0.4222	0.940
	1	1	1.0347	0.4267	0.4226	0.954	1.0365	0.4116	0.4157	0.942
2	-1	-1	-1.0566	0.4792	0.4798	0.952	-1.0248	0.4492	0.4673	0.946
	-1	1	-1.0481	0.5051	0.4640	0.946	1.0066	0.4719	0.4532	0.936
	1	-1	1.0798	0.4744	0.4712	0.952	-1.0425	0.4930	0.4642	0.940
	1	1	1.0849	0.4654	0.4653	0.954	1.0500	0.4419	0.4534	0.956

Table 3.2 The global mean and maximum MSEs ($\times 10^{-3}$)of the estimates $\widehat{S}_{ij}$ of the survival function S_{ij} from the proposed method using the true value of ρ for each parameter configuration. The three (i, j) combinations $(0, 0)$, $(0, 1)$ and $(1, 0)$ correspond to three different covariate combinations $(x_1, x_2) = (0, 0)$, $(0, 1)$, and $(1, 0)$, respectively.

			meanMSE			maxMSE		
ρ	β_1	β_2	S_{00}	S_{01}	S_{10}	S_{00}	S_{01}	S_{10}
0	-1	-1	2.3	6.0	10.4	6.3	11.8	18.2
		1	2.4	1.3	13.6	6.4	8.1	33.5
	1	-1	2.5	6.6	2.9	6.8	12.3	18.5
		1	2.6	1.2	2.9	7.5	7.4	17.8
0.5	-1	-1	3.4	8.6	13.7	5.8	15.6	19.5
		1	4.1	1.7	17.1	6.7	7.2	29.8
	1	-1	3.7	8.7	3.9	6.5	16.2	17.3
		1	4.0	1.7	3.6	6.3	6.7	13.8
1	-1	-1	4.7	10.0	14.7	5.6	24.7	27.2
		1	5.8	2.7	18.0	7.5	6.2	35.1
	1	-1	4.9	9.4	4.9	5.8	23.0	13.6
		1	6.1	2.6	4.9	7.9	6.3	12.7
1.5	-1	-1	5.9	10.3	15.0	9.7	27.6	32.2
		1	7.1	3.6	16.6	13.1	6.7	39.2
	1	-1	5.8	10.5	6.3	9.4	27.8	13.5
		1	6.7	3.6	6.3	12.9	5.8	13.2
2	-1	-1	6.3	9.7	14.2	12.8	29.7	33.6
		1	7.9	4.7	17.9	17.6	7.9	42.7
	1	-1	6.6	10.6	7.3	13.9	31.8	12.4
		1	7.9	4.4	7.3	17.2	6.8	10.7

Table 3.3 The global mean and maximum MSEs ($\times 10^{-3}$) of the estimates $\widehat{S}_{ij}$ of the survival function S_{ij} from the proposed method under the working GORH model with $\rho = 1$ for each parameter configuration. The three (i, j) combinations $(0, 0)$, $(0, 1)$ and $(1, 0)$ correspond to three different covariate combinations $(x_1, x_2) = (0, 0)$, $(0, 1)$, and $(1, 0)$, respectively.

			meanMSE			maxMSE		
ρ	β_1	β_2	S_{00}	S_{01}	S_{10}	S_{00}	S_{01}	S_{10}
0	-1	-1	2.9	5.5	9.4	6.9	10.2	18.2
		1	2.4	1.6	13.0	7.7	10.9	34.4
	1	-1	3.1	6.2	3.5	7.6	11.2	22.7
		1	2.5	1.6	3.4	8.7	10.8	22.7
0.5	-1	-1	3.4	8.2	12.6	5.8	13.2	16.4
		1	3.8	1.8	15.8	6.7	8.2	25.2
	1	-1	3.7	8.2	4.2	6.3	13.7	18.7
		1	3.7	1.8	3.7	7.0	7.1	15.4
1	-1	-1	4.7	10.0	14.7	5.6	24.7	27.2
		1	5.8	2.7	18.0	7.5	6.2	35.1
	1	-1	4.9	9.4	4.9	5.8	23.0	13.6
		1	6.1	2.6	4.9	7.9	6.3	12.7
1.5	-1	-1	6.2	10.6	15.6	10.6	29.7	34.8
		1	7.6	3.8	17.4	14.9	6.4	44.6
	1	-1	6.2	10.9	6.7	10.5	30.1	14.0
		1	7.2	3.8	6.6	14.5	5.6	12.7
2	-1	-1	7.2	10.5	15.4	15.8	34.9	39.6
		1	8.9	5.3	18.9	21.7	8.7	50.8
	1	-1	7.6	11.4	8.9	17.5	37.5	13.6
		1	8.8	5.0	8.3	20.9	8.5	10.9

Table 3.4 The estimated powers for testing $H_0 : \beta_j = 0$ vs. $H_a : \beta_j \neq 0$ from the proposed method using true ρ (True model) and using $\rho = 1$ (Working model). The power is calculated as the proportion of rejected null hypotheses based on the Wald tests for the 500 simulated data sets in each parameter configuration.

ρ	β_1	β_2	True model		Working model	
0	-1	-1	0.964	0.976	0.946	0.964
		1	0.968	0.980	0.952	0.964
	1	-1	0.980	0.960	0.966	0.936
		1	0.952	0.978	0.938	0.954
0.5	-1	-1	0.896	0.892	0.882	0.888
		1	0.878	0.920	0.882	0.908
	1	-1	0.878	0.894	0.874	0.890
		1	0.876	0.914	0.874	0.906
1.5	-1	-1	0.720	0.724	0.722	0.704
		1	0.736	0.714	0.732	0.718
	1	-1	0.682	0.710	0.688	0.704
		1	0.672	0.722	0.662	0.718
2	-1	-1	0.610	0.594	0.610	0.586
		1	0.642	0.598	0.632	0.598
	1	-1	0.634	0.614	0.642	0.604
		1	0.644	0.666	0.654	0.648

3.7 An illustrative example

Sponsord by the National Cancer Institute, the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial is a multicenter, randomized two-arm trial designed to evaluate the effect of routine screening for PLCO cancers on disease-specific mortality. The enrollment of the PLCO study started in 1993 and ended in 2001, and eligible participants were aged 55-74 and did not have personal history of any PLCO cancer. At the enrollment, participants reported their personal characteristics and they were randomized into two arms: intervention arm and control arm based on their age, sex, and center. Participants in the intervention arm received annual screenings but participants in the control arm did not.

Our data set comes from the baseline data at the enrollment from the prostate component of the PLCO study. The response variable of interest is the onset time of diabetes, which is either left-censored or right-censored at the age of enrollment depending on whether the participant had diabetes at enrollment. The considered covariates include education level (binary with 1 for not having a college education), hypertension (binary with 1 for yes), race (binary with 1 for others), aspirin (binary with 1 for regular use), ibuprofen (binary with 1 for regular use), smoke (binary with 1 for yes), obesity (binary with 1 for obese), and group (binary with 1 for intervention arm). After deleting the subjects with missing information in the response, age at enrollment, or these covariates, there are 43,395 participants in our data set. The detailed summary results about these variables can be found in Table 3.5. Among these subjects, 4,129 (9.5%) have left-censored diabetes onset times and 39,400 (90.5%) have right-censored diabetes onset times.

We applied our proposed approach to this data set with fixed $\rho = 0$, 1, and 2. For each case, we specified the order of the monotone splines to be 3 and used $m - 2$ equally spaced interior knots within $(49, 78)$ formed by thel observed minimum and maximum ages at enrollment, with m taking 5, 7, and 9. The initial values were taken to be 0.5 for all regression parameters and spline coefficients, and the convergence was claimed when the maximum of changes in all parameters values was smaller than 10^{-5}. It took about 2.5 minutes to run our approach on average for each combination of m and ρ.

Table 3.6 presents the estimated covariate effects from the proposed method for using different values of ρ and m. As seen in Table 3.6, the estimated covariate effects and their estimated standard errors are essentially identical across different values of m for any fixed ρ, which indicates that this analysis is robust to the number of interior knots using our method. The AIC crieria shows that the method using 3

Table 3.5 Summary information about the response and the covariates used in the analysis of the PLCO data.

Variable	Description	Code	Number	Percentage
Diabetes	Yes	1	4129	10%
	No	0	39400	90%
Age	Mean (SD)=62.7(5.3)			
	Max(Min)=78(49)			
Education	$\geq$ College	0	24856	57%
	$<$ High school	1	18673	43%
Hypertension	Yes	1	15341	35%
	No	0	28188	65%
Race	White, non-Hispanic	0	38507	89%
	Others	1	5022	11%
Aspirin	Yes	1	24513	56%
	No	0	19016	44%
Ibuprofen	Yes	1	11547	27%
	No	0	31982	73%
Smoke	Yes	1	7857	18%
	No	0	35672	82%
Obesity	Yes	1	10766	25%
	No	0	32763	75%
Group	Intervention	1	21108	49%
	Control	0	22421	51%

interior knots provides the best of fit for each ρ, and the model with $\rho = 2$ yields the smallest AIC value than those with other considered values of ρ when taking $m = 5$. More implementations of the proposed approach with different values of ρ and m will have to be run if one wishes to find the model that provides the best fit based on the AIC criteria. However, doing so is time consuming and is not needed based on our working model strategy, and the reasons are illustrated below.

Table 3.7 provides the p-values associated with the hypothesis tests on whether a covariate has a significant effect on diabetes from the proposed method with different ρ when taking 5 interior knots. As seen in Table 3.7, the p-values associated with

testing any covaraite effect are comparable for different ρ's, which indicates using the working model with $\rho = 1$ leads to the same conclusions as using other GORH models in detecting sigificant covariates. It is clear from Table 3.7 that education, hypertension, race, aspirin, smoke, and obesity all have a significant effect on diabetes, while ibuprofen and group do not have a significant effect. It is not surprising that group does not have a significant effect on diabetes since we are using the baseline data at randomization.

Figure 3.1 plots the estimated baseline survival functions for non-obese and obese participants controlling all other covariates to be 0 based on our method with different ρ's. The estimated survival functions show a clear difference between the two weight groups, which is consistent with the results in Table 3.7. It is also clear that the estimated survival functions obtained when using different ρ's are very close for each weight group, indicating using the working model with $\rho = 1$ yields accurate survival estimates as other GORH models.

3.8 Discussions

In this chapter, we investigate the identifiability issues of the GORH models and prove that all the parameters are identifiable in the usual regression settings. However, treating ρ as an unknown parameter and estimating it with other parameters may lead to serious problems in practice when analyzing right-censored data from existing literature research and when analyzing current status data based on our own experience. The main reason is that data have too little information about ρ and the observed likelihood as a function of ρ is quite flat. Note that this is different from modeling multivariate or clustered failure times under the gamma frailty PH model with the GORH model as the marginal distributions, where the dependence among the failure times does provide adequate information for estimating ρ.

Table 3.6 The estimated covariate effects and their corresponding standard errors from the proposed approach using different values of ρ and numbers of knots m in the analysis of the PLCO data. The Akaike Information Criterion (AIC) values from these models are also presented.

Model	Covariates	$m = 5$	$m = 7$	$m = 9$
$\rho = 0$	Education	0.139 (0.032)	0.150 (0.031)	0.150 (0.031)
	Hypertension	0.748 (0.032)	0.748 (0.032)	0.748 (0.032)
	Race	0.671 (0.040)	0.671 (0.040)	0.671 (0.040)
	Aspirin	0.209 (0.033)	0.209 (0.033)	0.209 (0.033)
	Ibuprofen	0.037 (0.035)	0.037 (0.035)	0.037 (0.035)
	Smoke	-0.166 (0.045)	-0.166 (0.046)	-0.166 (0.046)
	Obesity	0.690 (0.033)	0.690 (0.033)	0.690 (0.033)
	Group	-0.017 (0.031)	-0.017 (0.031)	-0.017 (0.031)
	AIC	25546.29	25549.83	25553.62
$\rho = 1$	Education	0.154 (0.034)	0.154 (0.034)	0.154 (0.034)
	Hypertension	0.800 (0.034)	0.800 (0.034)	0.800 (0.034)
	Race	0.734 (0.044)	0.735 (0.044)	0.735 (0.044)
	Aspirin	0.224 (0.035)	0.224 (0.035)	0.224 (0.035)
	Ibuprofen	0.045 (0.038)	0.045 (0.038)	0.045 (0.038)
	Smoke	-0.175 (0.048)	-0.176 (0.048)	-0.176 (0.048)
	Obesity	0.746 (0.036)	0.745 (0.036)	0.745 (0.036)
	Group	-0.017 (0.034)	-0.017 (0.034)	-0.017 (0.034)
	AIC	25538.03	25541.64	25545.47
$\rho = 2$	Education	0.168 (0.036)	0.168 (0.036)	0.168 (0.036)
	Hypertension	0.853 (0.036)	0.853 (0.036)	0.853 (0.036)
	Race	0.795 (0.048)	0.795 (0.048)	0.795 (0.048)
	Aspirin	0.239 (0.038)	0.238 (0.038)	0.238 (0.038)
	Ibuprofen	0.052 (0.041)	0.052 (0.041)	0.052 (0.041)
	Smoke	-0.184 (0.051)	-0.185 (0.051)	-0.185 (0.051)
	Obesity	0.800 (0.039)	0.800 (0.039)	0.800 (0.039)
	Group	-0.018 (0.036)	-0.018 (0.036)	-0.018 (0.036)
	AIC	25532.32	25536.00	25539.85

Assuming ρ is known, we propose a new estimation approach for analyzing current status data under the GORH models. Our approach is an generalization of McMahan et al. (2013) under the PH model. Specifically our proposed method adopts monotone splines of Ramsay (1988) to approximate the unknown nondecreasing function,

Table 3.7 The p-values associated with testing the significance of each covariate effect based on the Wald tests from the proposed method in the PLCO data analysis. The proposed method was implemented under different GORH models with $\rho = 0$, 1, and 2 using order 3 and 3 equally spaced interior knots for the monotone spline specifications.

	p-values		
Covariates	$\rho = 0$	$\rho = 1$	$\rho = 2$
Education	< 0.001	< 0.001	< 0.001
Hypertension	< 0.001	< 0.001	< 0.001
Race	< 0.001	< 0.001	< 0.001
Aspirin	< 0.001	< 0.001	< 0.001
Ibuprofen	0.290	0.240	0.203
Smoke	< 0.001	< 0.001	< 0.001
Obesity	< 0.001	< 0.001	< 0.001
Group	0.595	0.613	0.626

reducing the number of unknown parameters to finite, and an efficient EM algorithm is derived based on a three-stage data augmentation. Our approach enjoys several appealing properties, such as being easy to implement and robust to initial values, converging fast, and providing variance estimates in explicit form. Simulation study suggests that our approach performs well in estimating all the regression parameters and survival functions.

Treating ρ as known restricts the flexibility of the GORH model. To tackle this problem when ρ is unknown, we investigate a working model strategy which only requires one to implement our approach under the GORH models with $\rho = 1$. The results in our simulation study and real data application suggest that the use of the working model with $\rho = 1$ does not lose accuracy in estimating survival functions or power in detecting significant covariates even when it is a misspecified model. This suggests that the use of our method under the working GORH model with $\rho = 1$ does not actually restrict the great flexibility owned by the GORH models.

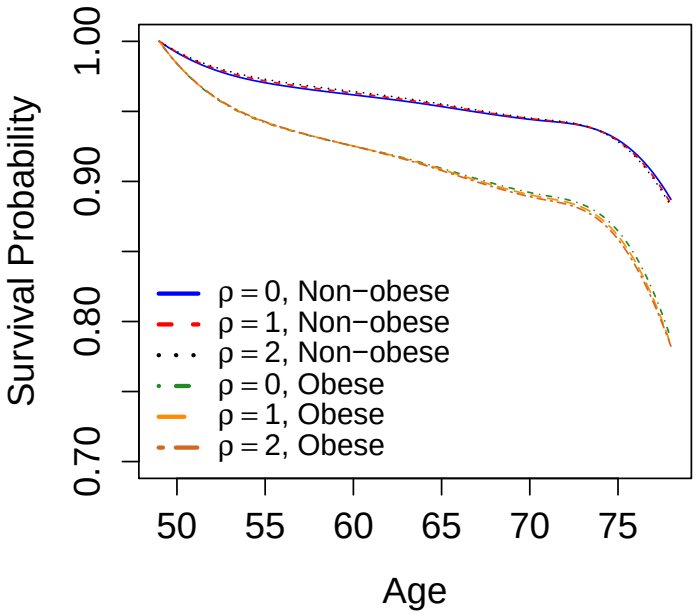

Figure 3.1 The estimated survival curves for different weight groups (non-obese and obese) controlling all other covariates at the baseline levels from the proposed method under different GORH models with $\rho = 0$, 1, and 2.

Chapter 4

Joint modeling of panel count data and interval-censored data with application to sexually transmitted infections

4.1 Introduction

Human sexual contacts are the primary path for spreading sexually transmitted infection (STI) diseases such as *Chlamydia trachomatis* (CT), *Neisseria gonorrhoeae* (NG), and *Trichomonas vaginalis* (TV). The proportion of young people who are aged 15-24 of getting infected is nearly half of the new sexually transmitted infections population in the United States (Tu et al., 2009), therefore it is meaningful to understand the young people sexual behaviors for preventing the infections. Young Women's Project (YWP) is an epidemiological study that was designed for studying sexual behaviors of young women. In the YWP, young women between 14 and 17 years old were recruited from three urban primary care clinics from the inner city of Indianapolis, Indiana. All the participants were required to visit the clinics every 3 months for up to over 6 years. At each visit the study subjects were queried about their sexual behavior such as the occurrence of sexual intercourse, condom use, condom failures, etc in the past 3 months. Besides, at each visit they were tested for the presence of CT, NG, and TV. If there was any of STI found, they would be treated promptly. In addition to the quarterly clinic visits, the study subjects also completed daily behavioral diaries

to record the subject's daily sexual activities to supplement the diagnostic test results obtained at pre-determined clinical visits. Intuitively, STI behavioral diaries contain important information about the potential times of infection. In the YMP study, the participated young women were examined for the presence of infections only at their follow-up visits. Thus, the true infection time is not exactly observed but is only known to fall within some interval that is formed by the last visit time without infection and the first visit time with infection. Hence the time to get first infection of any STI has the interval-censored data structure. In our analysis, we are interested in the time to the first infection of any STI since the enrollment under consideration. Harezlak and Tu (2006) proposed a multiple imputation method to resample true infection times of STI and estimated survival functions using auxiliary behavioral information provided by daily diaries. Condom non-use has long been recognized as a risk factor for STI acquisition. Hua et al. (2014) treated the number of condom non-use as panel count response and proposed a gamma frailty non-homogeneous Poisson process model for analyzing such data set. A lot of research work has been done about this study, for example, Tu et al. (2009) studied the time between first intercourse and first sexually transmitted infection with CT, NG and TV and time between repeated infections. Li et al. (2015) proposed a model-based sexually transmitted infection screening algorithm to identify individuals who are at increased STI risk. For more details about this study, please refer to Yu et al. (2009), Ghosh and Tu (2009) and Ott et al. (2011). Our research interest is to study the number of condom non-use over time and the time to a new STI as two correlated responses and to investigate the effects of the available covariates on these two types of responses. It is meaningful to study the two responses jointly because they share the same observational process.

Joint modeling of longitudinal and survival data has been popular in the past two decades. Tsiatis and Davidian (2004) reviewed the development of joint modeling of

longitudinal and time-to-event data. Wu et al. (2012) gave a brief overview of joint models for longitudinal and survival data and commonly used methods. In addition to joint modeling of longitudinal and survival data, Huang and Wang (2004) proposed a joint model of recurrent event processes and failure time. Cowling et al. (2006) analyzed event counts and survival times jointly using a maximum likelihood approach. Huang and Liu (2007) introduced a joint model for gap times and survival times fitted by an EM algorithm. Lee et al. (2011) performed a joint analysis of longitudinal and interval-censored failure time data using imputation method. However, there is no work about the joint modeling of panel count data and interval-censored data to the best of our knowledge.

In this chapter, we borrowed the idea of using non-homogeneous Poisson process model for panel count data (Yao et al., 2016) and the idea of using PH model for analyzing interval-censored data (Wang et al., 2015) to build a joint model to analyze the two types of data simultaneously. It is reasonable to study the two types of data jointly because the panel count response and failure time share the same observation process. The non-homogeneous Poisson process model and the PH model can be extended and connected via a shared gamma frailty. It is assumed that the panel count response and the failure time are conditionally independent given the frailty and covariates. Such modeling is very intuitive and allows one to construct the observed likelihood easily. A maximum likelihood approach is proposed for analyzing panel count data and interval-censored data under the proposed joint model. Monotone splines of Ramsay (1988) is adopted to model the baseline mean function and the baseline conditional cumulative hazard function, and all the parameters are estimated jointly through an efficient EM algorithm. The proposed joint analysis is more efficient than the corresponding univariate analysis of panel count data and interval-censored data separately. The parameter variances are estimated by using Louis's method

and numerical approximation of observed information matrix. Finally, the proposed model and approach are applied to the STI data set.

The remainder of this chapter is organized as follows. The methodological details of the proposed model and technique are provided in Section 4.2. These details include the use of monotone splines for approximating the baseline mean function and the baseline conditional cumulative hazard function, the EM algorithm derivation, and parameter variance estimation method. The performance of the proposed approach is illustrated in Section 4.3 through an extensive simulation study. In Section 4.4 the proposed method is used to analyze data from sexually transmitted infections study. Section 4.5 provides a summary discussion.

4.2 Proposed method

Data, model, and observed likelihood

Consider n subjects are involved in the study. Let $N_i(t)$ denote the recurrent event process that results in panel counts for subject i where $i = 1, \ldots, n$. The exact times of the recurrent events would be known if subject i is under continuous monitoring. However, $N_i(t)$ is only observed at discrete times $\{t_{ij}, j = 1, \ldots, K_i\}$. Thus, only the number of the recurrent events between adjacent observations times are available, leading to panel count data. Let $\boldsymbol{x}_i$ denote a vector of $p \times 1$ time-independent covariates associated with panel count data and $\{t_{ij}, j = 1, \ldots, K_i\}$ denote the actual observation times for subject i, where K_i is the number of observations and t_{iK_i} is the last observation time. This counting process $\{N_i(t)\}$ is observed only at t_{ij}'s. We assume that the counting process is conditionally independent of the observational process given the covariates. Conditional on latent variable ϕ_i, $N_i(t)$ is a non-homogeneous Poisson process with mean function $\mu_0(t)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i$, where $\mu_0(t)$

is an unspecified nondecreasing baseline mean function with $\mu_0(0) = 0$. This model implies

$$N_i(t)|\phi_i \sim \mathcal{P}\{\mu_0(t)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i\},$$

for any $t > 0$, where $\mathcal{P}(a)$ denotes the Poisson distribution with mean a, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression parameters.

For the same subject i, let T_i denote the failure time of interest. Let $\tilde{\boldsymbol{x}}_i$ denote a $q \times 1$ vector of time-independent covariates associated with the time-to-event process for subject i. We assume that conditional on the covariates, the failure time is independent of the observational process. Conditional on the same frailty ϕ_i used in the non-homogeneous Poisson process model, the conditional hazard function is specified as

$$\lambda(t|\tilde{\boldsymbol{x}}_i, \phi_i) = \phi_i\lambda_0(t)\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha}),$$

where $\lambda_0(t)$ is the unspecified and non-negative conditional baseline hazard function and $\boldsymbol{\alpha}$ is a vector of $q \times 1$ regression coefficients. The ϕ_i's are assumed to be independently and identically distributed from $\mathcal{G}a(\nu, \nu)$ with mean 1 and variance ν^{-1}. The mean constraint made on gamma frailty is for avoiding the non-identifiability issue because $\mu_0(t)$ and $\Lambda_0(t)$ are both unspecified. In addition, we assume that conditional on frailty and covariates, the panel count response is independent of the failure time. Under the gamma frailty PH model the conditional CDF of the failure time is given by $F(t|\tilde{\boldsymbol{x}}_i, \phi_i) = 1 - \exp\{-\Lambda_0(t)\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i\}$, where $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ is the conditional baseline cumulative hazard function. The resulting marginal model for the failure time T is referred to as the generalized odds-rate hazards (GORH) models in the literature with the following form for its survival function,

$$S(t|\boldsymbol{x}) = \{1 + \nu^{-1}\Lambda_0(t)\exp(\tilde{\boldsymbol{x}}_i\boldsymbol{\alpha})\}^{-\nu}.$$

Zhou et al. (2016) studied the GORH model for interval-censored data. The observed data is denoted by $\mathcal{D} = \{(t_{ij}, (L_i, R_i], N_i(t_{ij}), \boldsymbol{x}_i, \tilde{\boldsymbol{x}}_i), j = 1, \ldots, K_i, i = 1, \ldots, n\}$, where L_i and R_i denote the left and right bounds of the observed interval for the ith subject, respectively. Note that $L_i = 0 (R_i = \infty)$ indicates that the ith subject's failure time is left- (right-) censored. Define $Z_{ij} = N_i(t_{ij}) - N_i(t_{ij-1})$ which is the number of events occurred in the time interval $(t_{ij-1}, t_{ij}]$ for all i's and j's. By the independent increment property of Poisson model $Z_{ij}|\phi_i \sim \mathcal{P}[\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i]$. Under the conditional independence assumptions, the observed likelihood contributed by subject i is given by

$$\int_0^\infty P(L_i < T_i < R_i|\phi_i) \left\{ \prod_{j=1}^{K_i} P(Z_{ij}|\phi_i) \right\} g(\phi_i) d\phi_i,$$

where

$$\begin{aligned}
P(L_i < T_i < R_i|\phi_i) &\propto F(R_i|\tilde{\boldsymbol{x}}_i, \phi_i)^{\delta_{i1}} \{F(R_i|\tilde{\boldsymbol{x}}_i, \phi_i) - F(L_i|\tilde{\boldsymbol{x}}_i, \phi_i)\}^{\delta_{i2}} \\
&\quad \times \{1 - F(L_i|\tilde{\boldsymbol{x}}_i, \phi_i)\}^{\delta_{i3}}, \\
P(Z_{ij}|\phi_i) &= \exp[-\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i] \\
&\quad \times \frac{[\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i]^{Z_{ij}}}{Z_{ij}!}, \\
g(\phi_i) &= \frac{\nu^\nu}{\Gamma(\nu)} \phi_i^{\nu-1} \exp(-\nu\phi_i).
\end{aligned}$$

Note that δ_{i1}, δ_{i2} and δ_{i3} are indicator functions that distinguish the left censoring, the interval censoring and the right censoring for the status of failure time T_i, respectively. We have $\delta_{i1} + \delta_{i2} + \delta_{i3} = 1$. The observed likelihood L_{obs} for n subjects can be obtained

in the following closed form by integrating out the latent variable ϕ_i,

$$\begin{aligned} L_{obs} &= \prod_{i=1}^{n} \int_0^{\infty} P(L_i < T_i < R_i|\phi_i) \prod_{j=1}^{K_i} P(Z_{ij}|\phi_i) g(\phi_i) d\phi_i \\ &= \prod_{i=1}^{n} \frac{\nu^{\nu}\Gamma(\nu + Z_{i\cdot})}{\Gamma(\nu)} \times \prod_{j=1}^{K_i} \frac{[\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\}\exp(\boldsymbol{x}_i'\boldsymbol{\beta})]^{Z_{ij}}}{Z_{ij}!} \\ &\quad \times \left[\frac{\delta_{i1}}{\{\nu + \mu_0(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^{\nu+Z_{i\cdot}}} \right. \\ &\quad + \frac{\delta_{i2} + \delta_{i3}}{\{\nu + \Lambda_0(L_i)\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha}) + \mu_0(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^{\nu+Z_{i\cdot}}} \\ &\quad \left. - \frac{\delta_{i1} + \delta_{i2}}{\{\nu + \Lambda_0(R_i)\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha}) + \mu_0(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^{\nu+Z_{i\cdot}}} \right], \end{aligned}$$

where $Z_{i\cdot} = \sum_{j=1}^{K_i} Z_{ij}$ is the cumulative panel counts for subject i. The unknown parameters in the observed likelihood include the regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, frailty variance parameter ν, the baseline mean function $\mu_0(\cdot)$, and the baseline conditional cumulative hazard function $\Lambda_0(\cdot)$.

Modeling $\mu(\cdot)$ and $\Lambda(\cdot)$ with monotone splines

Both the baseline mean function $\mu_0(t)$ and the baseline conditional cumulative hazard function $\Lambda_0(t)$ are unspecified nondecreasing functions in this project. Following the previous chapters, monotone splines (Ramsay, 1988) is adopted to model them. Specifically, $\mu_0(t)$ can be expressed in the following form,

$$\mu_0(t) = \sum_{l=1}^{L} \gamma_l b_l(t),$$

and $\Lambda_0(t)$ can be written in the following manner,

$$\Lambda_0(t) = \sum_{m=1}^{M} \tilde{\gamma}_m I_m(t),$$

where $b_l(t)$'s and $I_m(t)$'s are integrated basis functions which can take values between 0 and 1, and γ_l's and $\tilde{\gamma}_m$'s are non-negative spline coefficients. To determine the

integrated basis functions, one needs to specify the knots and the degree. Knots are a sequence of increasing time points. The placement of knots determines the overall modeling flexibility, with more knots in a region for greater modeling flexibility in that region. The degree controls the overall smoothness of the basis functions. Degree equal to 1, 2 and 3 corresponds to linear, quadratic, and cubic basis functions, respectively. Once the knots and degree are specified, the basis functions are determined. The number of basis functions is the number of interior knots plus the degree.

Data augmentation for EM algorithm

The EM algorithm is adopted to find the maximum likelihood estimate of the unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \boldsymbol{\gamma}', \tilde{\boldsymbol{\gamma}}', \nu)'$ where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_L)'$ and $\tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_M)'$. The EM algorithm can be constructed based on introducing some latent variables to supplement the observed data. The discussion of two-stage data augmentation is formulated based on the two proposed models, non-homogeneous gamma frailty Poisson process model and gamma frailty PH model. This idea has been proposed in (Lin et al., 2015).

Consider a gamma frailty non-homogeneous Poisson process $\tilde{N}(t)$ with cumulative intensity function $\Lambda_0(t)\exp(\tilde{\boldsymbol{x}}'\boldsymbol{\alpha})\phi$. The process $\tilde{N}(t)$ is constructed for monitoring the occurrence of failure time T which can be considered as the time of the first jump in the process, i.e., $T = \inf\{t : \tilde{N}(t) > 0|\phi\}$. To show that T follows the gamma frailty PH model with a cumulative distribution function $F(t|\tilde{\boldsymbol{x}}, \phi) = 1 - \exp\{-\Lambda_0(t)\exp(\tilde{\boldsymbol{x}}'\boldsymbol{\alpha})\phi\}$, one should note that $P(T > t|\phi) = P\{\tilde{N}(t) = 0|\phi\} = \exp\{-\Lambda_0(t)\exp(\tilde{\boldsymbol{x}}'\boldsymbol{\alpha})\phi\} = 1 - F(t|\tilde{\boldsymbol{x}}, \phi)$, for any t, because $\tilde{N}(t)$ is a Poisson random variable with mean parameter $\Lambda_0(t)\exp(\tilde{\boldsymbol{x}}'\boldsymbol{\alpha})\phi$ conditional on ϕ.

Let $\tilde{N}_i(t)$ denote the latent Poisson process for subject i, which has conditional

cumulative intensity function $\Lambda_0(t)\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i$, for $i = 1, \ldots, n$. For subject i define $W_{i1} = \tilde{N}_i(\tilde{t}_{i1})$, where $\tilde{t}_{i1} = R_i 1_{(\delta_{i1}=1)} + L_i 1_{(\delta_{i1}=0)}$. When $\delta_{i1} = 0$ define $W_{i2} = \tilde{N}_i(\tilde{t}_{i2}) - \tilde{N}_i(\tilde{t}_{i1})$, where $\tilde{t}_{i2} = R_i 1_{(\delta_{i2}=1)} + L_i 1_{(\delta_{i3}=1)}$. W_{i1} and W_{i2} are Poisson random variables with mean parameters $\Lambda_0(\tilde{t}_{i1})\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i$ and $\{\Lambda_0(\tilde{t}_{i2}) - \Lambda_0(\tilde{t}_{i1})\}\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i$, respectively. Note that W_{i1} and W_{i2} are independent conditional on ϕ_i when $\delta_{i1} = 0$.

Under this construction if T_i is left-censored then $L_i = 0$ and $\tilde{t}_{i1} = R_i$, so $P(T_i \leq \tilde{t}_{i1}|\phi_i) = P\{\tilde{N}_i(\tilde{t}_{i1}) > 0|\phi_i\} = P(W_{i1} > 0|\phi_i) = 1 - P(W_{i1} = 0|\phi_i) = 1 - \exp\{-\Lambda_0(R_i)\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i\} = F(R_i|\tilde{\boldsymbol{x}}_i, \phi_i)$. If T_i is interval-censored, then $\tilde{t}_{i1} = L_i$ and $\tilde{t}_{i2} = R_i$,

$$
\begin{aligned}
P(\tilde{t}_{i1} < T_i < \tilde{t}_{i2}|\phi_i) &= P\{\tilde{N}_i(\tilde{t}_{i1}) = 0, \tilde{N}_i(\tilde{t}_{i2}) > 0|\phi_i\} \\
&= P(W_{i1} = 0, W_{i2} > 0|\phi_i) \\
&= \exp\{-\Lambda_0(\tilde{t}_{i1})\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i\} \\
&\quad \times[1 - \exp\{-\{\Lambda_0(\tilde{t}_{i2}) - \Lambda_0(\tilde{t}_{i1})\}\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i\}] \\
&= \exp\{-\Lambda_0(\tilde{t}_{i1})\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i\} - \exp\{-\Lambda_0(\tilde{t}_{i2})\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i\} \\
&= \exp\{-\Lambda_0(L_i)\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i\} - \exp\{-\Lambda_0(R_i)\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i\} \\
&= F(R_i|\tilde{\boldsymbol{x}}_i, \phi_i) - F(L_i|\tilde{\boldsymbol{x}}_i, \phi_i).
\end{aligned}
$$

If T_i is right-censored, then $\tilde{t}_{i1} = L_i$ and $\tilde{t}_{i2} = L_i$,

$$
\begin{aligned}
P(T_i > \tilde{t}_{i2}|\phi_i) &= P\{\tilde{N}_i(\tilde{t}_{i2}) = 0|\phi_i\} \\
&= P(W_{i1} = 0, W_{i2} = 0|\phi_i) \\
&= \exp\{-\Lambda_0(\tilde{t}_{i1})\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i\}\exp\{-\{\Lambda_0(\tilde{t}_{i2}) - \Lambda_0(\tilde{t}_{i1})\}\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i\} \\
&= \exp\{-\Lambda_0(\tilde{t}_{i2})\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i\} \\
&= \exp\{-\Lambda_0(L_i)\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i\} \\
&= 1 - F(L_i|\tilde{\boldsymbol{x}}_i, \phi_i).
\end{aligned}
$$

Based on the latent variables W_{i1}'s, W_{i2}'s and ϕ_i's, the augmented likelihood can be expressed as

$$L_{aug}(\boldsymbol{\theta}) = \prod_{i=1}^{n} g(\phi_i)\mathcal{P}_{W_{i1}}(W_{i1}|\phi_i)\mathcal{P}_{W_{i2}}(W_{i2}|\phi_i)^{\delta_{i2}+\delta_{i3}}\{\delta_{i1}1_{(W_{i1}>0)} + \delta_{i2}1_{(W_{i1}=0,W_{i2}>0)} + \delta_{i3}1_{(W_{i1}=0,W_{i2}=0)}\} \times \prod_{j=1}^{K_i}\mathcal{P}_{Z_{ij}}(Z_{ij}|\phi_i),$$

where $\mathcal{P}_A(\cdot)$ denotes the Poisson probability mass function for random variable A. One can obtain the observed likelihood by integrating the W_{i1}'s, W_{i2}'s and ϕ_i's out of the above augmented likelihood $L_{aug}(\boldsymbol{\theta})$.

To take advantage of monotone spline representation of $\mu_0(t)$ and $\Lambda_0(t)$, we consider a second stage of data augmentation. For each subject i, each of W_{i1}, W_{i2} and Z_{ij} can be decomposed as a sum of conditionally independent Poisson random variables, $W_{i1} = \sum_{m=1}^{M} W_{im1}, W_{i2} = \sum_{m=1}^{M} W_{im2}$ and $Z_{ij} = \sum_{l=1}^{L} Z_{ijl}$, where W_{im1}, W_{im2} and Z_{ijl} are Poisson random variables having mean parameters $\tilde{\gamma}_m I_m(\tilde{t}_{i1})\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i$, $\tilde{\gamma}_m\{I_m(\tilde{t}_{i2}) - I_m(\tilde{t}_{i1})\}\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\phi_i$ and $\gamma_l\{b_l(t_{ij}) - b_l(t_{ij-1})\}\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\phi_i$, respectively. Note that $\tilde{t}_{i1}$ and $\tilde{t}_{i2}$ are different time points. The augmented likelihood associated with the second stage of data augmentation is given by

$$L_c(\boldsymbol{\theta}) = \prod_{i=1}^{n} g(\phi_i)\left\{\prod_{j=1}^{K_i}\prod_{l=1}^{L}\mathcal{P}_{Z_{ijl}}(Z_{ijl}|\phi_i)\right\}\left\{\prod_{m=1}^{M}\mathcal{P}_{W_{im1}}(W_{im1}|\phi_i)\mathcal{P}_{W_{im2}}(W_{im2}|\phi_i)^{\delta_{i2}+\delta_{i3}}\right\}.$$

The augmented data likelihood can be regarded as the complete data likelihood with all the W_{i1}'s, W_{i2}'s, W_{im1}'s, W_{im2}'s and Z_{ijl}'s being considered as missing data. One can obtain the observed likelihood by integrating out all the latent variables from the complete likelihood.

The EM algorithm

In the EM algorithm each iteration involves two steps, the expectation step and the maximization step. We start the derivation of the EM algorithm by considering the

expectation of the logarithm of the complete data likelihood with respect to the latent variables (W_{im1}'s, W_{im2}'s, Z_{ijl}'s and ϕ_i's) conditional on the observed data $\mathcal{D}$ and the current parameter $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)'}, \boldsymbol{\alpha}^{(d)'}, \boldsymbol{\gamma}^{(d)'}, \tilde{\boldsymbol{\gamma}}^{(d)'}, \nu)'$. The logarithm of complete likelihood is as follows,

$$\begin{aligned}\log L_c(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \log g(\phi_i) + \sum_{i=1}^{n}\sum_{j=1}^{K_i}\sum_{l=1}^{L} \log \mathcal{P}(Z_{ijl}|\phi_i) \\ &\quad + \sum_{i=1}^{n}\sum_{m=1}^{M} \log \mathcal{P}(W_{im1}|\phi_i) + \sum_{i=1}^{n}\sum_{m=1}^{M} (\delta_{i2}+\delta_{i3}) \log \mathcal{P}(W_{im2}|\phi_i).\end{aligned}$$

It yields the Q function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = \mathrm{E}\{\log L_c(\boldsymbol{\theta})|\mathcal{D}, \boldsymbol{\theta}^{(d)}\}$, which can be expressed as $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)}) + H_2(\boldsymbol{\alpha}, \tilde{\boldsymbol{\gamma}}, \boldsymbol{\theta}^{(d)}) + H_3(\nu, \boldsymbol{\theta}^{(d)}) + H_4(\boldsymbol{\theta}^{(d)})$, where

$$\begin{aligned}H_1(\boldsymbol{\beta}, \boldsymbol{\theta}^{(d)}) &= \sum_{i=1}^{n} Z_{i\cdot} \boldsymbol{x}_i'\boldsymbol{\beta} - \sum_{i=1}^{n} \mu_0(t_{iK_i}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}), \\ &\quad + \sum_{i=1}^{n}\sum_{j=1}^{K_i}\sum_{l=1}^{L} \mathrm{E}(Z_{ijl}|\mathcal{D}, \boldsymbol{\theta}^{(d)}) \log(\gamma_l) \\ H_2(\boldsymbol{\alpha}, \boldsymbol{\theta}^{(d)}) &= \sum_{i=1}^{n}\sum_{m=1}^{M} \Big[\{\mathrm{E}(W_{im1}|\mathcal{D}, \boldsymbol{\theta}^{(d)}) + (\delta_{i2}+\delta_{i3})\mathrm{E}(W_{im2}|\mathcal{D}, \boldsymbol{\theta}^{(d)})\}\{\log(\tilde{\gamma}_m) + \tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha}\} \\ &\quad - \tilde{\gamma}_m \exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha}) \mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})\{(\delta_{i1}+\delta_{i2}) I_m(R_i) + \delta_{i3} I_m(L_i)\} \Big] + L(\boldsymbol{\theta}^{(d)}), \\ H_3(\nu, \boldsymbol{\theta}^{(d)}) &= \nu \sum_{i=1}^{n} \mathrm{E}\{\log(\phi_i)|\mathcal{D}, \boldsymbol{\theta}^{(d)}\} - \nu \sum_{i=1}^{n} \mathrm{E}(\phi_i|\mathcal{D}) + n\nu \log(\nu) - n \log \Gamma(\nu),\end{aligned}$$

where $H_4(\boldsymbol{\theta}^{(d)})$ does not involve the unknown parameter $\boldsymbol{\theta}$. All the conditional expectations in the Q function have closed forms and evaluated at parameter values of the current step d. To make the conditional expectations convenient to present, some notations are introduced. Define $a_i = \nu + Z_{i\cdot}$, $b_i = \nu + \mu_0(t_{iK_i}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, $c_i = \Lambda_0(R_i) \exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})$ and $d_i = \Lambda_0(L_i) \exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})$ where $i = 1, \ldots, n$, then the all conditional expectations involved in the algorithm can be expressed in the follows,

$$
\begin{aligned}
\mathrm{E}(\phi_i|\mathcal{D},\boldsymbol{\theta}^{(d)}) &= \frac{a_i}{b_i+c_i}\frac{1-(1+c_i/b_i)^{a_i+1}}{1-(1+c_i/b_i)^{a_i}}\delta_{i1} \\
&\quad +\frac{a_i}{b_i+c_i}\frac{1-(\frac{b_i+c_i}{b_i+d_i})^{a_i+1}}{1-(\frac{b_i+c_i}{b_i+d_i})^{a_i}}\delta_{i2}+\frac{a_i}{b_i+d_i}\delta_{i3}, \\
\mathrm{E}(Z_{ijl}|\mathcal{D},\boldsymbol{\theta}^{(d)}) &= \frac{\gamma_l\{b_l(t_{ij})-b_l(t_{ij-1})\}}{\mu_0(t_{ij})-\mu_0(t_{ij-1})}Z_{ij}, \\
\mathrm{E}\{\log(\phi_i)|\mathcal{D},\boldsymbol{\theta}^{(d)}\} &= \psi(a_i)-\frac{(b_i+c_i)^{a_i}\log(b_i)-b_i^{a_i}\log(b_i+c_i)}{(b_i+c_i)^{a_i}-b_i^{a_i}}\delta_{i1} \\
&\quad -\frac{(b_i+c_i)^{a_i}\log(b_i+d_i)-(b_i+d_i)^{a_i}\log(b_i+c_i)}{(b_i+c_i)^{a_i}-(b_i+d_i)^{a_i}}\delta_{i2} \\
&\quad -\{\log(b_i+d_i)\}\delta_{i3}, \\
\mathrm{E}(W_{im1}|\mathcal{D},\boldsymbol{\theta}^{(d)}) &= \{\Lambda_0(R_i)\}^{-1}\tilde{\gamma}_m I_m(R_i)\mathrm{E}(W_{i1}|\mathcal{D},\boldsymbol{\theta}^{(d)}), \\
\mathrm{E}(W_{im2}|\mathcal{D},\boldsymbol{\theta}^{(d)}) &= \{\Lambda_0(R_i)-\Lambda_0(L_i)\}^{-1}\tilde{\gamma}_m\{I_m(R_i)-I_m(L_i)\}\mathrm{E}(W_{i2}|\mathcal{D},\boldsymbol{\theta}^{(d)}),
\end{aligned}
$$

where $\psi(\cdot)=\Gamma'(\cdot)/\Gamma(\cdot)$ is the digamma function. The expected values of W_{i1} and W_{i2} given $\mathcal{D}$ and $\boldsymbol{\theta}^{(d)}$ can be expressed as

$$
\begin{aligned}
\mathrm{E}(W_{i1}|\mathcal{D},\boldsymbol{\theta}^{(d)}) &= \frac{a_i c_i}{b_i}\left\{1-\left(\frac{b_i}{b_i+c_i}\right)^{a_i}\right\}^{-1}\delta_{i1}, \\
\mathrm{E}(W_{i2}|\mathcal{D},\boldsymbol{\theta}^{(d)}) &= \frac{a_i(c_i-d_i)}{b_i+d_i}\left\{1-\left(\frac{b_i+d_i}{b_i+c_i}\right)^{a_i}\right\}^{-1}\delta_{i2}.
\end{aligned}
$$

In the M-step one needs to find $\boldsymbol{\theta}^{(d+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$. Consider the partial derivatives of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\theta}$, which are given by

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n} \left[Z_{i\cdot} - \mu_0(t_{iK_i}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \mathrm{E}(\phi_i|\mathcal{D}) \right] \boldsymbol{x}_i, \\
\frac{\partial Q}{\partial \boldsymbol{\alpha}} &= \sum_{i=1}^{n} \Bigg[\{ \mathrm{E}(W_{i1}|\mathcal{D}) + (\delta_{i2} + \delta_{i3}) \mathrm{E}(W_{i2}|\mathcal{D}) \} \\
&\quad - \{ (\delta_{i1} + \delta_{i2}) \Lambda_0(R_i) + \delta_{i3} \Lambda_0(L_i) \} \exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha}) \mathrm{E}(\phi_i|\mathcal{D}) \Bigg] \boldsymbol{x}_i, \\
\frac{\partial Q}{\partial \gamma_l} &= -\sum_{i=1}^{n} b_l(t_{iK_i}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \mathrm{E}(\phi_i|\mathcal{D}) + \sum_{i=1}^{n} \sum_{j=1}^{K_i} \frac{\mathrm{E}(Z_{ijl}|\mathcal{D})}{\gamma_l}, \\
\frac{\partial Q}{\partial \tilde{\gamma}_m} &= \sum_{i=1}^{n} \Bigg[\tilde{\gamma}_m^{-1} \{ \mathrm{E}(W_{im1}|\mathcal{D}) + \delta_{i2} \mathrm{E}(W_{im2}|\mathcal{D}) \} \\
&\quad - \{ (\delta_{i1} + \delta_{i2}) I_m(R_i) + \delta_{i3} I_m(L_i) \} \exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha}) \mathrm{E}(\phi_i|\mathcal{D})], \\
\frac{\partial Q}{\partial \nu} &= \sum_{i=1}^{n} \mathrm{E}\{\log(\phi_i)|\mathcal{D})\} - \sum_{i=1}^{n} \mathrm{E}(\phi_i|\mathcal{D}) + n\log(\nu) + n - n\psi(\nu).
\end{aligned}$$

Note that we do not need to calculate $\partial Q/\partial \nu$, since it does not involve any other parameters, so we can just maximize $H_3(\nu, \boldsymbol{\theta}^{(d)})$ directly in the M-step. Setting $\partial Q/\partial \gamma_l = 0$, we find that the solution γ_l is a function of $\boldsymbol{\beta}$, and the relation is as follows,

$$\gamma_l(\boldsymbol{\beta}) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{K_i} \mathrm{E}(Z_{ijl}|\mathcal{D})}{\sum_{i=1}^{n} b_l(t_{iK_i}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \mathrm{E}(\phi_i|\mathcal{D})}.$$

Similarly, setting $\partial Q/\partial \tilde{\gamma}_m = 0$ leads to a closed form of the solution $\tilde{\gamma}_m$ as a function of α, i.e.

$$\tilde{\gamma}_m(\boldsymbol{\alpha}) = \frac{\sum_{i=1}^{n} \{ \mathrm{E}(W_{im1}|\mathcal{D}) + \delta_{i2} \mathrm{E}(W_{im2}|\mathcal{D}) \}}{\sum_{i=1}^{n} \{ (\delta_{i1} + \delta_{i2}) I_m(R_i) + \delta_{i3} I_m(L_i) \} \exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha}) \mathrm{E}(\phi_i|\mathcal{D})}.$$

Our EM algorithm is summarized as follows. First set $d = 0$ and initialize $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)'}, \boldsymbol{\alpha}^{(d)'}, \boldsymbol{\gamma}^{(d)'}, \tilde{\boldsymbol{\gamma}}^{(d)'}, \nu^{(d)})'$. Then repeat the following five steps until convergence.

1. Obtain $\boldsymbol{\beta}^{(d+1)}$ by solving the following system of p equations,

$$\sum_{i=1}^{n} \left[Z_{i\cdot} - \sum_{l=1}^{L} \gamma_l^{*(d)}(\boldsymbol{\beta}) b_l(t_{iK_i}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) \right] \boldsymbol{x}_i = \mathbf{0},$$

where

$$\gamma_l^{*(d)}(\boldsymbol{\beta}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{K_i} \mathrm{E}(Z_{ijl}|\mathcal{D}, \boldsymbol{\theta}^{(d)})}{\sum_{i=1}^{n} b_l(t_{iK_i}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})}.$$

2. Update $\gamma_l^{(d+1)} = \gamma_l^{*(d)}(\boldsymbol{\beta}^{(d+1)})$, for $l = 1, \ldots, L$.

3. Obtain $\boldsymbol{\alpha}^{(d+1)}$ by solving the following system of q equations

$$\sum_{i=1}^{n}\{\mathrm{E}(W_{i1}|\mathcal{D}, \boldsymbol{\theta}^{(d)}) + \delta_{i2}\mathrm{E}(W_{i2}|\mathcal{D}, \boldsymbol{\theta}^{(d)})\}\tilde{\boldsymbol{x}}_i$$
$$= \sum_{i=1}^{n}\sum_{m=1}^{M}\{(\delta_{i1} + \delta_{i2})I_m(R_i) + \delta_{i3}I_m(L_i)\}\tilde{\gamma}_m^{*(d)}(\boldsymbol{\alpha}) \exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})\tilde{\boldsymbol{x}}_i,$$

where

$$\tilde{\gamma}_m^{*(d)} = \frac{\sum_{i=1}^{n}\{\mathrm{E}(W_{im1}|\mathcal{D}, \boldsymbol{\theta}^{(d)}) + \delta_{i2}\mathrm{E}(W_{im2}|\mathcal{D}, \boldsymbol{\theta}^{(d)})\}}{\sum_{i=1}^{n}\{(\delta_{i1} + \delta_{i2})I_m(R_i) + \delta_{i3}I_m(L_i)\} \exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})\mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})}.$$

4. Update $\tilde{\gamma}_m^{(d+1)} = \tilde{\gamma}_m^{*(d)}(\boldsymbol{\alpha}^{(d+1)})$, for $m = 1, \ldots, M$.

5. Calculate $\nu^{(d+1)}$ by maximizing $H_3(\nu)$ directly.

Solving the system of equations in the step 1 and step 3 of the algorithm can be accomplished using standard root finding routines, available in practically all existing statistical software packages. The second and fourth step of the algorithm is a simple updating of $\boldsymbol{\gamma}^{(d)}$ and $\tilde{\boldsymbol{\gamma}}^{(d)}$ in closed form. The step 5 maximizes an univariate function which is flat. Thus, the implementation of the EM algorithm is straightforward and computationally inexpensive.

Variance estimation

The proposed EM algorithm produces the point estimate $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\boldsymbol{\alpha}}', \widehat{\boldsymbol{\gamma}}', \widehat{\boldsymbol{\eta}}', \hat{\nu})'$ at the convergence. We can obtain the variance estimate of $\hat{\boldsymbol{\theta}}$ by taking inverse of the observed information matrix $I(\hat{\boldsymbol{\theta}})$, i.e.,

$$\mathrm{var}(\hat{\boldsymbol{\theta}}) = \left\{ -\frac{\partial^2 \log \mathcal{L}_{obs}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right\}^{-1}.$$

The observed information matrix can be obtained by using the missing information principle (Louis, 1982),

$$-\frac{\partial^2 \log \mathcal{L}_{obs}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \text{var}\left\{\frac{\partial \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}.$$

All the quantities involved in $\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ and $\text{var}\{\partial \log \mathcal{L}_c(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\}$ have closed-form expressions and can be evaluated easily from the output of our EM algorithm. The details of the formula for these quantities are presented in the Appendix C. These expressions make the variance easy to compute, which is another appealing characteristic of the proposed approach.

An alternative way to obtain $I^{-1}(\hat{\boldsymbol{\theta}})$ is that the observed information matrix can be numerically approximated by

$$\begin{aligned} I(s,l) \approx & -h_n^{-2}\Big[\log\{\mathcal{L}_{obs}(\hat{\boldsymbol{\theta}} + h_n\vec{e}_s + h_n\vec{e}_l)\} - \log\{\mathcal{L}_{obs}(\hat{\boldsymbol{\theta}} + h_n\vec{e}_s)\} \\ & - \log\{\mathcal{L}_{obs}(\hat{\boldsymbol{\theta}} + h_n\vec{e}_l)\} + \log\{\mathcal{L}_{obs}(\hat{\boldsymbol{\theta}})\}\Big], \end{aligned}$$

where $\vec{e}_s$ is a binary vector whose sth element is 1 with all others being 0 and h_n is a small tuning constant. In particular, as the tuning parameter h_n goes to zero (i.e. $h_n \to 0$) the approximation is expected to improve, although numerical instability can be encountered if h_n is taken too small. In general, h_n should be selected to be of order $n^{-1/2}$, and we have found that selecting a decreasing sequence of h_n and approximating $I(\hat{\boldsymbol{\theta}})$ at each allows one to establish a range of values for which the tuning parameter performs well; i.e., a range of h_n for which the approximation of $I(\hat{\boldsymbol{\theta}})$ is stable. Proceeding in this fashion provides a straightforward, reliable, and computationally efficient method of estimating the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$. In this project, we adopt this numerical approximation of observed information matrix to obtain all parameter variances.

4.3 Simulation study

We evaluate the performance of our proposed model and approach through simulation. We generated data from different joint models and analyzed them by applying the proposed approach. For comparison purpose, we performed two separate analysis under the gamma frailty non-homogeneous Poisson process (GFNPMS) model for panel count data and the GORH model for failure time data by using the same simulated data.

To generate panel count data, the number of observation times K_i was generated from Poisson(6)+1 to ensure that there was at least one observation time for subject i, and the observation gap times were independently sampled from an exponential distribution with a rate parameter 2. The counting process associated with subject i was generated from the following model,

$$N_i(t_{ij}) - N_i(t_{ij-1}) \sim \mathcal{P}[\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\} \exp(x_{i1}\beta_1 + x_{i2}\beta_2)\phi_i],$$

where the baseline mean function $\mu_0(t) = \log(1+t) + t^2, x_{i1} \sim \mathcal{N}(0, 0.5^2)$, and $x_{i2} \sim$ Bernoulli(0.5). The failure time T_i was generated from the following gamma frailty PH model,

$$F(t|\tilde{\boldsymbol{x}}_i, \phi_i) = 1 - \exp\{-\Lambda_0(t) \exp(x_{i1}\alpha_1 + x_{i2}\alpha_2)\phi_i\},$$

where the baseline conditional cumulative hazard function $\Lambda_0(t) = \log(1+t) + t^{1/2}$. Note that non-homogeneous Poisson model and gamma frailty PH model share the same covariates, that is, $\tilde{x}_{i1} = x_{i1}$ and $\tilde{x}_{i2} = x_{i2}$. Each T_i was generated by solving $F(t|\tilde{\boldsymbol{x}}_i, \phi_i) = u_i$ numerically, where $u_i \sim \mathcal{U}_{(0,1)}$. For interval-censored data, the number of observation times for each subject was generated according to 1 plus a Poisson random variable having mean parameter 6. The gap times between adjacent observations were sampled according to an exponential distribution with rate 2. Each

of the regression parameters took on values 1 or -1, with $\beta_1 = \alpha_1$ and $\beta_2 = \alpha_2$. The gamma frailty ϕ_i's were generated from $\mathcal{G}a(\nu, \nu)$ with ν taking $0.5, 1$ or 4. There are 12 different parameter configurations. Five hundred data sets were generated, and the size of each sample is 200. The average right-censoring rate varied from 8% to 47% across the 12 parameter configurations. The tolerance value ϵ for claiming convergence was taken to be 10^{-4} for all simulations. To fit the proposed model, the degree of the monotone splines was specified to be 3 and a knot set consisting of 6 equally spaced knots. The initial values for the EM algorithm were specified to be $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}^{(0)'}, \boldsymbol{\alpha}^{(0)'}, \boldsymbol{\gamma}^{(0)'}, \tilde{\boldsymbol{\gamma}}^{(0)'}, \nu^{(0)})' = (\mathbf{0.5}'_2, \mathbf{0.5}'_2, \mathbf{0.1}'_7, \mathbf{0.1}'_7, 0.01)'$. To approximate the observed information matrix, the tuning parameter h_n was taken to be 0.001. We tried some other h_n values such as 0.1 and 0.01 which provide us similar and stable results.

Table 4.1 presents the relative bias (RBias) of parameter estimates, the sample standard deviations (SSD) of the 500 point estimates, the estimated standard errors (ESE) obtained by the numerical approximation, and the 95% coverage probability (CP95) for each parameter configuration.

As seen in Table 4.1, the proposed approach performs well in all parameter configurations. The relative biases are close to zero, indicating that our proposed estimators are unbiased; the SSDs and ESEs are close, indicating that the variance estimates using numerical approximation are accurate; and the CP95s are close to 0.95, indicating that the asymptotical normality is valid. As the value of frailty variance parameter increases, the estimated standard errors of regression parameters decreases. Because when ν increases, the frailty variance decreases, and this leads to less variation in the observed data, which further leads to the decreased variances of the regression estimates.

To see the performance of the baseline mean function estimation, we consider

adjusted mean squared error (AMSE) at each observation time t,

$$\text{AMSE}\{\hat{\mu}_0(t)\} = \frac{1}{500}\sum_{j=1}^{500}\frac{\{\hat{\mu}_0^{(j)}(t) - \mu_0(t)\}^2}{\{\mu_0(t)\}^2},$$

where $\hat{\mu}_0^{(j)}$ is the estimate of μ_0 from the jth data set, $j = 1, \ldots, 500$. Here AMSE is used to adjust the potentially large scale of $\mu_0(t)$. All the AMSEs are small values, indicating that the proposed approach performs well in terms of baseline mean function estimation. To see the performance of the marginal survival function estimation, we consider the mean squared error at observation time t,

$$\text{MSE}\{\hat{S}(t|\boldsymbol{x})\} = \frac{1}{500}\sum_{j=1}^{500}\{\hat{S}^{(j)}(t|\boldsymbol{x}) - S(t|\boldsymbol{x})\}^2,$$

where $S(t|\boldsymbol{x})$ is the true survival function and is known, and $\hat{S}^{(j)}(t|\boldsymbol{x})$ is the estimate of $S(t|\boldsymbol{x})$ from our approach for the jth data set, $j = 1, \ldots, 500$. The global mean (maximum) squared error of $\hat{S}(t|\boldsymbol{x})$, denoted by meanMSE (maxMSE), is taken as the mean (maximum) of the local MSEs of $S(t|\boldsymbol{x})$ over the set of time points. The smaller these global MSEs are, the better estimation for the survival functions. Table 4.2 also presents the global mean and maximum MSEs of the estimated survival functions with different covariate combinations, $\tilde{\boldsymbol{x}} = (0,0), (0,1)$ and $(1,0)$. From Table 4.2, all the global mean and maximum MSEs of survival function estimates are very small for all parameter configurations, which suggests that our method provides accurate estimates of the survival functions.

In addition, we considered a simulation with larger sample size to see the performance of the proposed approach, when the sample size n was increased from 200 to 400. The regression parameters $(\beta_1, \beta_2) = (\alpha_1, \alpha_2) = (1, -1)$ and ν took the values of 0.5, 1 and 4.

Table 4.3 presented the results when the sample size $n = 400$. For each parameter configuration, the relative bias of all parameter estimates decrease and the estimated

Table 4.1 Simulation results from the joint analysis. Empirical relative bias (RBias) and standard deviation (SSD) of the 500 estimates of $\boldsymbol{\theta}$, the average of the estimated standard errors (ESE), and the empirical coverage probabilities associated with 95% Wald confidence intervals (CP95).

		$\nu = 0.5$				$\nu = 1$				$\nu = 4$			
Parameter	Est	RBias	SSD	ESE	CP95	RBias	SSD	ESE	CP95	RBias	SSD	ESE	CP95
$\beta_1 = 1$	$\hat{\beta}_1$	0.004	0.216	0.210	0.944	0.006	0.158	0.150	0.938	0.001	0.081	0.082	0.952
$\beta_2 = 1$	$\hat{\beta}_2$	0.001	0.215	0.207	0.946	0.006	0.151	0.148	0.942	0.000	0.079	0.081	0.946
$\alpha_1 = 1$	$\hat{\alpha}_1$	0.052	0.328	0.312	0.934	0.074	0.270	0.259	0.938	0.090	0.221	0.221	0.944
$\alpha_2 = 1$	$\hat{\alpha}_2$	0.038	0.343	0.305	0.914	0.053	0.263	0.251	0.946	0.079	0.226	0.210	0.920
	$\hat{\nu}$	0.012	0.053	0.051	0.948	0.025	0.110	0.106	0.938	0.026	0.528	0.512	0.954
$\beta_1 = 1$	$\hat{\beta}_1$	0.000	0.212	0.219	0.964	0.001	0.170	0.158	0.930	0.006	0.090	0.093	0.960
$\beta_2 = -1$	$\hat{\beta}_2$	0.015	0.214	0.214	0.942	0.005	0.168	0.157	0.940	0.004	0.093	0.092	0.958
$\alpha_1 = 1$	$\hat{\alpha}_1$	0.041	0.334	0.317	0.938	0.063	0.276	0.259	0.922	0.054	0.224	0.209	0.934
$\alpha_2 = -1$	$\hat{\alpha}_2$	0.062	0.346	0.308	0.920	0.025	0.273	0.251	0.938	0.039	0.218	0.200	0.926
	$\hat{\nu}$	0.016	0.058	0.057	0.956	0.026	0.133	0.120	0.944	0.035	0.657	0.641	0.956
$\beta_1 = -1$	$\hat{\beta}_1$	0.023	0.217	0.211	0.938	0.010	0.149	0.151	0.944	0.000	0.083	0.082	0.940
$\beta_2 = 1$	$\hat{\beta}_2$	0.001	0.207	0.207	0.948	0.001	0.143	0.149	0.938	0.006	0.082	0.081	0.944
$\alpha_1 = -1$	$\hat{\alpha}_1$	0.095	0.332	0.314	0.920	0.039	0.275	0.258	0.950	0.089	0.240	0.221	0.922
$\alpha_2 = 1$	$\hat{\alpha}_2$	0.054	0.322	0.304	0.932	0.069	0.277	0.251	0.922	0.068	0.226	0.210	0.916
	$\hat{\nu}$	0.014	0.054	0.051	0.934	0.021	0.109	0.105	0.944	0.028	0.506	0.513	0.964
$\beta_1 = -1$	$\hat{\beta}_1$	0.010	0.217	0.218	0.954	0.004	0.158	0.159	0.948	0.004	0.087	0.093	0.956
$\beta_1 = -1$	$\hat{\beta}_2$	0.001	0.220	0.214	0.934	0.005	0.157	0.157	0.952	0.005	0.089	0.092	0.954
$\beta_1 = -1$	$\hat{\alpha}_1$	0.029	0.345	0.317	0.948	0.050	0.285	0.260	0.932	0.050	0.223	0.210	0.936
$\beta_1 = -1$	$\hat{\alpha}_2$	0.039	0.312	0.308	0.946	0.063	0.253	0.252	0.932	0.035	0.199	0.199	0.946
	$\hat{\nu}$	0.012	0.055	0.057	0.950	0.021	0.121	0.120	0.956	0.040	0.653	0.644	0.952

Table 4.2 The global mean and maximum AMSEs ($\times 10^{-2}$) of the estimates of $\hat{\mu}_0$ if the baseline mean function μ_0; the global mean and maximum MSEs ($\times 10^{-3}$) of the estimates of $\hat{S}_{ij}$ of the survival function S_{ij}. The three (i, j) combinations $(0, 0), (0, 1)$ and $(1, 0)$ correspond to three different covariate combinations $(\tilde{x}_1, \tilde{x}_2) = (0, 0), (0, 1)$ and $(1, 0)$, respectively.

		meanAMSE	maxAMSE	meanMSE			maxMSE		
ν	$(\beta_1, \beta_2, \alpha_1, \alpha_2)$	μ_0	μ_0	S_{00}	S_{01}	S_{10}	S_{00}	S_{01}	S_{10}
0.5	(1,1,1,1)	2.6	3.5	2.5	1.5	2.8	5.0	2.5	6.0
	(1,-1,1,-1)	2.3	4.5	2.3	3.1	2.9	4.1	7.3	7.1
	(-1,1,-1,1)	2.2	3.0	2.6	1.6	6.9	5.1	2.6	12.3
	(-1,-1,-1,-1)	2.5	5.1	2.3	2.8	6.8	3.9	7.1	10.8
1	(1,1,1,1)	1.3	2.3	2.2	0.8	1.4	3.6	2.8	6.5
	(1,-1,1,-1)	1.5	3.6	1.7	3.3	1.4	3.7	6.7	7.0
	(-1,1,-1,1)	1.1	1.9	1.9	1.0	7.6	3.1	2.8	12.7
	(-1,-1,-1,-1)	1.3	3.8	1.9	3.3	7.2	3.5	8.2	11.7
4	(1,1,1,1)	0.4	1.4	1.1	0.0	0.0	3.1	2.7	6.0
	(1,-1,1,-1)	0.6	3.4	1.0	3.8	0.0	4.2	8.1	7.3
	(-1,1,-1,1)	0.4	1.4	1.2	0.0	10.3	3.2	2.6	16.7
	(-1,-1,-1,-1)	0.6	3.2	1.0	3.3	7.1	4.1	6.9	10.3

standard errors of all parameter estimates decrease compared to the results from the simulation study when the sample size $n = 200$.

Table 4.3 Simulation results from joint analysis with larger sample size $n = 400$. Empirical bias (Bias) and standard deviation (SSD) of the 500 estimates of $\boldsymbol{\theta}$, the average of the estimated standard errors (ESE), and the empirical coverage probabilities associated with 95% Wald confidence intervals (CP95). The regression parameter values are $(\beta_1, \beta_2, \alpha_1, \alpha_2) = (1, -1, 1, -1)$. The frailty variance parameter ν takes the value of 0.5, 1 and 4.

	$\nu = 0.5$				$\nu = 1$				$\nu = 4$			
Est	RBias	SSD	ESE	CP95	RBias	SSD	ESE	CP95	RBias	SSD	ESE	CP95
$\hat{\beta}_1$	0.003	0.157	0.152	0.948	0.011	0.113	0.111	0.944	0.001	0.063	0.065	0.942
$\hat{\beta}_2$	0.006	0.159	0.151	0.942	0.005	0.113	0.111	0.954	0.005	0.064	0.065	0.954
$\hat{\alpha}_1$	0.038	0.244	0.219	0.930	0.035	0.193	0.181	0.934	0.029	0.157	0.146	0.926
$\hat{\alpha}_2$	0.030	0.236	0.216	0.924	0.019	0.172	0.176	0.946	0.035	0.140	0.140	0.938
$\hat{\nu}$	0.008	0.040	0.040	0.964	0.015	0.076	0.084	0.968	0.025	0.463	0.443	0.938

We applied the GFNPMS model (Yao et al., 2016) to the simulated panel count data through using an R package `PCDSpline` and GORH model (Zhou et al., 2016) to the interval-censored data through using an R package `ICGOR`, respectively. Note that the frailty variance parameter ν in the GORH model cannot be estimated. We can run different GORH models and select the "best" GORH model according to some model selection criteria. Table 4.4 presents that, under the GFNPMS model, the accuracy of regression parameter are comparable with those under the joint analysis. However, for the frailty variance parameter estimate $\hat{\nu}$, GFNPMS model provides larger estimated standard errors. That means joint analysis is more efficient than GFNPMS model for using the panel count data alone. For the regression parameter estimates under the GORH model, both the relative bias and the estimated standard errors of all parameters are larger than those under the joint analysis. The CP95s from the GORH model are smaller than those from the joint analysis. These results indicate that joint analysis is more efficient than GORH model analysis.

4.4 Data application

After removing women who have no follow-up observations, there are 352 participants included in the data analysis. The panel count response is the number of condom

Table 4.4 Simulation results from the joint analysis, GFNPMS model and GORH model where the data were generated from the joint models. Empirical bias (Bias) and standard deviation (SSD) of the 500 estimates of $\boldsymbol{\theta}$, the average of the estimated standard errors (ESE), and the empirical coverage probabilities associated with 95% Wald confidence intervals (CP95).

	Joint analysis				GFNPMS model				GORH model			
	RBias	SSD	ESE	CP95	RBias	SSD	ESE	CP95	RBias	SSD	ESE	CP95
$\beta_1 = -1$	0.010	0.217	0.218	0.954	0.008	0.217	0.219	0.950	–	–	–	–
$\beta_2 = -1$	0.001	0.220	0.214	0.934	0.001	0.223	0.214	0.932	–	–	–	–
$\alpha_1 = -1$	0.029	0.345	0.317	0.948	–	–	–	–	0.040	0.383	0.370	0.936
$\alpha_2 = -1$	0.039	0.312	0.308	0.946	–	–	–	–	0.050	0.346	0.344	0.922
$\nu = 0.5$	0.012	0.055	0.057	0.950	0.024	0.058	0.060	0.950	–	–	–	–
$\beta_1 = -1$	0.004	0.158	0.159	0.948	0.006	0.158	0.160	0.950	–	–	–	–
$\beta_2 = -1$	0.005	0.157	0.157	0.952	0.006	0.158	0.157	0.950	–	–	–	–
$\alpha_1 = -1$	0.050	0.285	0.260	0.932	–	–	–	–	0.050	0.317	0.289	0.926
$\alpha_2 = -1$	0.063	0.253	0.252	0.932	–	–	–	–	0.080	0.281	0.272	0.934
$\nu = 1$	0.021	0.121	0.120	0.956	0.031	0.129	0.126	0.964	–	–	–	–
$\beta_1 = -1$	0.004	0.087	0.093	0.956	0.003	0.088	0.093	0.952	–	–	–	–
$\beta_2 = -1$	0.005	0.089	0.092	0.954	0.006	0.090	0.092	0.954	–	–	–	–
$\alpha_1 = -1$	0.050	0.223	0.210	0.936	–	–	–	–	0.049	0.227	0.219	0.938
$\alpha_2 = -1$	0.035	0.199	0.199	0.946	–	–	–	–	0.037	0.205	0.205	0.938
$\nu = 4$	0.040	0.653	0.644	0.952	0.052	0.679	0.675	0.952	–	–	–	–

non-use, and the failure time is the time to get first infection of any STI since the enrollment. Of the 352 subjects women, 47 (13.4%) have left-censored failure times, 196 (55.7%) have interval-censored failure times, 109 (30.9%) have right-censored failure times. The covariates of interest are the baseline age $(x_1, \tilde{x}_1)$, the baseline STI status $(x_2, \tilde{x}_2)$, the age at first sex $(x_3, \tilde{x}_3)$ and race $(x_4, \tilde{x}_4)$. Specifically, the race has two categories, African American (88.3%), others (11.7%). Table 4.5 presents the STI data covariates information. The African American is the baseline race. To model μ_0 and Λ_0, we took the degree to be 3 and the number of interior knots is 6. The convergence tolerance is 10^{-4}.

Table 4.6 presents the covariate effects (Point), estimated standard errors (ESE) and P-values from the joint modeling, the GFNPMS model and the GORH model, respectively. The baseline age has a positive significant effect on the number of condom non-use, and the first sex age has a negative significant effect on the number of condom non-use. The frailty variance parameter estimate $\hat{\nu}$ is significant, at the

Table 4.5 Demographic and behavioral characteristics of the Young Women's Project participants.

Covariates	Description	Code
Baseline age	Mean (SD)= 15.82(1.11) Max (Min)= 18(14)	
Baseline STI status	No (64.8%) Yes (35.2%)	0 1
First sex age	Mean (SD)=13.98(1.75) Max (Min)=20.3(10)	
Race	African American(88.3%) Others (11.7%)	0 1

level of significance 0.05, which means there exists an association between the number of condom non-use and the time to get first STI since the enrollment. The covariate effects under the GFNPMS model and GORH model are comparable with those under the joint model. However, there is some difference in the estimation of covariate effects from the joint analysis and the GORH model. We will do more investigation on this issue. In the future, we will run more models to choose the "best" one via comparing the AIC or BIC values.

Table 4.6 Sexually transmitted infections data analysis results under the joint model, GFNPMS model and GORH model. The frailty variance parameter ρ in the GORH model used the $\hat{\nu}^{-1}$ from the joint analysis.

	Joint analysis			GFNPMS model			GORH model		
Parameter	Point	ESE	P-value	Point	ESE	P-value	Point	ESE	P-value
β_1	0.537	0.043	< 0.001	0.540	0.062	< 0.001	–	–	–
β_2	0.058	0.166	0.728	0.053	0.179	0.767	–	–	–
β_3	-0.343	0.048	< 0.001	-0.345	0.052	< 0.001	–	–	–
β_4	0.337	0.233	0.148	0.344	0.249	0.168	–	–	–
α_1	0.016	0.084	0.851	–	–	–	0.155	0.128	0.279
α_2	0.521	0.226	0.021	–	–	–	1.257	0.279	< 0.001
α_3	-0.116	0.063	0.066	–	–	–	-0.235	0.080	0.010
α_4	-0.337	0.338	0.319	–	–	–	-0.987	0.413	0.025
ν	0.547	0.038	< 0.001	0.476	0.034	< 0.001	–	–	–

4.5 Discussion

In this chapter, we proposed a joint model for analyzing panel count data and interval-censored data. The joint model includes the non-homogeneous Poisson process model and the PH model which are connected via a shared gamma frailty. We developed a computationally efficient EM algorithm based on a Poisson data augmentation to jointly estimate all the unknown parameters. The proposed method is shown to work well in our simulation studies and outperforms the univariate analysis of panel count data and interval-censored failure time data separately. In the future, we aim to propose a more general model to better fit the STI data. For example, we could add a power to the frailty variable to make the model more flexible so that the frailty variable can contribute differently to the panel count response and failure time.

Bibliography

P. Andersen and B. Ronn. A nonparametric test for comparing two samples where all observations are either left- or right- censored. *Biometrics*, 51:323–329, 1995.

N. Balakrishnan and X. Zhao. New multi-sample nonparametric tests for panel count data. *The Annals of Statistics*, 37(3):1112–1149, 2009.

T. Banerjee, M. Chen, D. Dey, and S. Kim. Bayesian analysis of generalized odds-rate hazards models for survival data. *Lifetime Data Analysis*, 13:241–260, 2007.

S. Bennett. Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2(2):273–277, 1983.

R. A. Betensky, D. Rabinowitz, and A.A. Tsiatis. Computationally simple accelerated failure time regression for interval censored data. *Biometrika*, 88(3):703–711, 2001.

D.P. Byar, C. Blackard, and The Veterans Administration Cooperative Urological Research Group. Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage I bladder cancer. *Urology*, 10:556–561, 1977.

B. Cai, X. Lin, and L. Wang. Bayesian proportional hazards model for current status data with monotone splines. *Computational Statistics & Data Analysis*, 55:2644–2651, 2011.

T. Cai and R.A. Betensky. Hazard regression for interval-censored data with penalized spline. *Biometrics*, 59:570–579, 2003.

Mário Castro, Minghui Chen, Joseph Ibrahim, and John Klein. Bayesian transformation models for multivariate survival data. *Scandinavian Journal of Statistics*, 41:187–199, 2014.

D. Clayton and J. Cuzick. The em algorithm for cox's regression model using GLIM. *Journal of the Royal Statistical Society, Series C*, 34(2):148–156, 1985.

D.G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151, 1978.

B.J. Cowling, J.L. Hutton, and J.E.H. Shaw. Joint modelling of event counts and survival times. *Journal of the Royal Statistical Society. Series C*, 55:31–39, 2006.

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

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

D. Dabrowska and K. Doksum. Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association*, 83(403):744–749, 1988.

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

H. Fang, J. Sun, and M. Lee. Nonparametric survival comparisons for interval-censored continuous data. *Statistica Sinica*, 12:1073–1083, 2002.

D.M. Finkelstein. A proportional hazards model for interval-censored failure time data. *Biometrics*, 42:845–854, 1986.

R. Gentleman and C.J. Geyer. Maximum likelihood for interval censored data: Consistency and computation. *Biometrika*, 81(3):618–623, 1994.

P. Ghosh and W. Tu. Assessing sexual attitudes and behaviors of young women: a joint model with nonlinear time effects, time varying covariates, and dropouts. *Journal of the American Statistical Association*, 104(486):474–485, 2009.

W.B. Goggins, D.M. Finkelstein, D.A. Schoenfeld, and A.M. Zaslavsky. A Markov chain Monte Carlo EM algorithm for analyzing interval-censored data under the Cox proportional hazards model. *Biometrics*, 54:1498–1507, 1998.

P. Groeneboom and J.A. Wellner. *Information bounds and nonparametric maximum likelihood estimation*. DMV Seminar, Band 19, Birkhauser, New York, 1992.

P. Groeneboom, G. Jongbloed, and B. Witte. Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *The Annals of Statistics*, 38:352–387, 2010.

T. Hanson and M. Yang. Bayesian semiparametric proportional odds models. *Biometrics*, 63:88–95, 2007.

J. Harezlak and W. Tu. Estimation of survival functions in interval and right censored data using STD behavioural diaries. *Statistics in Medicine*, 25:4053–4064, 2006.

D. Harrington and T. Fleming. A class of rank test procedures for censored survival data. *Biometrika*, 69:553–566, 1982.

X. He, X. Tong, J. Sun, and R. Cook. Regression analysis of multivariate panel count data. *Biostatistics*, 9(2):234–248, 2008.

D.V. Hinkley and G. Runger. The analysis of transformed data. *Journal of the American Statistical Association*, 79:302–309, 1984.

X. Hu, J. Sun, and L.-J. Wei. Regression parameter estimation from panel counts. *Scandinavian Journal of Statistics*, 30:25–43, 2003.

L. Hua and Y. Zhang. Spline-based semiparametric projected generalized estimating equation method for panel count data. *Biostatistics*, 13(3):440–454, 2012.

L. Hua, Y. Zhang, and W. Tu. A spline-based semiparametric sieve likelihood method for over-dispersed panel count data. *The Canadian Journal of Statistics*, 42(2): 217–245, 2014.

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

C. Huang, M. Wang, and Y. Zhang. Analysing panel count data with informative observation times. *Biometrika*, 93(4):763–775, 2006.

J. Huang. Maximum likelihood estimation for proportional odds regression model with current status data. *Analysis of Censored Data, IMS Lecture notes-Monograph Series*, 27:129–146, 1995.

J. Huang. Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics*, 24(2):540–568, 1996.

J. Huang and A.J. Rossini. Sieve estimation for the proportional-odds failure-time regression model with interval censoring. *Journal of the American Statistical Association*, 92(439):960–967, 1997.

J. Huang and J.A. Wellner. Interval censored survival data: a review of recent progress. *Proceedings of the first Seattle symposium in biostatistics: survival analysis*, 1997.

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

J.D. Kalbfleisch and J.F. Lawless. The analysis of panel data under a markov assumption. *Journal of the American Statistical Association*, 80(392):863–871, 1985.

T. Lee, L. Zeng, D. Thompson, and C.B. Dean. Comparison of imputation methods for interval censored time-to-event data in joint modelling of tree growth and mortality. *The Canadian Journal of Statistics*, 39(3):438–457, 2011.

L. Li, T. Watkins, and Q. Yu. An EM algorithm for smoothing the self-consistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics*, 24:531–542, 1997.

N. Li, H. Zhao, and J. Sun. Semiparametric transformation models for panel count data with correlated observation and follow-up times. *Statistics in Medicine*, 32: 3039–3054, 2013.

Y. Li, H. Zhao, J. Sun, and K. Kim. Nonparametric tests for panel count data with unequal observation processes. *Computational Statistics & Data Analysis*, 73: 103–111, 2014.

Z. Li, H. Liu, and W. Tu. A sexually transmitted infection screening algorithm based on semiparametric regression models. *Statistics in Medicine*, 34:2844–2857, 2015.

K.-Y. Liang and S.L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

D.Y. Lin, David Oakes, and Z. Ying. Additive hazards regression with current status data. *Biometrika*, 85(2):289–298, 1998.

X. Lin and L. Wang. Semiparametric probit model for case 2 interval-censored failure time data. *Statistics in Medicine*, 29:972–981, 2010.

X. Lin, B. Cai, L. Wang, and Z. Zhang. A Bayesian proportional hazards model for general interval-censored data. *Lifetime Data Analysis*, 21:470–490, 2015.

T. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 44(2):226–233, 1982.

M. Lu, Y. Zhang, and J. Huang. Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika*, 94(3):705–718, 2007.

M. Lu, Y. Zhang, and J. Huang. Semiparametric estimation methods for panel count data using monotone B-splines. *Journal of the American Statistical Association*, 104(487):1060–1070, 2009.

C.S. McMahan, L. Wang, and J.M. Tebbs. Regression analysis for current status data using EM algorithm. *Statistics in Medicine*, 32:4552–4466, 2013. doi: 10.1002/sim.5863.

M. Ott, A. Katschke, W. Tu, and J. Fortenberry. Longitudinal associations among relationship factors, partner change, and sexually transmitted infection acquisition in adolescent women. *Sexually Transmitted Diseases*, 38(3):153–157, 2011.

R. Peto. Experimental survival curves for interval-censored data. *Applied Statistics*, 22:86–91, 1973.

D. Rabinowitz, R. Betensky, and A.A. Tsiatis. Using conditional logistic regression to fit proportional odds models to interval censored data. *Biometrics*, 56:511–518, 2000.

J.O. Ramsay. Monotone regression splines in action. *Statistical Science*, 3(4):425–461, 1988.

P. Rosenberg. Hazard function estimation using B-splines. *Biometrics*, 51:874–887, 1995.

A. Rossini and A. Tsiatis. A semiparametric proportional odds regression model for the analysis of current status data. *Journal of the American Statistical Association*, 91(434):713–721, 1996.

Glen A. Satten. Rank-based inference in the proportional hazards model for interval censored data. *Biometrika*, 83(2):355–370, 1996.

D. Scharfstein, A. Tsiatis, and P. Gilbert. Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Analysis*, 4:355–391, 1998.

X. Shen. Proportional odds regression and sieve maximum likelihood estimation. *Biometrika*, 85(1):165–177, 1998.

S. Shiboski. Generalized additive models for current status data. *Lifetime Data Analysis*, 4:29–50, 1998.

J. Sun. *The statistical analysis of interval-censored failure time data*. Springer-Verlag, New York, 2006.

J. Sun and H. Fang. A nonparametric test for panel count data. *Biometrika*, 90(1): 199–208, 2003.

J. Sun and J. Kalbfleisch. The analysis of current status data on point processes. *Journal of the American Statistical Association*, 88(424):1449–1454, 1993.

J. Sun and J. Kalbfleisch. Estimation of the mean function of point processes based on panel count data. *Statistica Sinica*, 5:279–290, 1995.

J. Sun and L.J. Wei. Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society. Series B (Methodological)*, 62:293–302, 2000.

J. Sun and X. Zhao. *Statistical analysis of panel count data*. New York: Springer, 2013.

J. Sun, Zhao Q., and X. Zhao. Generalized log-rank tests for interval censored failure time data. *Scandinavian Journal of Statistics*, 32:49–57, 2005.

J. Sun, X. Tong, and X. He. Regression analysis of panel count data with dependent observation times. *Biometrics*, 63(1053–1059), 2007.

M.A. Tanner. *Tools for statistical inference: methods for exploration of posterior distributions and likelihood functions*. Springer-Verlag, New York, 1996.

P. Thall and J. Lachin. Analysis of recurrent events: nonparametric methods for random-interval count data. *Journal of the American Statistical Association*, 83 (402):339–347, 1988.

P.F. Thall and S.C. Vail. Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 46:657–671, 1990.

A. A. Tsiatis and M. Davidian. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14:809–834, 2004.

W. Tu, B. Batteiger, S. Wiehe, S. Ofner, B. Van Der Pol, B. Katz, D. Orr, and D. Fortenberry. Time from first intercourse to first sexually transmitted infection diagnosis among adolescent women. *Archives of Pediatrics and Adolescent Medicine*, 163(12):1106–1111, 2009.

B.W. Turnbull. The empirical distribution with arbitrarily grouped censored and truncated data. *Journal of the Royal Statistical Society, Series B*, 38:290–295, 1976.

L. Wang and D. Dunson. Semiparametric bayes proportional odds models for current status data with under-reporting. *Biometrics*, 67:1111–1118, 2011.

L. Wang, C.S. McMahan, M.G. Hudgens, and Z.P. Qureshi. A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics*, 2015.

J.A. Wellner and Y. Zhan. A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association*, 92(439):945–959, 1997.

J.A. Wellner and Y. Zhang. Two estimators of the mean of a counting process with panel count data. *The Annals of Statistics*, 28(3):779–814, 2000.

J.A. Wellner and Y. Zhang. Two likelihood-based semiparametric estimation methods for panel count data with covariates. *The Annals of Statistics*, 35:2106–2142, 2007.

L. Wu, W. Liu, G. Yi, and Y. Huang. Analysis of longitudinal and survival data: joint modeling, inference methods and issues. *Journal of Probability and Statistics*, 2012:1–17, 2012.

R. Xu and D. Harrington. A semiparametric estimate of treatment effects with censored data. *Biometrics*, 57:875–885, 2001.

B. Yao, L. Wang, and X. He. Semiparametric regression analysis of panel count data allowing for within-subject correlation. *Computational Statistics & Data Analysis*, 97:47–59, 2016. doi: 10.1016/j.csda.2015.11.017.

Z. Yu, W. Tu, and M. Lee. A semi-parametric threshold regression analysis of sexually transmitted infections in adolescent women. *Statistics in Medicine*, 28:3029–3042, 2009.

D. Zeng, G. Yin, and J. Ibrahim. Semiparametric transformation models for survival data with cure fraction. *Journal of the American Statistical Association*, 101(474): 670–684, 2006.

H. Zhang, H. Zhao, J. Sun, D. Wang, and K. Kim. Regression analysis of multivariate panel count data with an informative observation process. *Journal of Multivariate Analysis*, 119:71–80, 2013.

Y. Zhang. A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika*, 89(1):39–48, 2002.

Y. Zhang. Nonparametric k-sample tests with panel count data. *Biometrika*, 93(4): 777–790, 2006.

Y. Zhang and M. Jamshidian. The gamma frailty poisson model for the nonparametric estimation of panel count data. *Biometrics*, 59:1099–1106, 2003.

Z. Zhang and J. Sun. Interval censoring. *Statistical Methods in Medical Research*, 19: 53–70, 2010.

Q. Zhao and J. Sun. Generalized log-rank test for mixed interval-censored failure time data. *Statistics in Medicine*, 23:1621–1629, 2004.

X. Zhao and X. Tong. Semiparametric regression analysis of panel count data with informative observation times. *Computational Statistics & Data Analysis*, 55: 291–300, 2011.

X. Zhao, X. Tong, and J. Sun. Robust estimation for panel count data with informative observation times. *Computational Statistics and Data Analysis*, 57:33–40, 2013.

J. Zhou, J. Zhang, and W. Lu. On efficient estimation of generalized odds-rate models with interval-censored data. *Manuscript*, 2016.

Appendix A

Chapter 2 Appendix and Supplementary Materials

A.1 Formula of the quantities involved in var($\widehat{\boldsymbol{\theta}}$)

The detailed expressions of the quantities involved $Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})$ and var$\{\partial \log L_c(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\}$ are presented below in order to obtain the variance estimate of $\widehat{\boldsymbol{\theta}}$. First, the second partial derivatives of $Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$, i.e., $\partial^2 Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$, are given by

$$\begin{aligned}
\partial^2 Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}' &= -\sum_{i=1}^{n} \mu_0(t_{iK_i}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \mathbf{E}(\phi_i | \mathcal{D}, \widehat{\boldsymbol{\theta}}) \boldsymbol{x}_i \boldsymbol{x}_i', \\
\partial^2 Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})/\partial \boldsymbol{\beta} \partial \gamma_l &= -\sum_{i=1}^{n} b_l(t_{iK_i}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \mathbf{E}(\phi_i | \mathcal{D}, \widehat{\boldsymbol{\theta}}) \boldsymbol{x}_i, \forall\ l, \\
\partial^2 Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})/\partial \gamma_l^2 &= -\sum_{i=1}^{n} \sum_{j=1}^{K_i} \frac{\mathbf{E}(Z_{ijl} | \mathcal{D}, \widehat{\boldsymbol{\theta}})}{{\gamma_l}^2}, \forall\ l, \\
\partial^2 Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})/\partial \nu^2 &= n\{\nu^{-1} - \psi'(\nu)\},
\end{aligned}$$

$\partial^2 Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})/\partial \boldsymbol{\beta} \partial \nu = 0$, $\partial^2 Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})/\partial \nu \partial \gamma_l = 0$ for each l, $\partial^2 Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})/\partial \gamma_l \partial \gamma_{l'} = 0$, for $l \neq l'$, where $\psi'(\nu)$ is the trigamma function with expression $\psi'(\nu) = \Gamma''(\nu)/\Gamma(\nu) - \{\Gamma'(\nu)/\Gamma(\nu)\}^2$ and the conditional expectations are given explicitly in Section 3.2.

Second, the necessary quantities in involved in $\text{var}\{\partial \log \mathcal{L}_c(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\}$ are given by

$$
\begin{aligned}
\text{cov}\Big(\frac{\partial \log \mathcal{L}_c}{\partial \boldsymbol{\beta}}, \frac{\partial \log \mathcal{L}_c}{\partial \boldsymbol{\beta}} | \mathcal{D}, \boldsymbol{\theta}\Big) &= \sum_{i=1}^{n} \{\mu_0(t_{iK_i}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta})\}^2 \text{var}(\phi_i | \mathcal{D}, \boldsymbol{\theta}) \boldsymbol{x}_i \boldsymbol{x}_i', \\
\text{cov}\Big(\frac{\partial \log \mathcal{L}_c}{\partial \boldsymbol{\beta}}, \frac{\partial \log \mathcal{L}_c}{\partial \gamma_l} | \mathcal{D}, \boldsymbol{\theta}\Big) &= \sum_{i=1}^{n} \mu_0(t_{iK_i}) b_l(t_{iK_i}) \{\exp(\boldsymbol{x}_i' \boldsymbol{\beta})\}^2 \text{var}(\phi_i | \mathcal{D}, \boldsymbol{\theta}) \boldsymbol{x}_i, \\
\text{cov}\Big(\frac{\partial \log \mathcal{L}_c}{\partial \boldsymbol{\beta}}, \frac{\partial \log \mathcal{L}_c}{\partial \nu} | \mathcal{D}, \boldsymbol{\theta}\Big) &= \sum_{i=1}^{n} \mu_0(t_{iK_i}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \\
&\quad \times \Big\{ \text{var}(\phi_i | \mathcal{D}, \boldsymbol{\theta}) - \text{cov}(\phi_i, \log(\phi_i) | \mathcal{D}, \boldsymbol{\theta}) \Big\} \boldsymbol{x}_i, \\
\text{cov}\Big(\frac{\partial \log \mathcal{L}_c}{\partial \gamma_l}, \frac{\partial \log \mathcal{L}_c}{\partial \nu} | \mathcal{D}, \boldsymbol{\theta}\Big) &= \sum_{i=1}^{n} b_l(t_{iK_i}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \\
&\quad \times \Big\{ \text{var}(\phi_i | \mathcal{D}, \boldsymbol{\theta}) - \text{cov}(\phi_i, \log(\phi_i) | \mathcal{D}, \boldsymbol{\theta}) \Big\}, \\
\text{cov}\Big(\frac{\partial \log \mathcal{L}_c}{\partial \gamma_l}, \frac{\partial \log \mathcal{L}_c}{\partial \gamma_l} | \mathcal{D}, \boldsymbol{\theta}\Big) &= \sum_{i=1}^{n} \{b_l(t_{iK_i}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta})\}^2 \text{var}(\phi_i | \mathcal{D}, \boldsymbol{\theta}) \\
&\quad + \frac{1}{\gamma_l^2} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \text{var}(Z_{ijl} | \mathcal{D}, \boldsymbol{\theta}), \\
\text{cov}\Big(\frac{\partial \log \mathcal{L}_c}{\partial \gamma_l}, \frac{\partial \log \mathcal{L}_c}{\partial \gamma_{l'}} | \mathcal{D}, \boldsymbol{\theta}\Big) &= \sum_{i=1}^{n} b_l(t_{iK_i}) b_{l'}(t_{iK_i}) \{\exp(\boldsymbol{x}_i' \boldsymbol{\beta})\}^2 \text{var}(\phi_i | \mathcal{D}, \boldsymbol{\theta}) \\
&\quad + \frac{1}{\gamma_l \gamma_{l'}} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \text{cov}(Z_{ijl}, Z_{ijl'} | \mathcal{D}, \boldsymbol{\theta}), \quad l \neq l', \\
\text{cov}\Big(\frac{\partial \log \mathcal{L}_c}{\partial \nu}, \frac{\partial \log \mathcal{L}_c}{\partial \nu} | \mathcal{D}, \boldsymbol{\theta}\Big) &= \sum_{i=1}^{n} \text{var}(\log \phi_i | \mathcal{D}, \boldsymbol{\theta}) - 2 \sum_{i=1}^{n} \text{cov}(\log(\phi_i), \phi_i | \mathcal{D}, \boldsymbol{\theta}) \\
&\quad + \sum_{i=1}^{n} \text{var}(\phi_i | \mathcal{D}, \boldsymbol{\theta}),
\end{aligned}
$$

where the conditional variance and covariance terms at the right hands have the

following form,

$$\begin{aligned}
\mathrm{var}(\phi_i|\mathcal{D},\boldsymbol{\theta}) &= \frac{\nu+Z_{i\cdot}}{\{\nu+\mu_0(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^2},\\
\mathrm{var}(Z_{ijl}|\mathcal{D},\boldsymbol{\theta}) &= \frac{\gamma_l\{b_l(t_{ij})-b_l(t_{ij-1})\}}{\mu_0(t_{ij})-\mu_0(t_{ij-1})}\left[1-\frac{\gamma_l\{b_l(t_{ij})-b_l(t_{ij-1})\}}{\mu_0(t_{ij})-\mu_0(t_{ij-1})}\right]Z_{ij},\\
\mathrm{cov}(Z_{ijl},Z_{ijl'}|\mathcal{D},\boldsymbol{\theta}) &= -\frac{\gamma_l\gamma_{l'}\{b_l(t_{ij})-b_l(t_{ij-1})\}\{b_{l'}(t_{ij})-b_{l'}(t_{ij-1})\}}{\{\mu_0(t_{ij})-\mu_0(t_{ij-1})\}^2}Z_{ij},\ l\neq l',\\
\mathrm{cov}(\phi_i,\log(\phi_i)|\mathcal{D},\boldsymbol{\theta}) &= \frac{\nu+Z_{i\cdot}}{\nu+\mu_0(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}\left\{\psi\left(\nu+Z_{i\cdot}+1\right)-\psi\left(\nu+Z_{i\cdot}\right)\right\},\\
\mathrm{var}(\log\phi_i|\mathcal{D},\boldsymbol{\theta}) &= \psi'(\nu+Z_{i\cdot}),\\
\mathrm{cov}(\phi_i,Z_{ijl}|\mathcal{D},\boldsymbol{\theta}) &= \mathrm{cov}(\log(\phi_i),Z_{ijl}|\mathcal{D},\boldsymbol{\theta})=0,
\end{aligned}$$

for all i, j, and l.

A.2 Derivation of the within-subject correlation.

Consider two non-overlapping intervals $(t_1,t_2]$ and $(t_3,t_4]$, and let Z_1 and Z_2 denote the count of the recurrent events within these two intervals, respectively, from the same subject with covariates $\boldsymbol{x}$. Below we provide the derivation of Pearson's correlation coefficient between Z_1 and Z_2 under the proposed gamma frailty Poisson process model.

Under the proposed model, it is known that $Z_1|\phi\sim\mathcal{P}(\lambda_1\phi)$ and $Z_2|\phi\sim\mathcal{P}(\lambda_2\phi)$ conditional on frailty $\phi\sim\mathcal{G}a(\nu,\nu)$, where $\lambda_1=\{\mu_0(t_2)-\mu_0(t_1)\}\exp(\boldsymbol{x}'\boldsymbol{\beta})$ and $\lambda_2=\{\mu_0(t_4)-\mu_0(t_3)\}\exp(\boldsymbol{x}'\boldsymbol{\beta})$ are the mean numbers of the recurrent events occurring within $(t_1,t_2]$ and $(t_3,t_4]$, respectively. First, using the law of iterated conditional expectations, one obtains

$$\mathrm{E}(Z_1)=\mathrm{E}\{\mathrm{E}(Z_1|\lambda_1\phi)\}=\mathrm{E}(\lambda_1\phi)=\lambda_1,$$

$$\mathrm{E}(Z_2)=\mathrm{E}\{\mathrm{E}(Z_2|\lambda_2\phi)\}=\mathrm{E}(\lambda_2\phi)=\lambda_2,$$

$$\mathrm{var}(Z_1)=\mathrm{E}\{\mathrm{var}(Z_1|\phi)\}+\mathrm{var}\{\mathrm{E}(Z_1|\phi)\}=\mathrm{E}(\lambda_1\phi)+\mathrm{var}(\lambda_1\phi)=\lambda_1+\lambda_1^2\nu^{-1},$$

$$\text{var}(Z_2) = \text{E}\{\text{var}(Z_2|\phi)\} + \text{var}\{\text{E}(Z_2|\phi)\} = \text{E}(\lambda_2\phi) + \text{var}(\lambda_2\phi) = \lambda_2 + \lambda_2^2\nu^{-1},$$

and

$$\begin{aligned}\text{E}(Z_1Z_2) &= \text{E}\{\text{E}(Z_1Z_2|\phi)\} = \text{E}\{\text{E}(Z_1|\phi)\text{E}(Z_2|\phi)\} = \text{E}\{(\lambda_1\phi)(\lambda_2\phi)\} = \lambda_1\lambda_2\text{E}(\phi^2) \\ &= \lambda_1\lambda_2(1+\nu^{-1}).\end{aligned}$$

Then the correlation between Z_1 and Z_2 is

$$\begin{aligned}\rho(Z_1, Z_2) &= \frac{\text{cov}(Z_1, Z_2)}{\sqrt{\text{var}(Z_1)\text{var}(Z_2)}} \\ &= \frac{\text{E}(z_1z_2) - \text{E}(z_1)\text{E}(z_2)}{\sqrt{\text{var}(Z_1)\text{var}(Z_2)}} \\ &= \frac{\lambda_1\lambda_2(1+\nu^{-1}) - \lambda_1\lambda_2}{\sqrt{(\lambda_1+\lambda_1^2\nu^{-1})(\lambda_2+\lambda_2^2\nu^{-1})}} \\ &= \{(1+\lambda_1^{-1}\nu)(1+\lambda_2^{-1}\nu)\}^{-1/2}.\end{aligned}$$

A.3 Prove $\boldsymbol{\beta}^{(d+1)}$ is the unique global maximizer of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$

Let

$$W(\boldsymbol{\beta}) = \sum_{i=1}^n \left[Z_{i\cdot} - \sum_{l=1}^L \left\{ \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i|\mathcal{D},\boldsymbol{\theta}^{(d)})\{\sum_{c=1}^n \sum_{j=1}^{K_c} \mathbf{E}(Z_{cjl}|\mathcal{D},\boldsymbol{\theta}^{(d)})\}b_l(t_{iK_i})}{\sum_{c=1}^n b_l(t_{cK_c})\exp(\boldsymbol{x}_c'\boldsymbol{\beta})\mathbf{E}(\phi_c|\mathcal{D},\boldsymbol{\theta}^{(d)})} \right\} \right] \boldsymbol{x}_i$$

denote the estimating function in step 1 of the proposed EM algorithm in Section 3.2. We now show that $\boldsymbol{\beta}^{(d+1)}$ is the unique solution of $W(\boldsymbol{\beta}) = 0$.

In the following, we write the conditional expectations $\mathbf{E}(\phi_i) = \mathbf{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})$ and $\mathbf{E}(Z_{ijl}) = \mathbf{E}(Z_{ijl}|\mathcal{D}, \boldsymbol{\theta}^{(d)})$ for each i, j, and l. Also, define
$\xi_l = \sum_{i=1}^n b_l(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i)$ and $\eta_l = \sum_{c=1}^n \sum_{j=1}^{K_c} \mathbf{E}(Z_{cjl})\}b_l(t_{iK_i})$ for $l = 1, \cdots, L$. Note that ξ_l's are all functions of $\boldsymbol{\beta}$ but η_l's are all constants. Taking partial derivative

of $W(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, one obtains

$$\begin{aligned} V(\boldsymbol{\beta}) &= \frac{\partial W(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ &= -\sum_{i=1}^{n}\sum_{l=1}^{L} \xi_l^{-1} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i) b_l(t_{iK_i})\boldsymbol{x}_i\boldsymbol{x}_i' \\ &\quad + \sum_{i=1}^{n}\sum_{l=1}^{L} \xi_l^{-2} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i) b_l(t_{iK_i}) \sum_{c=1}^{n} \exp(\boldsymbol{x}_c'\boldsymbol{\beta})\mathbf{E}(\phi_c) b_l(t_{cK_c})\boldsymbol{x}_i\boldsymbol{x}_c' \\ &= -W_1 + W_2 \end{aligned}$$

where

$$\begin{aligned} W_1 &= \sum_{i=1}^{n}\sum_{l=1}^{L} \xi_l^{-1} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i) b_l(t_{iK_i})\boldsymbol{x}_i\boldsymbol{x}_i' \\ &= \sum_{i=1}^{n}\sum_{l=1}^{L} \xi_l^{-2} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i) b_l(t_{iK_i})\boldsymbol{x}_i\boldsymbol{x}_i' \sum_{c=1}^{n} \exp(\boldsymbol{x}_c'\boldsymbol{\beta})\mathbf{E}(\phi_c) b_l(t_{cK_c}) \end{aligned}$$

and

$$W_2 = \sum_{i=1}^{n}\sum_{l=1}^{L} \xi_l^{-2} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i) b_l(t_{iK_i}) \sum_{c=1}^{n} \exp(\boldsymbol{x}_c'\boldsymbol{\beta})\mathbf{E}(\phi_c) b_l(t_{cK_c})\boldsymbol{x}_i\boldsymbol{x}_c'.$$

Notice that switching i and c does not change the values of W_1 and W_2. Thus, one can write

$$\begin{aligned} W_1 &= \sum_{i=1}^{n}\sum_{l=1}^{L}\sum_{c=1}^{n} \xi_l^{-2} \exp(\boldsymbol{x}_c'\boldsymbol{\beta}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_c)\mathbf{E}(\phi_i) b_l(t_{cK_c}) b_l(t_{iK_i})\boldsymbol{x}_i\boldsymbol{x}_i' \\ &= \frac{1}{2}\sum_{c=1}^{n}\sum_{l=1}^{L} \xi_l^{-2} \exp(\boldsymbol{x}_c'\boldsymbol{\beta}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_c)\mathbf{E}(\phi_i) b_l(t_{cK_c}) b_l(t_{iK_i})(\boldsymbol{x}_i\boldsymbol{x}_i' + \boldsymbol{x}_c\boldsymbol{x}_c') \end{aligned}$$

and

$$\begin{aligned} W_2 &= \sum_{i=1}^{n}\sum_{l=1}^{L}\sum_{c=1}^{n} \xi_l^{-2} \exp(\boldsymbol{x}_c'\boldsymbol{\beta})\mathbf{E}(\phi_c) b_l(t_{cK_c}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i) b_l(t_{iK_i})\boldsymbol{x}_i\boldsymbol{x}_c' \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{L}\sum_{c=1}^{n} \xi_l^{-2} \exp(\boldsymbol{x}_c'\boldsymbol{\beta})\mathbf{E}(\phi_c) b_l(t_{cK_c}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i) b_l(t_{iK_i})(\boldsymbol{x}_i\boldsymbol{x}_c' + \boldsymbol{x}_c\boldsymbol{x}_i'). \end{aligned}$$

Thus,

$$\begin{aligned} V(\boldsymbol{\beta}) &= -W_1 + W_2 \\ &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{L}\sum_{c=1}^{n} \exp(\boldsymbol{x}_c'\boldsymbol{\beta})\mathbf{E}(\phi_c) b_l(t_{cK_c}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i) b_l(t_{iK_i})(\boldsymbol{x}_i - \boldsymbol{x}_c)(\boldsymbol{x}_i - \boldsymbol{x}_c)'. \end{aligned}$$

Consider $\boldsymbol{z}'V(\boldsymbol{\beta})\boldsymbol{z}$ for any $\boldsymbol{z} \in R^p$, where p is the dimension of $\boldsymbol{\beta}$. One obtains

$$\begin{aligned}\boldsymbol{z}'V(\boldsymbol{\beta})\boldsymbol{z} &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{L}\sum_{c=1}^{n}\exp(\boldsymbol{x}_c'\boldsymbol{\beta})\mathbf{E}(\phi_c)b_l(t_{cK_c})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i)b_l(t_{iK_i})\boldsymbol{z}'(\boldsymbol{x}_i-\boldsymbol{x}_c)(\boldsymbol{x}_i-\boldsymbol{x}_c)'\boldsymbol{z} \\ &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{L}\sum_{c=1}^{n}\exp(\boldsymbol{x}_c'\boldsymbol{\beta})\mathbf{E}(\phi_c)b_l(t_{cK_c})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathbf{E}(\phi_i)b_l(t_{iK_i})\{\boldsymbol{z}'(\boldsymbol{x}_i-\boldsymbol{x}_c)\}^2.\end{aligned}$$

Since all the terms of $\exp(\boldsymbol{x}_c'\boldsymbol{\beta})$'s, $\mathbf{E}(\phi_c)$'s, $b_l(t_{cK_c})$'s are positive, $\boldsymbol{z}'V(\boldsymbol{\beta})\boldsymbol{z}$ is always non-positive. For nonzero $\boldsymbol{z}$, $\boldsymbol{z}'V(\boldsymbol{\beta})\boldsymbol{z}$ takes zero only when $\boldsymbol{z}'(\boldsymbol{x}_i - \boldsymbol{x}_c) = 0$ for all i and c, in which case all subjects have the same predictors. In that case, there is nonidentifiability between the regression parameters and the unspecified baseline mean function $\mu_0(\cdot)$. Thus, as long as the model is identifiable, $\boldsymbol{z}'V(\boldsymbol{\beta})\boldsymbol{z}$ is always negative for any nonzero $\boldsymbol{z}$. In other words, $V(\boldsymbol{\beta})$ is negative definite. This indicates that there is a unique solution to the estimating equation $W(\boldsymbol{\beta}) = 0$, and the solution is $\boldsymbol{\beta}^{(d+1)}$ following the definition of $\boldsymbol{\beta}^{(d+1)}$.

Comment: A modified version of this proof can show that $\boldsymbol{\theta}^{(d+1)}$ is actually the unique maximizer of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ by examining $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ directly instead of examining the estimating equation.

A.4 Numerical comparison of the proposed method to the approach of Hua et al. (2014)

Hua et al. (2014) studied the same model for panel count data as ours and proposed a two-stage estimation approach for inferences. Discussion about the difference between their approach and ours can be found in Section 6 of the paper. Below we provide the comparison of the two methods via simulations using the same simulation setups in Hua et al. (2014).

The general simulation setups are as follows according to Hua et al. (2014). (i) Suppose that six follow-up observations are scheduled at times $\boldsymbol{t}^\circ = \{t_j^\circ : t_j^\circ = 2j, j = 1, \ldots, 6\}$. The actual observation times t_{ij}° are generated from a normal distribution,

$N(t_j^\circ, 1/3)$, to allow for some deviation from the pre-scheduled times. Let $\xi_{ij} = 1_{[t_{ij-1}^\circ < t_{ij}^\circ]}$, for $i = 1, \ldots, 6$ and $t_{i0}^\circ = 0$. (ii) Let $\delta_{ij} = 1$ indicating the jth visit happens and zero otherwise with probability $P(\delta_{ij} = 1) = 1/(1 + e^{t_{ij}^\circ - 10})$, indicating the probability of missing visit increases as follow-up proceeds which is likely the case in many clinical follow-up studies. Hence the ith subject has $K_i = \sum_{j=1}^{6} \xi_{ij}\delta_{ij}$ observations at $\boldsymbol{t}_i = (t_{i1}, t_{i2}, \ldots, t_{iK_i})$. K_i's could be different from subject to subject. (iii) The covariate vector $\boldsymbol{x}_i = (x_{i1}, x_{i2}, x_{i3})$ is generated by $x_{i1} \sim$ Uniform $(0, 1)$, $x_{i2} \sim N(0, 1)$ and $x_{i3} \sim$ Bernoulli(0.5). The true values of regression parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ are taken to be $(-1, 0.5, 1.5)'$. (iv) Given the observed data $(\boldsymbol{x}_i, K_i, \boldsymbol{t}_i)$, the panel counts are generated using

$$\boldsymbol{N}_i = (N_i(t_{i1}), N_i(t_{i2}), \ldots, N_i(t_{iK_i}))$$

and the following three scenarios are considered for the counting process $\boldsymbol{N}_i(t)$.

Scenario 1. Data are generated from a gamma frailty nonhomogeneous Poisson process model. The gamma frailties $\phi_1, \phi_2, \ldots, \phi_n$ are randomly drawn from the Gamma distribution, $\mathcal{G}a(0.5, 0.5)$. Conditional on the gamma frailty ϕ_i and covariates $\boldsymbol{x}_i$, the panel counts associated with subject i are generated from the following model,

$$N_i(t_{ij}) - N_i(t_{ij-1})|\phi_i \sim \mathcal{P}[2\phi_i\{(t_{ij})^{1/2} - (t_{ij-1})^{1/2}\}\exp(\boldsymbol{x}_i'\boldsymbol{\beta})]. \tag{A.1}$$

This specifies the baseline mean function $\mu_0(t) = 2t^{1/2}$ and true frailty variance parameter $\nu = 0.5$ in our model.

Scenario 2. Data are generated from a lognormal frailty nonhomogeneous Poisson process model. The frailties ϕ_i's are generated from a lognormal distribution with mean 1 and variance 2. Conditional on the frailty ϕ_i, the panel counts associated with subject i are also drawn from a nonhomogeneous Poisson process as in A.1.

Scenario 3. Data are generated from a mixture Poisson distribution. Generate a discrete frailty term ϕ_i from $\{0.6, 1, 1.4\}$ with probabilities 0.25, 0.5 and 0.25, respectively. Given ϕ_i, the panel counts are generated from a nonhomogeneous Poisson process model,

$$N_i(t_{ij}) - N_i(t_{ij-1})|\phi_i \sim \mathcal{P}[(2+\phi_i)\{(t_{ij})^{1/2} - (t_{ij-1})^{1/2}\} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})].$$

In all scenarios, we generated 1000 data sets, each with 100 subjects. Note that Scenario 1 corresponds to the true case and the other two scenarios correspond to misspecified cases for our method and that in Hua et al. (2014). Table A.1 summarizes the simulation results for the regression parameters in all scenarios from the proposed method (GFNPMS) and the approach in Hua et al. (2014) method (HZT), respectively. The summary results include their bias(Bias), Monte Carlo standard deviation (MC-sd), the average of the estimated standard errors (ASE), and the 95% coverage probability (CP95) based on the Wald 95% confidence intervals. The summary results for HZT were taken directly from Tables 1∼3 in Hua et al. (2014) with some adjustment. We switched their summary results for $\hat{\beta}_1$ and $\hat{\beta}_2$ because there are substantial evidences that the authors may have misplaced their results for $\hat{\beta}_1$ and $\hat{\beta}_2$ in Hua et al. (2014), likely due to mislabeling x_1 and x_2 in their simulations. As seen from Table A.1 here, the two methods produced comparative results in all these setups.

Table A.1 Simulation results from the proposed method (GFNPMS) and the HZT approach in Hua et al. (2014). Bias is defined as the average of the point estimates minus the true value, MC-sd the Monte Carlo standard deviation of the point estimates, ASE the average of the estimated standard errors, and CP95 the 95% coverage probability based on the Wald confidence intervals.

		GFNPMS				HZT			
Scenario	Est	Bias	MC-sd	ASE	CP95	Bias	MC-sd	ASE	CP95
Scenario 1	$\hat{\beta}_1$	-0.0173	0.5630	0.5186	0.919	0.0016	0.5113	0.5755	0.941
	$\hat{\beta}_2$	-0.0029	0.1498	0.1544	0.952	0.0028	0.1527	0.1425	0.921
	$\hat{\beta}_3$	-0.0008	0.3042	0.2992	0.952	0.0029	0.3072	0.3017	0.916
Scenario 2	$\hat{\beta}_1$	0.0115	0.4830	0.3830	0.882	-0.0030	0.4833	0.4547	0.904
	$\hat{\beta}_2$	-0.0033	0.1413	0.1154	0.893	0.0074	0.1412	0.1211	0.913
	$\hat{\beta}_3$	0.0042	0.2776	0.2229	0.886	-0.0113	0.2783	0.2560	0.915
Scenario 3	$\hat{\beta}_1$	-0.0125	0.1666	0.1712	0.948	-0.0008	0.1375	0.1311	0.925
	$\hat{\beta}_2$	-0.0011	0.0474	0.0503	0.960	0.0019	0.0442	0.0402	0.922
	$\hat{\beta}_3$	0.0025	0.1074	0.1048	0.942	0.0013	0.0814	0.0788	0.945

A.5 Evidence of the robustness of the proposed approach to the knot placement

Tables A.2 and A.3 present the estimation results from the proposed method using different number of knots for the two data applications. From the results in these tables, it is clear that the proposed method is robust to the number of equally-spaced knots.

Table A.2 Results of the bladder tumor data analysis from the proposed method using different numbers (m) of equally-spaced interior knots. Summarized results include the point estimates and their corresponding standard errors (in parenthesis) for $\beta_1 \sim \beta_4$ and ν as well as the AIC value in each model.

m	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\nu}$	AIC
3	0.364	0.014	0.062	-1.146	0.356	1614.207
	(0.105)	(0.119)	(0.409)	(0.433)	(0.063)	—
4	0.356	0.016	0.039	-1.145	0.353	1598.330
	(0.105)	(0.119)	(0.410)	(0.434)	(0.062)	—
5	0.348	0.016	-0.003	-1.154	0.351	1589.554
	(0.105)	(0.119)	(0.411)	(0.435)	(0.062)	—
6	0.342	0.015	-0.031	-1.152	0.350	1578.834
	(0.106)	(0.120)	(0.411)	(0.436)	(0.061)	—
7	0.341	0.017	-0.029	-1.142	0.350	1576.821
	(0.106)	(0.120)	(0.410)	(0.435)	(0.062)	—
8	0.338	0.017	-0.026	-1.138	0.351	1571.158
	(0.106)	(0.120)	(0.409)	(0.434)	(0.062)	—
9	0.336	0.012	-0.033	-1.140	0.351	1564.837
	(0.106)	(0.120)	(0.409)	(0.435)	(0.062)	—
10	0.334	0.013	-0.028	-1.137	0.352	1564.231
	(0.106)	(0.120)	(0.409)	(0.434)	(0.062)	—
11	0.335	0.015	-0.034	-1.134	0.352	1568.283
	(0.106)	(0.120)	(0.408)	(0.434)	(0.062)	—

Table A.3 Results of the skin cancer data analysis from the proposed method using different numbers (m) of equally-spaced interior knots. Summarized results include the point estimates and their corresponding standard errors (in parenthesis) for $\beta_1 \sim \beta_4$ and ν as well as the AIC value in each model.

m	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\nu}$	AIC
3	-0.031	0.116	-0.0008	0.252	1.273	10454.39
	(0.143)	(0.015)	(0.0065)	(0.145)	(0.205)	—
4	-0.031	0.116	-0.0008	0.252	1.273	10457.54
	(0.143)	(0.015)	(0.0065)	(0.145)	(0.205)	—
5	-0.031	0.116	-0.0008	0.252	1.273	10464.82
	(0.143)	(0.015)	(0.0065)	(0.145)	(0.204)	—
6	-0.031	0.116	-0.0008	0.252	1.274	10461.64
	(0.143)	(0.015)	(0.0065)	(0.145)	(0.205)	—
7	-0.031	0.116	-0.0008	0.250	1.276	10470.23
	(0.143)	(0.015)	(0.0065)	(0.145)	(0.205)	—
8	-0.031	0.116	-0.0009	0.249	1.276	10493.35
	(0.143)	(0.015)	(0.0065)	(0.145)	(0.205)	—

Appendix B

Chapter 3 Appendix and Supplementary Materials

B.1 Proofs of Theorem 1 and Theorem 2

Proof of Theorem 1: Let $\boldsymbol{\theta}_j = (\rho_j, \beta_j, \Lambda_{0j})$ for $j = 1, 2$. Suppose $S(t|\boldsymbol{x}, \boldsymbol{\theta}_1) = S(t|\boldsymbol{x}, \boldsymbol{\theta}_2)$ for any $t \geq 0$ and $\boldsymbol{x}$. To show the identifiability, we need to prove $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. First, by taking $\boldsymbol{x}$ to be 0 and 1 respectively in the equation $S(t|\boldsymbol{x}, \boldsymbol{\theta}_1) = S(t|\boldsymbol{x}, \boldsymbol{\theta}_2)$, one obtains

$$\{1 + \rho_1 \Lambda_{01}(t)\}^{-\rho_1^{-1}} = \{1 + \rho_2 \Lambda_{02}(t)\}^{-\rho_2^{-1}}$$

and

$$\{1 + \rho_1 \Lambda_{01}(t) \exp(\beta_1)\}^{-\rho_1^{-1}} = \{1 + \rho_2 \Lambda_{02}(t) \exp(\beta_2)\}^{-\rho_2^{-1}} \tag{B.1}$$

for any $t \geq 0$. The first equation leads to (3.2)when there are no covariates. Plugging (3.2) into (B.1), one obtains

$$\{1 + \rho_1 \Lambda_{01}(t) \exp(\beta_1)\}^{-\rho_1^{-1}} = \left(1 + \exp(\beta_2)[\{1 + \rho_1 \Lambda_{01}(t)\}^{\rho_2 \rho_1^{-1}} - 1]\right)^{-\rho_2^{-1}} \tag{B.2}$$

for any $t \geq 0$. Recall $\Lambda_{01}(t) \to \infty$ as $t \to \infty$. For large t, the left side of (B.2) has the order of $\rho_1^{-\rho_1^{-1}} \{\Lambda_{01}(t)\}^{-\rho_1^{-1}} \exp(-\rho_1^{-1}\beta_1)$ and the right side has the order of $\rho_1^{-\rho_1^{-1}} \{\Lambda_{01}(t)\}^{-\rho_1^{-1}} \exp(-\rho_2^{-1}\beta_2)$. This leads to $\rho_1^{-1}\beta_1 = \rho_2^{-1}\beta_2$. On the other hand, taking logarithm at both sides of equation (B.2) yields

$$\rho_1^{-1} \log\{1 + \rho_1 \Lambda_{01}(t) \exp(\beta_1)\} = \rho_2^{-1} \log\left(1 + \exp(\beta_2)[\{1 + \rho_1 \Lambda_{01}(t)\}^{\rho_2 \rho_1^{-1}} - 1]\right) \tag{B.3}$$

for all $t > 0$. Further taking derivative of both sides of (B.3) with respect to t, we have

$$\frac{\lambda_{01}(t)\exp(\beta_1)}{1+\rho_1\Lambda_{01}(t)\exp(\beta_1)} = \frac{\lambda_{01}(t)\exp(\beta_2)\{1+\rho_1\Lambda_{01}(t)\}^{\rho_2\rho_1^{-1}-1}}{1+\exp(\beta_2)[\{1+\rho_1\Lambda_{01}(t)\}^{\rho_2\rho_1^{-1}}-1]} \tag{B.4}$$

for all $t > 0$ possibly except a zero probability set with countable values. Letting $t \to 0$ in the equation (B.4), one obtains $\exp(\beta_1) = \exp(\beta_2)$. Thus, we have $\beta_1 = \beta_2$. Since we already have $\rho_1^{-1}\beta_1 = \rho_2^{-1}\beta_2$, we get $\rho_1 = \rho_2$. In this case, equation (3.2) reduces to $\Lambda_{02}(t) = \Lambda_{01}(t)$ for all $t \geq 0$. This suggests the GORH models are identifiable in the regression setting of containing only a binary covariate.

Proof of Theorem 2: Without losing generality, assume that 0 and 1 are both possible values for all covariates. This is reasonable and feasible to achieve through reparameterization after standardizing continuous covariates and adopting multiple binary covariates for categorical variables. Let p denote the number of covariates and $\boldsymbol{e_h}$ denote the vector with 1 for the hth element and 0 for other elements, for $h = 1, \ldots, p$. Let $\boldsymbol{\theta}_j = (\rho_j, \boldsymbol{\beta}_j, \Lambda_{0j})$, $j = 1, 2$, be two sets of parameters under the GORH models defined in (3.1). Taking $\boldsymbol{x}$ equal to $\mathbf{0}$ and $\boldsymbol{e_h}$ in the equation of $S(t|\boldsymbol{x}, \boldsymbol{\theta}_1) = S(t|\boldsymbol{x}, \boldsymbol{\theta}_2)$ and using the same techniques as in the proof of Theorem 1, we obtain $\beta_{1h} = \beta_{2h}$ for all $h = 1, \ldots, p$ and further $\rho_1 = \rho_2$ and $\Lambda_{01} = \Lambda_{02}$.

B.2 Formula of the quantities involved in var($\widehat{\boldsymbol{\theta}}$)

For notation simplicity, we use E(Y) to denote the conditional expection of a general quantity Y conditional on the observed data $\mathcal{D}$. Similarly, the covariances and variances below refer to the conditional covariances and variances given the observed data $\mathcal{D}$. In Louis's method, the quantities involved in $\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ have the

following expressions,

$$\begin{aligned}
\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' &= -\sum_{i=1}^{n} \Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{E}(\phi_i)\boldsymbol{x}_i\boldsymbol{x}_i', \\
\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})/\partial\boldsymbol{\beta}\partial\gamma_l &= -\sum_{i=1}^{n} b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{E}(\phi_i|\mathcal{D})\boldsymbol{x}_i, \\
\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})/\partial\gamma_l^2 &= -\frac{1}{\gamma_l^2}\sum_{i=1}^{n} \mathrm{E}(Z_{il}|\mathcal{D}), \quad \text{for each} \quad l, \\
\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})/\partial\gamma_l\partial\gamma_l' &= 0 \ \text{ for each } \quad l \neq l',
\end{aligned}$$

and the quantities in $\mathrm{var}(\partial \log \mathcal{L}_c/\partial\boldsymbol{\theta})$ are given by the following expressions,

$$\begin{aligned}
\mathrm{cov}\Big(\frac{\partial \log \mathcal{L}_c}{\partial\boldsymbol{\beta}}, \frac{\partial \log \mathcal{L}_c}{\partial\boldsymbol{\beta}}\Big) &= \sum_{i=1}^{n} \Big[\mathrm{var}(Z_i) - 2\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{cov}(Z_i, \phi_i) \\
&\quad + \Big\{\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\Big\}^2 \mathrm{var}(\phi_i)\Big]\boldsymbol{x}_i\boldsymbol{x}_i', \\
\mathrm{cov}\Big(\frac{\partial \log \mathcal{L}_c}{\partial\boldsymbol{\beta}}, \frac{\partial \log \mathcal{L}_c}{\partial\gamma_l}\Big) &= \sum_{i=1}^{n} \Big\{\gamma_l^{-1}\mathrm{cov}(Z_i, Z_{il}) - b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{cov}(Z_i, \phi_i)\Big\}\boldsymbol{x}_i \\
&\quad + \sum_{i=1}^{n} \Big\{b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{var}(\phi_i) \\
&\quad - \gamma_l^{-1}\mathrm{cov}(Z_{il}, \phi_i)\Big\}\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_i, \\
\mathrm{cov}\Big(\frac{\partial \log \mathcal{L}_c}{\partial\gamma_l}, \frac{\partial \log \mathcal{L}_c}{\partial\gamma_l}\Big) &= \sum_{i=1}^{n} \Big\{b_l^2(C_i)e^{2\boldsymbol{x}_i'\boldsymbol{\beta}}\mathrm{var}(\phi_i) - 2\gamma_l^{-1}b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{cov}(Z_{il}, \phi_i) \\
&\quad + \gamma_l^{-2}\mathrm{var}(Z_{il})\Big\}, \\
\mathrm{cov}\Big(\frac{\partial \log \mathcal{L}_c}{\partial\gamma_l}, \frac{\partial \log \mathcal{L}_c}{\partial\gamma_{l'}}\Big) &= \sum_{i=1}^{n} \Big\{(\gamma_l\gamma_{l'})^{-1}\mathrm{cov}(Z_{il}, Z_{il'}) - \gamma_{l'}^{-1}b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{cov}(\phi_i, Z_{il'}) \\
&\quad - \gamma_l^{-1}b_{l'}(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{cov}(\phi_i, Z_{il}) \\
&\quad + b_l(C_i)b_{l'}(C_i)\exp(2\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{var}(\phi_i)\Big\}, \ l \neq l'.
\end{aligned}$$

The conditional variance and covariance terms have the following forms,

$$\begin{aligned}
\mathrm{var}(\phi_i) &= \frac{\rho}{\{1+\rho\Lambda_0(t)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^2} I\{\delta_i = 0\} \\
&\quad + \Bigg[\frac{1+\rho}{\{1+\rho\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^2} \frac{\{1+\rho\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^{\rho^{-1}+2}-1}{\{1+\rho\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^{\rho^{-1}}-1} \\
&\quad - \mathrm{E}^2(\phi_i) \Bigg] I\{\delta_i = 1\}, \\
\mathrm{var}(Z_i) &= \mathrm{E}(Z_i)\Big\{1+(1+\rho)\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}) - \mathrm{E}(Z_i)\Big\}, \\
\mathrm{var}(Z_{il}) &= \Big\{\gamma_l b_l(C_i)\Lambda_0^{-1}(C_i)\Big\}^2\{\mathrm{var}(Z_i) - \mathrm{E}(Z_i)\} + \gamma_l b_l(C_i)\Lambda_0^{-1}(C_i)\mathrm{E}(Z_i), \\
\mathrm{cov}(Z_i, \phi_i) &= \{(1+\rho) - \mathrm{E}(\phi_i)\}\mathrm{E}(Z_i), \\
\mathrm{cov}(Z_i, Z_{il}) &= \gamma_l b_l(C_i)\Lambda_0^{-1}(C_i)\{\mathrm{var}(Z_i) + \mathrm{E}^2(Z_i)\} - \mathrm{E}(Z_i)\mathrm{E}(Z_{il}), \\
\mathrm{cov}(Z_{il}, \phi_i) &= \{(1+\rho) - \mathrm{E}(\phi_i)\}\mathrm{E}(Z_{il}), \\
\mathrm{cov}(Z_{il}, Z_{il'}) &= \gamma_l\gamma_{l'} b_l(C_i) b_{l'}(C_i)\Lambda_0^{-2}(C_i)\{\mathrm{var}(Z_i) - \mathrm{E}(Z_i)\}, \quad l \neq l',
\end{aligned}$$

where the expressions of $\mathrm{E}(\phi_i)$'s, $\mathrm{E}(Z_i)$'s, and $\mathrm{E}(Z_{il})$'s are given in Section 3.4.

B.3 Prove $\boldsymbol{\beta}^{(d+1)}$ is the unique global maximizer of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$

An EM algorithm involves two steps, the expectation step (E-step) and the maximization step (M-step). In the E-step of the algorithm one takes the expectation of log-complete likelihood $\log L_c(\boldsymbol{\theta})$ with respect to all the latent variables including ϕ_i's, $Z_i's$ and $Z_{il}'s$ conditional on the observed data $\mathcal{D}$ and the current parameter $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)}, \boldsymbol{\gamma}^{(d)})$ at dth iteration. It yields that $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = \mathrm{E}\log L_c(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)}) + H_2(\boldsymbol{\theta}^{(d)})$, where

$$\begin{aligned}
H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)}) =& \sum_{i=1}^{n} \mathrm{E}(Z_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})\boldsymbol{x}_i'\boldsymbol{\beta} - \sum_{i=1}^{n}\Lambda_0(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{E}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) \qquad \text{(B.5)} \\
&+ \sum_{i=1}^{n}\sum_{l=1}^{L}\mathrm{E}(Z_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)})\log(\gamma_l),
\end{aligned}$$

and $H_2(\boldsymbol{\theta}^{(d)})$ is a function of $\boldsymbol{\theta}^{(d)}$ but free of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$.

Setting $\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})/\partial \gamma_l = 0$, one obtains that the solution of γ_l is a function of $\boldsymbol{\beta}$ in the following form,

$$\gamma_l(\boldsymbol{\beta}) = \frac{\sum_{i=1}^n \mathrm{E}(Z_{il})}{\sum_{i=1}^n b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{E}(\phi_i)}.$$

So it means that we only need to show that $\boldsymbol{\beta}^{(d+1)}$ is a unique global maximizer. We plug the $\gamma_l(\boldsymbol{\beta})$ into $H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})$ so that $H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})$ can be rewritten as

$$\begin{aligned}
&\widetilde{H}_1(\boldsymbol{\beta}, \boldsymbol{\theta}^{(d)}) \\
&= \sum_{i=1}^n \mathrm{E}(Z_i)\boldsymbol{x}_i'\boldsymbol{\beta} - \sum_{i=1}^n \sum_{l=1}^L \left\{ \frac{\sum_{i=1}^n \mathrm{E}(Z_{il})}{\sum_{i=1}^n b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{E}(\phi_i)} \right\} b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{E}(\phi_i) \\
&+ \sum_{i=1}^n \sum_{l=1}^L \mathrm{E}(Z_{il}) \log \left\{ \frac{\sum_{i=1}^n \mathrm{E}(Z_{il})}{\sum_{i=1}^n b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{E}(\phi_i)} \right\},
\end{aligned}$$

Now our goal is to show that Hessian matrix $\partial^2 \widetilde{H}_1(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'$ is positive definite for all $\boldsymbol{\beta}$ so that the solution of $H_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})$ is unique global maximizer. For notational convenience we define $d_{il} = E(Z_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)})$, $d_l = \sum_{i=1}^n d_{il} = \sum_{i=1}^n E(Z_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)})$, $g_{il} = b_l(C_i)\mathrm{E}(\phi_i)$ and $f_l = \sum_{i=1}^n b_l(C_i)\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mathrm{E}(\phi_i) = \sum_{i=1}^n g_{il}\exp(\boldsymbol{x}_i'\boldsymbol{\beta})$ for all i and l. Using the defined notations $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ can be expressed as

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = \sum_{i=1}^n \mathrm{E}(Z_i)\boldsymbol{x}_i'\boldsymbol{\beta} - \sum_{i=1}^n \sum_{l=1}^L \left(\frac{d_l}{f_l}\right) g_{il}\exp(\boldsymbol{x}_i'\boldsymbol{\beta}) + \sum_{i=1}^n \sum_{l=1}^L d_{il}\log\left(\frac{d_l}{f_l}\right) + H_2(\boldsymbol{\theta}^{(d)}),$$

where $H_2(\boldsymbol{\theta}^{(d)})$ is free of parameters. Note that f_l is a function of $\boldsymbol{\beta}$, but d_{il}, d_l and g_{il} are not free of $\boldsymbol{\beta}$. The first partial derivative of $H_1(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ is

$$\begin{aligned}
\frac{\partial \widetilde{H}_1(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \sum_{l=1}^L d_{il}\boldsymbol{x}_i + \sum_{i=1}^n \sum_{l=1}^L \left(\frac{d_l}{f_l^2}\right) g_{il}\exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \left(\sum_{j=1}^n g_{jl}\exp(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_j\right) \\
&\quad - \sum_{i=1}^n \sum_{l=1}^L \left(\frac{d_l}{f_l}\right) g_{il}\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_i - \sum_{i=1}^n \sum_{l=1}^L \left(\frac{d_{il}}{f_l}\right) \left(\sum_{j=1}^n g_{jl}\exp(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_j\right).
\end{aligned}$$

The two terms in the middle of the above expression are canceled out because

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{l=1}^{L} d_l f_l^{-1} g_{il} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_i &= \sum_{i=1}^{n}\sum_{l=1}^{L} d_l f_l^{-2} g_{il} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_i \left(\sum_{j=1}^{n} g_{jl} \exp(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_j\right) \\
&= \sum_{i=1}^{n}\sum_{l=1}^{L}\sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \exp(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_i \\
&= \sum_{l=1}^{L}\sum_{i=1}^{n}\sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \exp(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_i \\
&= \sum_{l=1}^{L}\sum_{j=1}^{n}\sum_{i=1}^{n} d_l f_l^{-2} g_{jl} g_{il} \exp(\boldsymbol{x}_j'\boldsymbol{\beta}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_j \\
&= \sum_{l=1}^{L}\sum_{i=1}^{n} d_l f_l^{-2} g_{il} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \left(\sum_{j=1}^{n} g_{jl} \exp(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_j\right)
\end{aligned}$$

So one obtains that

$$\begin{aligned}
\frac{\partial \widetilde{H}_1(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n}\sum_{l=1}^{L} d_{il}\boldsymbol{x}_i - \sum_{i=1}^{n}\sum_{l=1}^{L} \left(\frac{d_{il}}{f_l}\right) \left(\sum_{j=1}^{n} g_{jl} \exp(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_j\right) \\
&= \sum_{i=1}^{n}\sum_{l=1}^{L} d_{il}\boldsymbol{x}_i - \sum_{i=1}^{n}\sum_{l=1}^{L} d_{il} f_l^{-1} \left(\sum_{j=1}^{n} g_{jl} \exp(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_j\right) \\
&= \sum_{l=1}^{L}\sum_{i=1}^{n} d_{il}\boldsymbol{x}_i - \sum_{l=1}^{L} d_l f_l^{-1} \left(\sum_{j=1}^{n} g_{jl} \exp(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_j\right) \\
&= \sum_{l=1}^{L}\sum_{i=1}^{n} d_{il}\boldsymbol{x}_i - \sum_{l=1}^{L}\sum_{i=1}^{n} d_l f_l^{-1} g_{il} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\boldsymbol{x}_i
\end{aligned}$$

The second order partial derivative of $\widetilde{H}_1(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$,

$$\begin{aligned}
\frac{\partial^2 \widetilde{H}_1(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= \sum_{l=1}^{L} \sum_{i=1}^{n} d_l f_l^{-2} g_{il} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \boldsymbol{x}_i \left(\sum_{j=1}^{n} g_{jl} \exp(\boldsymbol{x}_j' \boldsymbol{\beta}) \boldsymbol{x}_j' \right) \\
&\quad - \sum_{l=1}^{L} \sum_{i=1}^{n} d_l f_l^{-1} g_{il} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \boldsymbol{x}_i \boldsymbol{x}_i' \\
&= \sum_{l=1}^{L} \sum_{i=1}^{n} \sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \exp(\boldsymbol{x}_j' \boldsymbol{\beta}) \boldsymbol{x}_i \boldsymbol{x}_j' \\
&\quad - \sum_{l=1}^{L} \sum_{i=1}^{n} d_l f_l^{-2} g_{il} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \boldsymbol{x}_i \boldsymbol{x}_i' \left(\sum_{j=1}^{n} g_{jl} \exp(\boldsymbol{x}_j' \boldsymbol{\beta}) \right) \\
&= \sum_{l=1}^{L} \sum_{i=1}^{n} \sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \exp(\boldsymbol{x}_j' \boldsymbol{\beta}) \boldsymbol{x}_i \boldsymbol{x}_j' \\
&\quad - \sum_{l=1}^{L} \sum_{i=1}^{n} \sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \exp(\boldsymbol{x}_j' \boldsymbol{\beta}) \boldsymbol{x}_i \boldsymbol{x}_i'
\end{aligned}$$

Rewrite (1) and (2) as follows,

$$\begin{aligned}
(1) &= \sum_{l=1}^{L} \sum_{i=1}^{n} \sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \exp(\boldsymbol{x}_j' \boldsymbol{\beta}) \boldsymbol{x}_i \boldsymbol{x}_j' \\
&= \sum_{l=1}^{L} \sum_{j=1}^{n} \sum_{i=1}^{n} d_l f_l^{-2} g_{jl} g_{il} \exp(\boldsymbol{x}_j' \boldsymbol{\beta}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \boldsymbol{x}_j \boldsymbol{x}_i' \\
&= \frac{1}{2} \sum_{l=1}^{L} \sum_{i=1}^{n} \sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \exp(\boldsymbol{x}_j' \boldsymbol{\beta}) \boldsymbol{x}_i \boldsymbol{x}_j' \\
&\quad + \frac{1}{2} \sum_{l=1}^{L} \sum_{j=1}^{n} \sum_{i=1}^{n} d_l f_l^{-2} g_{jl} g_{il} \exp(\boldsymbol{x}_j' \boldsymbol{\beta}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \boldsymbol{x}_j \boldsymbol{x}_i' \\
&= \frac{1}{2} \sum_{l=1}^{L} \sum_{i=1}^{n} \sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \exp(\boldsymbol{x}_j' \boldsymbol{\beta}) (\boldsymbol{x}_i \boldsymbol{x}_j' + \boldsymbol{x}_j \boldsymbol{x}_i')
\end{aligned}$$

Similarly, one can obtain that

$$(2) = -\frac{1}{2} \sum_{l=1}^{L} \sum_{i=1}^{n} \sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \exp(\boldsymbol{x}_j' \boldsymbol{\beta}) (\boldsymbol{x}_i \boldsymbol{x}_i' + \boldsymbol{x}_j \boldsymbol{x}_j').$$

Thus, the second order partial derivative of $H_1(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ can be expressed

as

$$\begin{aligned}
\frac{\partial^2 \widetilde{H}_1(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} &= \frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n}\sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \exp(\boldsymbol{x}_j'\boldsymbol{\beta})(\boldsymbol{x}_i\boldsymbol{x}_j' + \boldsymbol{x}_j\boldsymbol{x}_i') \\
&\quad -\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n}\sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \exp(\boldsymbol{x}_j'\boldsymbol{\beta})(\boldsymbol{x}_i\boldsymbol{x}_i' + \boldsymbol{x}_j\boldsymbol{x}_j') \\
&= -\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n}\sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \exp(\boldsymbol{x}_j'\boldsymbol{\beta}) \\
&\quad (\boldsymbol{x}_i\boldsymbol{x}_i' + \boldsymbol{x}_j\boldsymbol{x}_j' - \boldsymbol{x}_i\boldsymbol{x}_j' - \boldsymbol{x}_j\boldsymbol{x}_i') \\
&= -\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n}\sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \exp(\boldsymbol{x}_j'\boldsymbol{\beta})(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)'
\end{aligned}$$

The Hessian matrix $\partial^2\widetilde{H}_1(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'$ is negative definite if and only if $\boldsymbol{z}'\{\partial^2\widetilde{H}_1(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'\}z < 0$ where $\boldsymbol{z} \in R^p$.

$$\begin{aligned}
\boldsymbol{z}'\frac{\partial^2 \widetilde{H}_1(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\boldsymbol{z} &= -\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n}\sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \exp(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{z}'(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)'\boldsymbol{z} \\
&= -\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n}\sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \exp(\boldsymbol{x}_j'\boldsymbol{\beta})\{\boldsymbol{z}'(\boldsymbol{x}_i - \boldsymbol{x}_j)\}^2 \\
&= -\frac{1}{2}\sum_{l=1}^{L}\sum_{i=1}^{n}\sum_{j=1}^{n} d_l f_l^{-2} g_{il} g_{jl} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \exp(\boldsymbol{x}_j'\boldsymbol{\beta})\{\sum_{s=1}^{p} \boldsymbol{z}_s'(\boldsymbol{x}_{is} - \boldsymbol{x}_{js})\}^2
\end{aligned}$$

Since d_l, f_l, g_{il} and $\exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, for all i and j are positive, $\partial^2 H_1(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'$ is non-positive definite. Note that $\boldsymbol{z}$, $\boldsymbol{z}'\{\partial^2\widetilde{H}_1(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'\}z = 0$ only when $\boldsymbol{z}'(\boldsymbol{x}_i - \boldsymbol{x}_j) = 0$ for all i and j, which is only possible when all subjects have the same value for a particular predictor. In that case, the corresponding regression parameters and cumulative baseline hazard function $\Lambda_0(t)$ are not identifiable. In other words, as long as the model is identifiable and proper, we have $\boldsymbol{z}'\{\partial^2\widetilde{H}_1(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'\}z < 0$ for all $\boldsymbol{z} \in R^p/\{\mathbf{0}\}$. Therefore $\partial^2\widetilde{H}_1(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'$ is negative definite. This guarantees that $\boldsymbol{\beta}^{(d+1)}$ is the unique global maximizer of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$.

Appendix C

Chapter 4 Appendix and Supplementary Materials

C.1 Formula of the quantities involved in var$(\widehat{\boldsymbol{\theta}})$

The detailed expressions of the quantities involved $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ and $\text{var}\{\partial \log L_c(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\}$ are presented below in order to obtain the variance estimate of $\hat{\boldsymbol{\theta}}$. So the first part is the second partial derivative of $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$.

$$
\begin{aligned}
\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= -\sum_{i=1}^{n} \mu_0(t_{iK_i}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \mathrm{E}(\phi_i | \mathcal{D}, \hat{\boldsymbol{\theta}}) \boldsymbol{x}_i \boldsymbol{x}_i', \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} &= -\sum_{i=1}^{n} \{(\delta_{i1} + \delta_{i2}) \Lambda_0(R_i) + \delta_{i3} \Lambda_0(L_i)\} \exp(\tilde{\boldsymbol{x}}_i' \boldsymbol{\alpha}) \mathrm{E}(\phi_i | \mathcal{D}, \hat{\boldsymbol{\theta}}) \tilde{\boldsymbol{x}}_i \tilde{\boldsymbol{x}}_i', \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \gamma_l} &= -\sum_{i=1}^{n} b_l(t_{iK_i}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \mathrm{E}(\phi_i | \mathcal{D}, \hat{\boldsymbol{\theta}}) \boldsymbol{x}_i, \forall l, \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha} \partial \tilde{\gamma}_m} &= -\sum_{i=1}^{n} \{(\delta_{i1} + \delta_{i2}) I_m(R_i) + \delta_{i3} I_m(L_i)\} \exp(\tilde{\boldsymbol{x}}_i' \boldsymbol{\alpha}) \mathrm{E}(\phi_i | \mathcal{D}, \hat{\boldsymbol{\theta}}) \tilde{\boldsymbol{x}}_i \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \gamma_l^2} &= -\sum_{i=1}^{n} \sum_{j=1}^{K_i} \frac{\mathrm{E}(Z_{ijl} | \mathcal{D}, \hat{\boldsymbol{\theta}})}{\gamma_l^2}, \forall l, \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \tilde{\gamma}_m^2} &= -\sum_{i=1}^{n} \frac{\mathrm{E}(W_{im1} | \mathcal{D}, \hat{\boldsymbol{\theta}}) + (\delta_{i2} + \delta_{i3}) \mathrm{E}(W_{im2} | \mathcal{D}, \hat{\boldsymbol{\theta}})}{\tilde{\gamma}_m^2}, \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \nu^2} &= n\{\nu^{-1} - \psi'(\nu)\},
\end{aligned}
$$

$\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})/\partial \boldsymbol{\beta} \partial \nu = 0$, $\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})/\partial \boldsymbol{\beta} \partial \tilde{\gamma}_m = 0$, $\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})/\partial \boldsymbol{\alpha} \partial \nu = 0$,

$\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})/\partial \boldsymbol{\alpha} \partial \gamma_l = 0$, $\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})/\partial \nu \partial \gamma_l = 0$, $\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})/\partial \nu \partial \tilde{\gamma}_m = 0$,

$\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})/\partial \gamma_l \partial \gamma_{l'} = 0$, $\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})/\partial \tilde{\gamma}_m \partial \tilde{\gamma}_{m'} = 0$, for $l \neq l'$ and $m \neq m'$. $\psi(\cdot)$ and $\psi'(\cdot)$

are digamma function and trigamma function, respectively.

Next we are going to derive the necessary quantities involved in $\mathrm{var}\{\partial \log L_c(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\}$. To make all the formulas clear to present, we define the following notations. Define $\tau_i = \{(\delta_{i1}+\delta_{i2})\Lambda_0(R_i) + \delta_{i3}\Lambda_0(L_i)\}\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})$, $\tau_{im} = \{(\delta_{i1}+\delta_{i2})I_m(R_i) + \delta_{i3}I_m(L_i)\}\exp(\tilde{\boldsymbol{x}}_i'\boldsymbol{\alpha})$. Define $p_{im} = \{\Lambda_0(R_i)\}^{-1}\tilde{\gamma}_m I_m(R_i)$ and $q_{im} = \{\Lambda_0(R_i) - \Lambda_0(L_i)\}^{-1}\tilde{\gamma}_m\{I_m(R_i) - I_m(L_i)\}$, then

$$
\begin{aligned}
\mathrm{cov}\left(\frac{\partial \log \mathcal{L}_c}{\partial \boldsymbol{\beta}}, \frac{\partial \log \mathcal{L}_c}{\partial \boldsymbol{\beta}}\right) &= \sum_{i=1}^{n}\{\mu_0(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^2 \mathrm{var}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\boldsymbol{x}_i\boldsymbol{x}_i', \\
\mathrm{cov}\left(\frac{\partial \log \mathcal{L}_c}{\partial \boldsymbol{\alpha}}, \frac{\partial \log \mathcal{L}_c}{\partial \boldsymbol{\alpha}}\right) &= \sum_{i=1}^{n}\Big[\mathrm{var}(W_{i1}|\mathcal{D},\hat{\boldsymbol{\theta}}) + \delta_{i2}\mathrm{var}(W_{i2}|\mathcal{D},\hat{\boldsymbol{\theta}}) + \tau_i^2\mathrm{var}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}}) \\
&\quad -2\tau_i\{\mathrm{cov}(W_{i1},\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}}) + \delta_{i2}\mathrm{cov}(W_{i2},\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\}\Big]\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_i', \\
\mathrm{cov}\left(\frac{\partial \log \mathcal{L}_c}{\partial \boldsymbol{\beta}}, \frac{\partial \log \mathcal{L}_c}{\partial \gamma_l}\right) &= \sum_{i=1}^{n}\mu_0(t_{iK_i})b_l(t_{iK_i})\{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^2\mathrm{var}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\boldsymbol{x}_i, \\
\mathrm{cov}\left(\frac{\partial \log \mathcal{L}_c}{\partial \boldsymbol{\alpha}}, \frac{\partial \log \mathcal{L}_c}{\partial \tilde{\gamma}_m}\right) &= \sum_{i=1}^{n}\Big[\tilde{\gamma}_m^{-1}\{\mathrm{cov}(W_{i1},W_{im1}|\mathcal{D},\hat{\boldsymbol{\theta}}) + \delta_{i2}\mathrm{cov}(W_{i2},W_{im2}|\mathcal{D},\hat{\boldsymbol{\theta}})\} \\
&\quad -\tau_{im}\{\mathrm{cov}(W_{i1},\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}}) + \delta_{i2}\mathrm{cov}(W_{i2},\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\} \\
&\quad -\tilde{\gamma}_m^{-1}\tau_i\{\mathrm{cov}(W_{im1},\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}}) + \delta_{i2}\mathrm{cov}(W_{im2},\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\} \\
&\quad +\tau_i\tau_{im}\mathrm{var}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\Big]\tilde{\boldsymbol{x}}_i, \\
\mathrm{cov}\left(\frac{\partial \log \mathcal{L}_c}{\partial \boldsymbol{\beta}}, \frac{\partial \log \mathcal{L}_c}{\partial \nu}\right) &= \sum_{i=1}^{n}\mu_0(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \\
&\quad \times\big\{\mathrm{var}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}}) - \mathrm{cov}(\phi_i,\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}})\big\}\boldsymbol{x}_i, \\
\mathrm{cov}\left(\frac{\partial \log \mathcal{L}_c}{\partial \boldsymbol{\alpha}}, \frac{\partial \log \mathcal{L}_c}{\partial \nu}\right) &= \sum_{i=1}^{n}\Big[\mathrm{cov}(W_{i1},\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}) - \mathrm{cov}(W_{i1},\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}}) \\
&\quad +\delta_{i2}\{\mathrm{cov}(W_{i2},\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}) - \mathrm{cov}(W_{i2},\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\} \\
&\quad +\tau_i\{\mathrm{var}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}}) - \mathrm{cov}(\phi_i,\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}})\}\Big]\tilde{\boldsymbol{x}}_i, \\
\mathrm{cov}\left(\frac{\partial \log \mathcal{L}_c}{\partial \gamma_l}, \frac{\partial \log \mathcal{L}_c}{\partial \nu}\right) &= \sum_{i=1}^{n}b_l(t_{iK_i})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\Big\{\mathrm{var}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}}) - \mathrm{cov}(\phi_i,\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}})\Big\},
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{cov}\left(\frac{\partial \log \mathcal{L}_c}{\partial \tilde{\gamma}_m}, \frac{\partial \log \mathcal{L}_c}{\partial \nu}\right) &= \sum_{i=1}^{n} \Big[\tilde{\gamma}_m^{-1}\Big\{\operatorname{cov}(W_{im1}, \log(\phi)|\mathcal{D}, \hat{\boldsymbol{\theta}}) - \operatorname{cov}(W_{im1}, \phi_i|\mathcal{D}, \hat{\boldsymbol{\theta}}) \\
&\quad +\delta_{i2}\{\operatorname{cov}(W_{im2}, \log(\phi_i)|\mathcal{D}, \hat{\boldsymbol{\theta}}) - \operatorname{cov}(W_{im2}, \phi_i|\mathcal{D}, \hat{\boldsymbol{\theta}})\}\Big\} \\
&\quad +\tau_{im}\Big\{\operatorname{var}(\phi_i|\mathcal{D}, \hat{\boldsymbol{\theta}}) - \operatorname{cov}(\phi_i, \log(\phi_i)|\mathcal{D}, \hat{\boldsymbol{\theta}})\Big\}\Big], \\
\operatorname{cov}\left(\frac{\partial \log \mathcal{L}_c}{\partial \gamma_l}, \frac{\partial \log \mathcal{L}_c}{\partial \gamma_l}\right) &= \sum_{i=1}^{n} \{b_l(t_{iK_i}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^2 \operatorname{var}(\phi_i|\mathcal{D}, \hat{\boldsymbol{\theta}}) \\
&\quad +\frac{1}{\gamma_l^2} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \operatorname{var}(Z_{ijl}|\mathcal{D}, \hat{\boldsymbol{\theta}}), \\
\operatorname{cov}\left(\frac{\partial \log \mathcal{L}_c}{\partial \gamma_l}, \frac{\partial \log \mathcal{L}_c}{\partial \gamma_{l'}}\right) &= \sum_{i=1}^{n} b_l(t_{iK_i}) b_{l'}(t_{iK_i}) \{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^2 \operatorname{var}(\phi_i|\mathcal{D}, \hat{\boldsymbol{\theta}}) \\
&\quad +(\gamma_l \gamma_{l'})^{-1} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \operatorname{cov}(Z_{ijl}, Z_{ijl'}|\mathcal{D}, \hat{\boldsymbol{\theta}}), \\
\operatorname{cov}\left(\frac{\partial \log \mathcal{L}_c}{\partial \tilde{\gamma}_m}, \frac{\partial \log \mathcal{L}_c}{\partial \tilde{\gamma}_{m'}}\right) &= \sum_{i=1}^{n} \Big[(\tilde{\gamma}_m \tilde{\gamma}_{m'})^{-1}\{\operatorname{cov}(W_{im1}, W_{im'1}) + \delta_{i2}\operatorname{cov}(W_{im2}, W_{im'2})\} \\
&\quad -\tilde{\gamma}_m^{-1}\tau_{im'}\{\operatorname{cov}(W_{im1}, \phi_i|\mathcal{D}, \hat{\boldsymbol{\theta}}) + \delta_{i2}\operatorname{cov}(W_{im2}, \phi_i|\mathcal{D}, \hat{\boldsymbol{\theta}})\} \\
&\quad -\tilde{\gamma}_{m'}^{-1}\tau_{im}\{\operatorname{cov}(W_{im'1}, \phi_i|\mathcal{D}, \hat{\boldsymbol{\theta}}) + \delta_{i2}\operatorname{cov}(W_{im'2}, \phi_i|\mathcal{D}, \hat{\boldsymbol{\theta}}) \\
&\quad +\tau_{im}\tau_{im'}\operatorname{var}(\phi_i|\mathcal{D}, \hat{\boldsymbol{\theta}})\}\Big], \\
\operatorname{cov}\left(\frac{\partial \log \mathcal{L}_c}{\partial \nu}, \frac{\partial \log \mathcal{L}_c}{\partial \nu}\right) &= \sum_{i=1}^{n} \operatorname{var}(\log(\phi_i)|\mathcal{D}, \hat{\boldsymbol{\theta}}) - 2\sum_{i=1}^{n} \operatorname{cov}(\phi_i, \log(\phi_i)|\mathcal{D}, \hat{\boldsymbol{\theta}}) \\
&\quad +\sum_{i=1}^{n} \operatorname{var}(\phi_i|\mathcal{D}, \hat{\boldsymbol{\theta}})
\end{aligned}
$$

where the conditional variance and covariance terms have the following form,

$$
\begin{aligned}
\mathrm{var}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}}) &= \left[\frac{a_i(a_i+1)}{b_i^2}\frac{1-\left(\frac{b_i}{b_i+c_i}\right)^{a_i+2}}{1-\left(\frac{b_i}{b_i+c_i}\right)^{a_i}} - \mathrm{E}^2(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\right]\delta_{i1} \\
&\quad + \left[\frac{a_i(a_i+1)}{(b_i+d_i)^2}\frac{1-\left(\frac{b_i+d_i}{b_i+c_i}\right)^{a_i+2}}{1-\left(\frac{b_i+d_i}{b_i+c_i}\right)^{a}} - \mathrm{E}^2(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\right]\delta_{i2} + \frac{a_i}{(b_i+d_i)^2}\delta_{i3}, \\
\mathrm{var}\{\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}\} &= \Bigg[\psi'(a_i) + \{\psi(a_i)-\log(b_i+c_i)\}^2 \\
&\quad + \frac{\{\psi(a_i)-\log(b_i)\}^2 - \{\psi(a_i)-\log(b_i+c_i)\}^2}{1-\left(\frac{b_i}{b_i+c_i}\right)^{a_i}} \\
&\quad - \mathrm{E}^2\{\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}\}\Bigg]\delta_{i1} + \Bigg[\psi'(a_i) + \{\psi(a_i)-\log(b_i+c_i)\}^2 \\
&\quad + \frac{\{\psi(a_i)-\log(b_i+d_i)\}^2 - \{\psi(a_i)-\log(b_i+c_i)\}^2}{1-\left(\frac{b_i+d_i}{b_i+c_i}\right)^{a_i}} \\
&\quad - \mathrm{E}^2\{\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}\}\Bigg]\delta_{i2} \\
&\quad + \big[\psi'(a_i) + \{\psi(a_i)-\log(b_i+d_i)\}^2 - \mathrm{E}\{\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}\}\big]\delta_{i3}, \\
\mathrm{cov}(\phi_i,\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}) &= \Bigg[\frac{a_i}{b_i}\times\frac{\psi(a_i+1)-\log(b_i)-\left(\frac{b_i}{b_i+c_i}\right)^{a_i+1}\{\psi(a_i+1)-\log(b_i+c_i)\}}{1-\left(\frac{b_i}{b_i+c_i}\right)^{a_i}} \\
&\quad - \mathrm{E}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\mathrm{E}\{\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}\}\Bigg]\delta_{i1} + \Bigg[\frac{a_i}{b_i+d_i} \\
&\quad \times\frac{\psi(a_i+1)-\log(b_i+d_i)-\left(\frac{b_i+d_i}{b_i+c_i}\right)^{a_i+1}\{\psi(a_i+1)-\log(b_i+c_i)\}}{1-\left(\frac{b_i+d_i}{b_i+c_i}\right)^{a_i}} \\
&\quad - \mathrm{E}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\mathrm{E}\{\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}\}\Bigg]\delta_{i2} \\
&\quad + \left[\frac{a_i}{b_i+d_i}\{\psi(a_i+1)-\log(b_i+d_i)\} - \mathrm{E}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\mathrm{E}\{\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}\}\right]\delta_{i3}, \\
\mathrm{var}(W_{i1}|\mathcal{D},\hat{\boldsymbol{\theta}}) &= \left[\frac{a_ib_ic_i + a_i(a_i+1)c_i^2}{b^2}\left\{1-\left(\frac{b_i}{b_i+c_i}\right)^{a_i}\right\}^{-1} - \mathrm{E}^2(W_{i1}|\mathcal{D},\hat{\boldsymbol{\theta}})\right]\delta_{i1},
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(W_{i2}|\mathcal{D},\hat{\boldsymbol{\theta}}) &= \left[\frac{a_i(b_i+d_i)(c_i-d_i)-a_i(a_i+1)(c_i-d_i)^2}{(b_i+d_i)^2}\right. \\
&\quad \left.\times\left\{1-\left(\frac{b_i+d_i}{b_i+c_i}\right)^{a_i}\right\}^{-1}-\mathrm{E}^2(W_{i2}|\mathcal{D},\hat{\boldsymbol{\theta}})\right]\delta_{i2}, \\
\mathrm{cov}(W_{i1},\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}}) &= \left[\frac{a_i(a_i+1)c_i}{b_i^2}\left\{1-\left(\frac{b_i}{b_i+c_i}\right)^{a_i}\right\}^{-1}\right. \\
&\quad \left.-\mathrm{E}(W_{i1}|\mathcal{D},\hat{\boldsymbol{\theta}})\mathrm{E}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\right]\delta_{i1}, \\
\mathrm{cov}(W_{i2},\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}}) &= \left[\frac{a_i(a_i+1)(c_i-d_i)}{(b_i+d_i)^2}\left\{1-\left(\frac{b_i+d_i}{b_i+c_i}\right)^{a_i}\right\}^{-1}\right. \\
&\quad \left.-\mathrm{E}(W_{i2}|\mathcal{D},\hat{\boldsymbol{\theta}})\mathrm{E}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\right]\delta_{i2}, \\
\mathrm{cov}(W_{i1},\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}) &= \left[\frac{a_ic_i\{\psi(a_i+1)-\log(b_i)\}}{b_i}\left\{1-\left(\frac{b_i}{b_i+c_i}\right)^{a_i}\right\}^{-1}\right. \\
&\quad \left.-\mathrm{E}(W_{i1}|\mathcal{D},\hat{\boldsymbol{\theta}})\mathrm{E}\{\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}\}\right]\delta_{i1}, \\
\mathrm{cov}(W_{i2},\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}) &= \left[\frac{a_i(c_i-d_i)\{\psi(a_i+1)-\log(b_i+d_i)\}}{b_i+d_i}\left\{1-\left(\frac{b_i+d_i}{b_i+c_i}\right)^{a_i}\right\}^{-1}\right. \\
&\quad \left.-\mathrm{E}(W_{i2}|\mathcal{D},\hat{\boldsymbol{\theta}})\mathrm{E}\{\log(\phi_i)|\mathcal{D},\hat{\boldsymbol{\theta}}\}\right]\delta_{i2}, \\
\mathrm{cov}(W_{i1},W_{im1}|\mathcal{D},\hat{\boldsymbol{\theta}}) &= \left[p_{im}\left\{\mathrm{var}(W_{i1}|\mathcal{D},\hat{\boldsymbol{\theta}})+\mathrm{E}^2(W_{i1}|\mathcal{D},\hat{\boldsymbol{\theta}})\right\}\right. \\
&\quad \left.-\mathrm{E}(W_{i1}|\mathcal{D},\hat{\boldsymbol{\theta}})\mathrm{E}(W_{im1}|\mathcal{D},\hat{\boldsymbol{\theta}})\right]\delta_{i1}, \\
\mathrm{cov}(W_{i2},W_{im2}|\mathcal{D},\hat{\boldsymbol{\theta}}) &= \left[q_{im}\left\{\mathrm{var}(W_{i2}|\mathcal{D},\hat{\boldsymbol{\theta}})+\mathrm{E}^2(W_{i2}|\mathcal{D},\hat{\boldsymbol{\theta}})\right\}\right. \\
&\quad \left.-\mathrm{E}(W_{i2}|\mathcal{D},\hat{\boldsymbol{\theta}})\mathrm{E}(W_{im2}|\mathcal{D},\hat{\boldsymbol{\theta}})\right]\delta_{i2}, \\
\mathrm{cov}(W_{im1},\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}}) &= \left[\frac{p_{im}a_i(a_i+1)c_i}{b_i^2}\left\{1-\left(\frac{b_i}{b_i+c_i}\right)^{a_i}\right\}^{-1}\right. \\
&\quad \left.-\mathrm{E}(W_{im1}|\mathcal{D},\hat{\boldsymbol{\theta}})\mathrm{E}(\phi_i|\mathcal{D},\hat{\boldsymbol{\theta}})\right]\delta_{i1},
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{cov}(W_{im2}, \phi_i | \mathcal{D}, \hat{\boldsymbol{\theta}}) &= \left[\frac{q_{im} a_i (a_i + 1)(c_i - d_i)}{(b_i + d_i)^2} \left\{ 1 - \left(\frac{b_i + d_i}{b_i + c_i} \right)^{a_i} \right\}^{-1} \right. \\
&\quad \left. - \mathrm{E}(W_{im2} | \mathcal{D}, \hat{\boldsymbol{\theta}}) \mathrm{E}(\phi_i | \mathcal{D}, \hat{\boldsymbol{\theta}}) \right] \delta_{i2}, \\
\operatorname{cov}(W_{im1}, \log(\phi_i) | \mathcal{D}, \hat{\boldsymbol{\theta}}) &= \left[\frac{p_{im} a_i c_i \{ \psi(a_i + 1) - \log(b_i) \}}{b_i} \left\{ 1 - \left(\frac{b_i}{b_i + c_i} \right)^{a_i} \right\}^{-1} \right. \\
&\quad \left. - \mathrm{E}(W_{im1} | \mathcal{D}, \hat{\boldsymbol{\theta}}) \mathrm{E}\{ \log(\phi_i) | \mathcal{D}, \hat{\boldsymbol{\theta}} \} \right] \delta_{i1}, \\
\operatorname{cov}(W_{im2}, \log(\phi_i) | \mathcal{D}, \hat{\boldsymbol{\theta}}) &= \left[\frac{q_{im} a_i (c_i - d_i) \{ \psi(a_i + 1) - \log(b_i + d_i) \}}{b_i + d_i} \right. \\
&\quad \times \left\{ 1 - \left(\frac{b_i + d_i}{b_i + c_i} \right)^{a_i} \right\}^{-1} \\
&\quad \left. - \mathrm{E}(W_{im2} | \mathcal{D}, \hat{\boldsymbol{\theta}}) \mathrm{E}\{ \log(\phi_i) | \mathcal{D}, \hat{\boldsymbol{\theta}} \} \right] \delta_{i2}, \\
\operatorname{cov}(W_{im1}, W_{im'1} | \mathcal{D}, \hat{\boldsymbol{\theta}}) &= p_{im} p_{im'} \{ \operatorname{var}(W_{i1} | \mathcal{D}, \hat{\boldsymbol{\theta}}) - \mathrm{E}(W_{i1} | \mathcal{D}, \hat{\boldsymbol{\theta}}) \} \delta_{i1}, \quad m \neq m', \\
\operatorname{cov}(W_{im2}, W_{im'2} | \mathcal{D}, \hat{\boldsymbol{\theta}}) &= q_{im} q_{im'} \{ \operatorname{var}(W_{i2} | \mathcal{D}, \hat{\boldsymbol{\theta}}) - \mathrm{E}(W_{i2} | \mathcal{D}, \hat{\boldsymbol{\theta}}) \} \delta_{i2}, \quad m \neq m', \\
\operatorname{cov}(Z_{ijl}, Z_{ijl'} | \mathcal{D}, \hat{\boldsymbol{\theta}}) &= - \frac{\gamma_l \gamma_{l'} \{ b_l(t_{ij}) - b_l(t_{ij-1}) \} \{ b_{l'}(t_{ij}) - b_{l'}(t_{ij-1}) \}}{\{ \mu_0(t_{ij}) - \mu_0(t_{ij-1}) \}^2} Z_{ij}, \\
\operatorname{var}(W_{im1} | \mathcal{D}, \hat{\boldsymbol{\theta}}) &= \{ p_{im}(1 - p_{im}) \mathrm{E}(W_{i1} | \mathcal{D}, \hat{\boldsymbol{\theta}}) + p_{im}^2 \operatorname{var}(W_{i1} | \mathcal{D}, \hat{\boldsymbol{\theta}}) \} \delta_{i1}, \\
\operatorname{var}(W_{im2} | \mathcal{D}, \hat{\boldsymbol{\theta}}) &= \{ q_{im}(1 - q_{im}) \mathrm{E}(W_{i2} | \mathcal{D}, \hat{\boldsymbol{\theta}}) + q_{im}^2 \operatorname{var}(W_{i2} | \mathcal{D}, \hat{\boldsymbol{\theta}}) \} \delta_{i2}.
\end{aligned}
$$