STRATIFIED SEMIPARAMETRIC REGRESSION ANALYSIS OF PARTLY INTERVAL CENSORED FAILURE TIME DATA WITH MISSING AND MIS-MEASURED LONGITUDINAL COVARIATES

by

Gang Cheng

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2024

Approved by:

____

Dr. Yanqing Sun and Dr. Qingning Zhou

____

Dr. Yinghao Pan

____

Dr. Asis Nasipuri

ABSTRACT

GANG CHENG. Stratified Semiparametric Regression Analysis of Partly Interval Censored Failure Time Data with Missing and Mis-Measured Longitudinal Covariates. (Under the direction of DR. YANQING SUN AND DR. QINGNING ZHOU)

Survival analysis has important applications across various fields, including medicine, finance, actuarial science, and social studies. In the modeling process, we often encounter challenges related to censored data, where the exact event times are not directly observable. Instead, we only know that the events occurred within specific time intervals. Also, the covariates in the model may subject to missingness and mis-measurement. In this dissertation, we investigate (partly) interval censored data with missing and mis-measured covariates under semi-parametric models.

In the first project, we proposed an inverse probability weighting (IPW) estimator for transformation models with interval-censored data and missing covariates. To estimate the model parameters, we developed a combined approach that integrates the EM algorithm with inverse probability weighting. Additionally, we introduced a variance estimation procedure using weighted bootstrap. We demonstrated that the proposed estimator is consistent and asymptotically normal through theoretical justification and numerical simulations. Finally, we applied our approach to data from the HVTN 703/704 HIV clinical trial.

In the second project, we extended the approach from the first project to accommodate covariates subject to both missing data and measurement error. We employed a measurement error induced hazard approach to construct the baseline hazard function. To estimate the true covariates from the mis-measured covariates, we utilized a linear mixed-effects model. For model estimation, we applied a method similar to that used in the first project. Extensive simulations demonstrated that the resulting estimator is consistent and asymptotically normal. Finally, we applied the developed method to data from the HVTN 703/704 HIV clinical trial, accounting for measurement error in the covariate log(VRC).

ACKNOWLEDGMENTS

First, I would like to express my deepest gratitude to my dissertation advisors, Dr. Yanqing Sun and Dr. Qingning Zhou, for their exceptional guidance, patience, and encouragement throughout my Ph.D. journey at UNC Charlotte. Their expertise in biostatistics was invaluable in helping me navigate the challenges and complexities of my dissertation research. Without their unwavering support, the completion of this dissertation would not have been possible. Their passion for research and tireless dedication have profoundly inspired me, motivating me to pursue my career goals. Their optimism has also left a lasting impact, teaching me to approach challenges with positivity and perseverance.

To my committee members, Drs. Yinghao Pan and Asis Nasipuri, I extend my heartfelt gratitude for your constructive feedback and invaluable suggestions throughout this process. I would also like to thank all the faculty and staff in the Mathematics and Statistics Department for their support. A special note of appreciation goes to Dr. Shaozhong Deng, whose guidance on course selection and academic planning has been instrumental in shaping my career at UNC Charlotte.

Finally, I would like to take this opportunity to express my deepest gratitude to my parents, Zhu Cheng and Shifeng Hou, for their unwavering support and unconditional love. This achievement would not have been possible without their encouragement and sacrifices.

Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

Survival analysis is a statistical method employed to analyze time-to-event data, commonly used in medical, epidemiological, and social sciences research. It revolves around understanding the duration until a particular event of interest occurs, such as death, failure of a machine, or relapse of a disease. Unlike traditional statistical methods that focus solely on observing if an event occurs, survival analysis also considers the time it takes for an event to happen, accommodating censoring (where some individuals do not experience the event within the study period).

1.1 Survival Data and Examples

Survival data, also known as failure time data, refers to observations that track the time until a particular event of interest occurs, such as machine failure, disease relapse, or death. However, due to various factors such as loss to follow-up or study termination, complete observations of survival time are not always feasible. These incomplete observations, termed censored failure times, pose a significant challenge in survival analysis, where the precise event occurrence is not fully known for all participants. Censored failure data includes left censoring, right censoring, and interval censoring, representing situations where the precise event time is only partially known. Left censoring pertains to events occurring before the study commencement with unknown exact times, as observed in cancer studies where some patients decease prior to the study's initiation. Interval censoring is observed when the event time falls within a specific range, and it is a common scenario in HIV antibody trials. In such trials, antibody concentrations are measured periodically, and instances of HIV occurrences are known to happen between these scheduled examinations. Right censoring indicates that the event time occurs after a certain observation period due to reasons like the study's conclusion or participant dropout before the study's end. Left truncation in survival analysis refers to a scenario where individuals with event times earlier than the start of the study are excluded from analysis. This exclusion occurs because the observation of events is limited to those occurring after the study has commenced. Left truncation can lead to biased estimates of survival probabilities since individuals who have already experienced the event of interest before the study began are not included, resulting in an underestimation of the true survival probabilities. Effectively handling censorship or truncation in are crucial for accurate data analysis and meaningful statistical interpretation, particularly in the context of survival analysis.

1.2 Hazard Functions and Survival Models

We usually study survival times through survival function and hazard function. Let T be the survival time. The survival function $S(t)$ represents the unconditional probability that the event time occurs after a specified time t. On the other hand, the hazard function $\lambda(t)$ models the instantaneous rate at which the event occurs at time t, provided that the individual has survived up to time t. The hazard function $\lambda(t)$ is modeled as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T \leq t + \Delta t | T \geq t)$$

Survival analysis is widely applied across numerous fields, including medical research, epidemiology, biostatistics, engineering, social sciences, and market research. In medical research, it's instrumental in investigating patient outcomes, disease progression, and treatment effectiveness. The purpose of many survival analysis are studying effects of covariates on the hazard function. Let $Z = (Z_1, Z_2, ..., Z_p)$ be the $p-$dimensional covariates. We define the conditional hazard function of T at t given covariate $Z = z$ defined as:

$$\lambda(t|Z = z) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T \leq t + \Delta t | T \geq t, Z = z)$$

An established and fundamental tool in survival analysis is the Cox Proportional Hazards Model (Cox (1972)), with a regression model given by

$$\lambda(t|Z) = \lambda_0(t) \exp(\beta^\top Z) \tag{1.1}$$

where β is a $p-$dimensional vector of regression parameters, and $\lambda_0(t)$ represents an arbitrary and unspecified baseline hazard function. The Cox model assumes proportional hazards across individuals over time, and a consistent estimator of β can be obtained by maximizing the partial likelihood (Cox (1972)). Another framework for analyzing failure time data is the transformation model Zeng and Lin (2006), Zeng et al. (2016), Zhou et al. (2021) which applies transformation function to the cumulative hazard function. It takes the form

$$\Lambda_j(t|X(\cdot), Z) = G\left[\int_0^t \exp\{\beta^\top X(s) + \gamma^\top Z\} d\Lambda_j(s)\right] \tag{1.2}$$

where β, γ are p, q dimension unknown regression parameters, $\Lambda(t) = \int_0^t \lambda(s)ds$ is an increasing function with $\Lambda(0) = 0$. $G(\cdot)$ is a pre-specified transformation function that is strictly increasing and three times continuously differentiable with $G(0) = 0$, $G^{'}(0) > 0$ and $G(\infty) = \infty$. For the choices of G, the class of frailty-induced transformation models which take the form:

$$G(x) = -\log \int_0^\infty \exp(-x\xi) f(\xi) d\xi \tag{1.3}$$

where $f(\xi)$ is the density function of a non-negative random variable ξ with support $[0, \infty)$. If ξ follows gamma distribution with mean equals to one and variance equals to r, then $G(x) = \frac{\log(1+rx)}{r}(r \geq 0)$ with $G(x) = x$ corresponds to $r = 0$. By treating ξ as missing, the frailty induced transformation is extremely useful in developing EM algorithm (Zeng and Lin (2006), Zeng et al. (2016), Zhou et al. (2021)). The class of Box-Cox transformations $G(x) = \frac{(1+x)^\rho - 1}{\rho}$ can be obtained from the positive stable distribution with parameter $0 < \rho < 1$. The function $G(x) = \log(1 + x)$ can be considered as a member of the Box-Cox transformations with $\rho = 0$. Transformation models have several advantages which include: (i) flexible to handle non-linear relationships between the covariates and the failure time; (ii) can include time varying covariates and handle non-proportional hazards. Furthermore, when X is time independent, the equation (1.2) can take the form

$$\log \Lambda(t) = -(\beta^\top X + \gamma^\top Z) + \log G^{-1}[-\log \epsilon_0]$$

where ϵ_0 follows a uniform distribution (Zeng and Lin (2006)). Specifically, $G(x) = x$ yield proportional hazards model and $G(x) = x$ yield proportional odds model. Besides the usual transformation model, we proposed a stratified transformation model, which further flexes the traditional transformation model by allowing different baseline cumulative hazard for different stratum. It takes the form:

$$\Lambda_j(t|X(\cdot), Z) = G\left[\int_0^t \exp\{\beta^\top X(s) + \gamma^\top Z\} d\Lambda_j(s)\right] \tag{1.4}$$

where $j = 1, \cdots, J$ and J represents number of stratums in the data.

1.3 Missing Data and Measurement Errors

In many scenarios, obtaining complete covariates information for all study subjects is not feasible, which results in missing data. Missing data problems are common in the survival data analysis which includes missing covariates, missing failure causes. One straightforward approach of handling missing data problem is complete-case (CC) analysis, which use only observations with full information. Complete-case analysis (CC) may lead to biased or misleading results when the missingness of data depends on observed data, but not on unobserved data (termed as missing at random (MAR), Rubin (1976)). An example is when we estimating proportional hazards model (1.1), the estimator of β will be biased if we simply use complete case analysis. In order to get reliable estimator of β, we need to adjust bias from complete case analysis. One approach of handling the MAR problem is weighting complete cases by the inverse probability weight Horvitz and Thompson (1952), which is commonly used to correct the bias introduced by missing at random Rubin (1976). By applying the inverse probability weighting, the complete cases are enlarged to represent the missing data. Besides for missing data, the covariates measured with error is another issue when we do model estimating. In the AIDS Clinical Trial Group (ACGT) 175 clinical trial on HIV-infected patients (Hammer et al. (1996)), the effects of baseline CD4 cell (zidovudine alone, zidovudine + didanosine, zidovudine + zalcitabine, didanosine alone) on time to the incidence of AIDS are of interest. Measurements of the baseline CD4 counts are subject to measurement error because of instrumental contamination and biological variation Song and Ma (2008). Using the contaminated covariates directly, termed as naive method, will lead to biased estimation results (Sun et al. (2023), Tsiatis and Davidian (2001)).

This dissertation addresses the challenge posed by the interplay between the semi-parametric transformation model and missing covariates and covariates subject to measurement error. We propose a flexible method capable of accommodating the semi-parametric transformation model and various missing data patterns and covariate measurement errors. The effectiveness of the model is validated through extensive simulations and theoretical justification. Subsequently, this method is applied in pertinent research on HIV prevention. This dissertation is organized as follows. Chapter 2 introduces stratified transformation models with missing covariates (covariates measured accurately). Chapter 3 describes stratified transformation models with missing covariates and covariates measurement errors. Tables, lists, figures and theorem proofs are relegate in chapter 4, 5.

CHAPTER 2: THE STRATIFIED SEMIPARAMETRIC TRANSFORMATION MODELS WITH MISSING COVARIATES

2.1 Introduction and Literature Review

Epidemiological and biomedical studies frequently yield failure time data subject to various forms of censoring. For instance, in clinical trials for AIDS among HIV-infected individuals, researchers typically track the time until onset of AIDS. The onset of AIDS is usually determined at periodic scheduled clinic visits. Consequently, the exact time of AIDS onset is known only within an interval defined by the last visit without AIDS criteria and the first visit where AIDS-defining conditions are observed, resulting in interval-censored failure time data. Left-censored observations arise if a subject presents with AIDS-defining conditions at their initial clinic visit, while right-censored observations occur if a subject never develops AIDS-defining conditions up to their final visit. Many studies follow a similar data structure, where blood tests are conducted during regular visits, and the event of interest occurs between the last negative test and the first positive one. We analyze failure time data encompassing a mixture of exact and censored observations, including left, interval, and/or right-censored data. Such datasets are commonly referred to as partly interval-censored data in scientific literature. Our research is inspired by the HVTN-703/HVTN-704 trials, a randomized trial comparing HIV-infected individuals across four regions (Corey et al. (2021)). Participants in these trials underwent regular measurements of HIV antibody VRC concentration every 4 weeks and were monitored for AIDS infection occurrence over an 86-week period. The study aims to investigate the association between age, time-varying HIV antibody VRC concentrations, and HIV infection onset. However, the data present challenges: left-censored observations occur if HIV infection predates the study, interval-censored observations arise when infection occurs between consecutive examinations, and right-censored observations signify cases where infection hasn't occurred by the study's end. Additionally, complicating matters, only a portion of participants have VRC measurements available, introducing potential estimation bias due to missing data. Thus, the analysis must address the complexities of both censored observations and missing covariate data, posing significant analytical challenges.

Extensive research has been conducted on interval censored and partly interval-censored data across various models. Among these, the simplest and most studied type is referred to as case-1 or current-status data, involving only one monitoring time per subject. When there are either two or k monitoring times per subject, the resulting data are termed as case-2 or case-k interval censoring (Huang and Wellner (1997). Schick and Yu (2000)) proposed mixed-case interval censoring, accommodating varying numbers of monitoring times among subjects. Finkelstein (1986), Huang (1996) studied maximum likelihood estimator (MLE) for the proportional hazards model with interval censored data. Zeng et al. (2016) proposed an efficient algorithm using Expectation-Maximization (EM) for estimating transformation models with interval-censored data. utilized an EM algorithm to estimate semiparametric transformation models in the presence of interval-censored data and missing covariates. Zhang et al. (2010) introduced a spline based sieve semiparametric maximum likelihood method to estimate the proportional hazards with interval censored data. Kim (2003) studied the maximum likelihood estimation for the proportional hazards model with partly interval censored data using generalized Gauss-Seidel algorithm with midpoint imputation for the interval censored data. Gao et al. (2017) studied generalized Buckley-James estimator for partly interval censored data with failure time under accelerate failure time model. Gao et al. (2019) proposed an EM algorithm for the partly interval censored data under the asymptomatic disease and symptomatic disease and random effects. There are also extensive literature about missing covariate with censored data. Chen (2001) used a local averaging of the observed covariates approach to provide an effective and unified approach to analysis a class of sampling designs with proportional hazards model. Qi et al. (2005) studied the weighted estimators for proportional hazards model with missing using IPW and AIPW for the right censored data. Breslow and Wellner (2007), Saegusa and Wellner (2013) developed the asymptotic properties of weighted likelihood estimators under two phase sampling design. However, methodologies for fitting transformation models to (partly) interval data with missing covariates, along with corresponding asymptotic properties, remain relatively unexplored. This chapter presents a comprehensive class of transformation models that integrate multi-baseline hazard functions and time-varying coefficients. These models provide significant flexibility and encompass special cases, including those discussed in previous works such as Zeng et al. (2016) and Zhou et al. (2021). To mitigate bias introduced by missing data under the Missing at Random (MAR) assumption, we employ inverse probability weighting.

The rest of this chapter is organized as follows: Section 2.2 introduces the data structure, models, and model assumptions. Section 2.3 presents the weighted EM algorithm for partly interval censored failure times data with liear transformation function. Section 2.4 introduces a weighted bootstrap variance estimating procedure. The proposed methods' finite-sample performance is assessed through simulation studies in Section 2.6. In Section 2.7, we present the results of applying our proposed model to the HVTN-703/704 data. Finally, Section 2.8 provides concluding remarks for chapter 2.

2.2 Data, Model and Assumption

Let T_{ji} denote the failure time of interest for the i-th subject in the j-th stratum, where $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$. Here, J represents the total number of strata, and n_j is the number of subjects in the j-th stratum. Let $X_{ji}(\cdot)$ and $Z_{ji}(\cdot)$ be vectors of possibly time-dependent covariates for the i-th subject in the j-th stratum. We consider partly interval-censored failure time data, which include exact, left-, interval-, and/or right-censored observations. Let Δ_{1ji} indicate whether the failure time T_{ji} is observed exactly: $\Delta_{1ji} = 1$ if T_{ji} is observed exactly, and $\Delta_{1ji} = 0$ otherwise. Let Δ_{2ji} indicate whether the observation is strictly interval-censored or left-censored. Let $(L_{ji}, R_{ji}]$ denote the smallest observed interval that includes T_{ji}, where L_{ji} is the last monitoring time at which the failure has not yet occurred, and R_{ji} is the first monitoring time at which the failure has occurred. If $L_{ji} < R_{ji} < \infty$, then $\Delta_{2ji} = 1$ and $T_{ji} \in (L_{ji}, R_{ji}]$. If $R_{ji} = \infty$, then $\Delta_{1ji} = \Delta_{2ji} = 0$. In addition to the partly interval-censored data, we consider a study design where covariates are collected in two phases. In the first phase, covariates $Z_{ji}(\cdot)$ are collected for all study subjects. In the second phase, covariates $X_{ji}(\cdot)$ are collected only for a subset of the study group. Let η_{ji} be the selection indicator for the second-phase covariates: $\eta_{ji} = 1$ if the observation is selected into the second phase (i.e., both $X_{ji}(\cdot)$ and $Z_{ji}(\cdot)$ are available), and $\eta_{ji} = 0$ otherwise (i.e., $X_{ji}(\cdot)$ is not available). We assume that the missing data mechanism depends only on the first-phase data $Z_{ji}(\cdot)$, Δ_{1ji}, Δ_{2ji}, L_{ji}, R_{ji}, and T_{ji}, but not on the missing variable $X_{ji}(\cdot)$. This is termed the missing-at-random (MAR) assumption Rubin (1976), which can be expressed as:

$$X_{ji}(\cdot) \perp\!\!\!\perp \eta_{ji} \mid Z_{ji}(\cdot), \Delta_{1ji}, \Delta_{2ji}, L_{ji}, R_{ji}, T_{ji}.$$

The failure times that are left/interval-censored and/or exactly observed are considered as the cases. Our data have the following structure, let

$$O_{ji} = (\Delta_{1ji}, \Delta_{2ji}, \Delta_{1ji}T_{ji}, (1-\Delta_{1ji})L_{ji}, (1-\Delta_{1ji})R_{ji}, \eta_{ji}, \eta_{ji}X_{ji}(\cdot), Z_{ji}(\cdot)) \tag{2.1}$$

be the observed data for the i-th subject in the j-th stratum, where $i = 1, ..., n_j$ and $j = 1, ..., J$. Under the stratified semiparametric transformation model, we assume that the cumulative hazard function for T_{ji} conditional on $X_{ji}(\cdot)$ and $Z_{ji}(\cdot)$ takes the form

$$\Lambda_j(t|X_{ji}(\cdot), Z_{ji}(\cdot)) = G\bigg(\int_0^t \exp\{\beta^\top X_{ji}(s) + \gamma^\top Z_{ji}(s)\} d\Lambda_j(s)\bigg) \tag{2.2}$$

where β and γ are vectors of unknown regression coefficients and $\Lambda_j(\cdot)$ is an unknown increasing function. $G(\cdot)$ is a pre-specified transformation function that is strictly increasing and three times continuously differentiable with $G(0) = 0$, $G^{'}(0) > 0$ and $G(\infty) = \infty$ (Zeng and Lin (2006)). $G^{'}(x)$ denotes $\frac{dG(x)}{dx}$. The choice of $G(x) = x$ yield the proportional hazards model while $G(x) = \log(1+x)$ yield the proportional odds model. In order to investigate interval censored data, Zeng et al. (2016) considered taking frailty induced hazard transformation model

$$G(x) = -\log \int_0^\infty \exp(-x\xi) f(\xi) d\xi \tag{2.3}$$

where $f(\xi)$ is the density function of a non-negative random variable ξ with support $[0, \infty)$. The gamma density of ξ with unit mean and variance r yields the logarithmic transformations $G(x) = \frac{\log(1+rx)}{r}(r \geq 0)$ with $r = 0$ corresponding to $G(x) = x$ and $r = 1$ corresponding to $G(x) = \log(1+x)$. Zeng and Lin (2006), Zeng et al. (2016), Zeng et al. (2017) devised EM-type algorithms using frailty induced transformations by treating ξ as missing.

Suppose the event time T_{ji} is monitored at a sequence of positive time-points $U_{ji,1} < \cdots < U_{ji,M_{ji}}$, we assume that $\{U_{jik} : j = 1, \cdots, J; i = 1, \cdots n_j; k = 1, \cdots M_{ji}\}$ are independent of $\{T_{ji} : j = 1, \cdots J, i = 1, \cdots J_i\}$. Let $(L_{ji}, R_{ji}]$ be the shortest time interval that contain T_{ji} which means $L_{ji} = \max\{U_{jik} : U_{jik} < T_{ji}, k = 0, \cdots, M_{ji}\}$ and $R_{ji} = \min\{U_{jik} : U_{jik} \geq T_{ji}, k = 1, \cdots, M_{ji} + 1\}$ where $U_{ji0} = 0$ and $U_{ji,m_{j_i}+1} = \infty$. Then the full data likelihood function for (β, γ) and $\Lambda = (\Lambda_1, \cdots, \Lambda_J)$ is

$$
\begin{aligned}
L_n(\beta,\gamma,\Lambda) = \prod_{j=1}^{J}\prod_{i=1}^{n_j} \Bigg[& \Lambda_j\{T_{ji}\} e^{\beta^\top X_{ji}(T_{ji}) + \gamma^\top Z_{ji}(T_{ji})} G'\Bigg(\int_0^{T_{ji}} e^{\beta^\top X_{ji}(s)+\gamma^\top Z_{ji}(s)} d\Lambda_j(s) \Bigg) \\
& \exp\Bigg\{ - G\Bigg(\int_0^{T_{ji}} e^{\beta^\top X_{ji}(s)+\gamma^\top Z_{ji}(s)} d\Lambda_j(s) \Bigg) \Bigg\} \Bigg]^{\Delta_{1ji}} \\
& \Bigg[\exp\Bigg\{ - G\Bigg(\int_0^{L_{ji}} e^{\beta^\top X(s)+\gamma^\top Z(s)} d\Lambda_j(s) \Bigg) \Bigg\} - \exp\Bigg\{ - G\Bigg(\int_0^{R_{ji}} e^{\beta^\top X(s)+\gamma^\top Z(s)} d\Lambda_j(s) \Bigg) \Bigg\} \Bigg]^{\Delta_{2ji}} \\
& \Bigg[\exp\Bigg\{ - G\Bigg(\int_0^{L_{ji}} e^{\beta^\top X(s)+\gamma^\top Z(s)} d\Lambda_j(s) \Bigg) \Bigg\} \Bigg]^{1-\Delta_{1ji}-\Delta_{2ji}}
\end{aligned} \tag{2.4}
$$

2.3 Model Estimation

We consider the non-parametric maximum likelihood estimation of β, γ, Λ. Let $0 = t_{j0} < t_{j1} < \cdots < t_{j_{m_j}} < \infty$ be the ordered unique values of $\{\Delta_{1ji}T_{ji}, \Delta_{2ji}L_{ji}, \Delta_{2ji}R_{ji}, (1 - \Delta_{1ji} - \Delta_{2ji})L_{ji}I(R_{ji} = \infty)\}$. For the selection of the subcohort, we consider independent Bernoulli sampling with selection probability $q_1, q_2, q_3 \in (0, 1)$, where q_1, q_2, q_3 denote the selection probability of X for exact observation, interval censored observation and right censored observation. Thus, under our design, the probability that we observe the covariate X_{ji} is

$$
P(\eta_{ji} = 1|Z_{ji}) = \Delta_{1ji}q_1(Z_{ji}) + \Delta_{2ji}q_2(Z_{ji}) + (1 - \Delta_{1ji} - \Delta_{2ji})q_3(Z_{ji}) \tag{2.5}
$$

for $i = 1, \cdots, n_j, J = 1, \cdots, J$. For simplicity, we let the selection probability depend on the baseline phase one covariates, such that $q_1(Z_{ji}) = q_1(Z_{ji}(0))$, $q_2(Z_{ji}) = q_2(Z_{ji}(0))$ and $q_3(Z_{ji}) = q_3(Z_{ji}(0))$. We employ inverse probability weighting to construct the likelihood function. The inverse probability weight is defined as

$$\omega_{ji} = \frac{\eta_{ji}}{P(\eta_{ji} = 1|Z_{ji})} = \frac{\eta_{ji}}{\Delta_{1ji}q_{1j}(Z_{ji}) + \Delta_{2ji}q_{2j}(Z_{ji}) + (1 - \Delta_{1ji} - \Delta_{2ji})q_{3j}(Z_{ji})}$$

We treat Λ_j as a step function with non-negative jumps at the $t'_{jk}s$. Let λ_{jk} be the jump size of the estimator for Λ_j at t_{jk} for $k = 1, \cdots, m_j$ and define $\lambda_0 = 0$. Then the likelihood function in (2.4) can be written as

$$\begin{aligned} L_n(\beta,\gamma,\Lambda) = \prod_{j=1}^{J}\prod_{i=1}^{n_j}\Bigg[&\Lambda_j(T_{ji})e^{\beta^\top X_{ji}(T_{ji})+\gamma^\top Z_{ji}(T_{ji})}G'\Bigg(\sum_{t_{jk}\leq T_{ji}}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\Bigg) \\ &\exp\Bigg\{-G\Bigg(\sum_{t_{jk}\leq T_{ji}}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\Bigg)\Bigg\}\Bigg]^{\Delta_{1ji}\omega_{ji}} \\ &\Bigg[\exp\Bigg\{-G\Bigg(\sum_{t_{jk}\leq R_{ji}}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\Bigg)\Bigg\} - \exp\Bigg\{-G\Bigg(\sum_{t_{jk}\leq L_{ji}}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\Bigg)\Bigg\}\Bigg]^{\Delta_{2ji}\omega_{ji}} \\ &\Bigg[\exp\Bigg\{-G\Bigg(\sum_{t_{jk}\leq L_{ji}}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\Bigg)\Bigg\}\Bigg]^{(1-\Delta_{1ji}-\Delta_{2ji})\omega_{ji}} \end{aligned} \tag{2.6}$$

where $X_{jik} = X_{ji}(t_k)$ and $\Lambda_j\{T_{ji}\}$ denotes the jump size of Λ_j at T_{ji}. We introduce a latent variable ξ_{ji} and density $f(\xi_{ji})$ as in (2.3), By the properties $\exp\{-G(x)\} = \int_0^\infty \exp(-x\xi)f(\xi)d\xi$ and $G'(x)\exp\{-G(x)\} = \int_0^\infty \xi\exp(-x\xi)f(\xi)d\xi$. The likelihood function (2.6) can be written as below

$$\begin{aligned} L_n(\beta,\gamma,\Lambda) = \prod_{j=1}^{J}\prod_{i=1}^{n_j}\Bigg\{\Bigg[&\int_{\xi_{ji}}\exp\Bigg(-\xi_{ji}\sum_{t_{jk}<T_{ji}}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\Bigg)\xi_{ji}\Lambda_j(T_{ji})e^{\beta^\top X_{ji}(T_{ji})+\gamma^\top Z_{ji}(T_{ji})} \\ &\exp\Bigg(-\xi_{ji}\Lambda_j(T_{ji})e^{\beta^\top X_{ji}(T_{ji})+\gamma^\top Z_{ji}(T_{ji})}\Bigg)f(\xi_{ji})d\xi_{ji}\Bigg]^{\Delta_{1ji}\omega_{ji}} \\ &\Bigg[\int_{\xi_{ji}}\exp\Bigg(-\xi_{ji}\sum_{t_{jk}\leq L_i}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\Bigg) \\ &\Bigg\{1-\exp\Bigg(-\xi_{ji}\sum_{L_{ji}<t_{jk}\leq R_i}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\Bigg)\Bigg\}^{I(R_{ji}<\infty)}f(\xi_{ji})d\xi_{ji}\Bigg]^{\Delta_{2ji}\omega_{ji}} \\ &\Bigg[\int_{\xi_{ji}}\exp\Bigg(-\xi_{ji}\sum_{t_{jk}\leq L_i}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\Bigg)f(\xi_{ji})d\xi_{ji}\Bigg]^{(1-\Delta_{1ji}-\Delta_{2ji})\omega_{ji}} \end{aligned} \tag{2.7}$$

Zeng et al. (2016) and Zhou et al. (2021) developed an EM algorithm for maximum likelihood estimation. We show in Lemma 4.1 that an EM-algorithm can be used to maximize the weighted likelihood. In particular, we introduce the latent variables $W_{jik}(j = 1, \cdots, J; i = 1, \cdots, n_j; k = 1, \cdots, m_j)$, which conditional on ξ_{ji}, are independent Poisson random variables with means $\xi_{ji}\lambda_{jk}\exp(\beta^\top X_{jik}+\gamma^\top Z_{jik})$, i.e.

$$W_{jik}|\xi_{ji} \sim Pois\Big(\xi_{ji}\lambda_{jk}\exp(\beta^\top X_{jik}+\gamma^\top Z_{jik})\Big)$$

and let

$$\begin{cases} A_{ji} = \Delta_{1ji}\sum_{t_{jk}<T_{ji}} W_{jik} \\ B_{ji} = \Delta_{1ji}\sum_{t_{jk}=T_{ji}} W_{jik} \\ C_{ji} = \Delta_{2ji}\sum_{t_{jk}\le L_{ji}} W_{jik} \\ D_{ji} = \Delta_{2ji}\sum_{L_{ji}<t_{jk}\le R_{ji}} W_{jik} \\ E_{ji} = (1-\Delta_{1ji}-\Delta_{2ji})\sum_{t_{jk}\le L_{ji}} W_{jik} \end{cases}$$

for $j = 1, \cdots J; i = 1, \dots, n_j$. Then the observed data consist of

$$\begin{cases} (T_{ji},\, \eta_{ji}X_{ji},\, Z_{ji},\, A_{ji}=0,\, B_{ji}=1) & \text{if } \Delta_{1ji}=1 \\ (L_{ji},\, R_{ji},\, \eta_{ji}X_{ji},\, Z_{ji},\; C_{ji}=0,\, D_{ji}>0) & \text{if } \Delta_{2ji}=1 \\ (L_{ji},\, \eta_{ji}X_{ji},\, Z_{ji},\, E_{ji}=0) & \text{if } 1-\Delta_{1ji}-\Delta_{2ji}=1 \end{cases} \tag{2.8}$$

The equation (2.6) can be re-written as

$$\begin{aligned} L_n(\beta,\gamma,\Lambda) = \prod_{j=1}^{J}\prod_{i=1}^{n_j}\Big\{ & \Big[\int_{\xi_{ji}} P(\sum_{t_{jk}<T_{ji}} W_{jik}=0|\xi_{ji})P(I(t_{jk}=T_{ji})W_{jik}=1|\xi_{ji})f(\xi_{ji})d\xi_{ji}\Big]^{\Delta_{1ji}\omega_{ji}} \\ & \Big[\int_{\xi_{ji}} P(\sum_{t_{jk}\le L_{ji}} W_{jik}=0|\xi_{ji})P(\sum_{L_{ji}<t_{jk}\le R_{ji}} W_{jik}>0|\xi_{ji})^{I(R_{ji}<\infty)}f(\xi_{ji})d\xi_{ji}\Big]^{\Delta_{2ji}\omega_{ji}} \\ & \Big[\int_{\xi_{ji}} P(\sum_{t_{jk}\le L_{ji}} W_{jik}=0|\xi_{ji})f(\xi_{ji})d\xi_{ji}\Big]^{(1-\Delta_{1ji}-\Delta_{2ji})\omega_{ji}}\Big\} \\ = \prod_{j=1}^{J}\prod_{i=1}^{n_j}\Big\{ & \Big[\int_{\xi_{ji}} P(A_{ji}=0|\xi_{ji})P(B_{ji}=1|\xi_{ji})f(\xi_{ji})d\xi_{ji}\Big]^{\Delta_{1ji}\omega_{ji}} \\ & \Big[\int_{\xi_{ji}} P(C_{ji}=0|\xi_{ji})P(D_{ji}>0|\xi_{ji})f(\xi_{ji})d\xi_{ji}\Big]^{\Delta_{2ji}\omega_{ji}} \\ & \Big[\int_{\xi_{ji}} P(E_{ji}=0|\xi_{ji})f(\xi_{ji})d\xi_{ji}\Big]^{(1-\Delta_{1ji}-\Delta_{2ji})\omega_{ji}}\Big\} \end{aligned} \tag{2.9}$$

We maximize the equation (2.9) through an EM algorithm by treating ξ_{ji} and W_{jik} as missing data.

The corresponding weighted log-likelihood function based on the complete data is given by

$$
\begin{aligned}
l_n^{\omega}(\beta,\Lambda) = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\omega_{ji}\Bigg(\sum_{k=1}^{m_j} I(t_{jk}\le R_{ji}^{*})[W_{jik}\log\{\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\} \\
-\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}} - \log(W_{jik}!)] + \log f(\xi_{ji})\Bigg)
\end{aligned}
\tag{2.10}
$$

where $R_{ji}^{*} = \Delta_{1ji}T_{ji} + \Delta_{2ji}R_{ji} + (1-\Delta_{1ji}-\Delta_{2ji})L_{ji}$; In the M-step, we calculate

$$
\lambda_{jk} = \frac{\sum_{i=1}^{n_j}\omega_{ji}I(t_{jk}\le R_{ji}^{*})\hat{\mathbb{E}}W_{jik}}{\sum_{i=1}^{n_j}\omega_{ji}I(t_{jk}\le R_{ji}^{*})\hat{\mathbb{E}}(\xi_{ji})e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}} \qquad (k=1,2,...,m_j, \quad j=1,...,J) \tag{2.11}
$$

where $\hat{\mathbb{E}}(\cdot)$ denotes the posterior mean given the observed data. Then plug in λ_{jk} (2.11) into the equation (2.10) and get the update of β,γ by solving the following equations:

$$
\frac{\partial\hat{\mathbb{E}}l_n^{\omega}(\beta,\lambda)}{\partial\beta} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\omega_{ji}\Bigg(\sum_{k=1}^{m_j} I(t_{jk}\le R_{ji}^{*})\hat{\mathbb{E}}(W_{jik})\Bigg[X_{jik} - \frac{\sum_{i=l}^{n_j}\omega_{jl}I(t_{jk}\le R_{jl}^{*})\hat{\mathbb{E}}(\xi_{jl})e^{\beta^\top X_{jlk}+\gamma^\top Z_{jlk}}X_{jlk}}{\sum_{l=1}^{n_j}\omega_{jl}I(t_{jk}\le R_{jl}^{*})\hat{\mathbb{E}}(\xi_{jl})e^{\beta^\top X_{jlk}+\gamma^\top Z_{jlk}}}\Bigg]\Bigg)
$$

$$
\frac{\partial\hat{\mathbb{E}}l_n^{\omega}(\beta,\lambda)}{\partial\gamma} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\omega_{ji}\Bigg(\sum_{k=1}^{m_j} I(t_{jk}\le R_{ji}^{*})\hat{\mathbb{E}}(W_{jik})\Bigg[Z_{jik} - \frac{\sum_{l=1}^{n_j}\omega_{jl}I(t_{jk}\le R_{jl}^{*})\hat{\mathbb{E}}(\xi_{jl})e^{\beta^\top X_{jlk}+\gamma^\top Z_{jlk}}Z_{jlk}}{\sum_{l=1}^{n_j}\omega_{jl}I(t_{jk}\le R_{jl}^{*})\hat{\mathbb{E}}(\xi_{jl})e^{\beta^\top X_{jlk}+\gamma^\top Z_{jlk}}}\Bigg]\Bigg)
$$

In E-step, we calculate the posterior means $\hat{\mathbb{E}}(W_{jik})$ and $\hat{\mathbb{E}}(\xi_{ji})$. Define $S_{jiT} = \Delta_{1ji}\sum_{t_{jk}\le T ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma Z_{jik}}$, $S_{jiL} = (1-\Delta_{j1i})\sum_{t_{jk}\le L_{ji}}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma Z_{jik}}$ and $S_{jiR} = (1-\Delta_{1ji})\sum_{t_{jk}\le R_{ji}}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma Z_{jik}}$. For $\Delta_{1ji}=1$, we have

$$
\begin{aligned}
\hat{\mathbb{E}}(\xi_{ji}) &= E(\xi_{ji}|A_{ji}=0,B_{ji}=1) \\
&= \int_{\xi_{ji}}\xi_{ji}P(\xi_{ji}|A_{ji}=0,B_{ji}=1)d\xi_{ji} \\
&= G^{'}(S_{jiT}) - \frac{G^{''}(S_{jiT})}{G^{'}(S_{jiT})}
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{\mathbb{E}}(W_{jik}) &= E(W_{jik}|A_{ji}=0,B_{ji}=1) \\
&= \begin{cases} 1, & t_{jk}=T_{ji} \\ 0, & t_{jk}<T_{ji} \end{cases}
\end{aligned}
$$

For $\Delta_{2ji} = 1$, we have

$$
\begin{aligned}
\hat{\mathbb{E}}(\xi_{ji}) &= \int_{\xi_{ji}} \xi_{ji} f(\xi_{ji}|C_{ji} = 0, D_{ji} > 0) d\xi_{ji} \\
&= \int_{\xi_{ji}} \xi_{ji} \frac{(\exp\{-\xi_{ji} S_{jiL}\} - \exp\{-\xi_{ji} S_{jiR}\}) f(\xi_{ji})}{\exp\{-G(S_{jiL})\} - \exp\{-G(S_{jiR})\}} d\xi_{ji} \\
&= \frac{G^{'}(S_{jiL}) \exp\{-G(S_{jiL})\} - G^{'}(S_{jiR}) \exp\{-G(S_{jiR})\}}{\exp\{-G(S_{jiL})\} - \exp\{-G(S_{jiR})\}}
\end{aligned}
$$

and for $\hat{\mathbb{E}}(W_{jik})$, when $t_{j_k} \leq L_{ji}$, we have

$$
\hat{\mathbb{E}}(W_{jik}) = 0
$$

and $L_{ji} < t_{j_k} \leq R_{ji}$,

$$
\begin{aligned}
\hat{\mathbb{E}}(W_{jik}) &= E_{\xi_{ji}}\{E(W_{jik}|\xi_{ji}, C_{ji} = 0, D_{ji} > 0)|C_{ji} = 0, D_{ji} > 0\} \\
&= E_{\xi_{ji}}\{\frac{\xi_{ji}\lambda_{jk} e^{\beta^\top X_{jik}}}{1 - \exp\{-\xi_{ji}(S_{jiR} - S_{jiL})\}}|C_{ji} = 0, D_{ji} > 0\} \\
&= \frac{\lambda_{jk} \exp\{\beta^\top X_{jik}\}}{\exp\{-G(S_{jiL})\} - \exp\{-G(S_{jiR})\}} G^{'}(S_{jiL}) \exp\{-G(S_{jiL})\}
\end{aligned}
$$

For right censored observation, $\Delta_{1ji} = \Delta_{2ji} = 0$, we have

$$
\begin{aligned}
\hat{\mathbb{E}}(\xi_{ji}) &= E(\xi_{ji}|O_{ji}) \\
&= \int_{\xi_{ji}} \xi_{ji} f(\xi_{ji}|O_{ji}) d\xi_{ji} \\
&= G^{'}(S_{jiL})
\end{aligned}
$$

The conditional expectation of W_{jik} is

$$
\begin{aligned}
\hat{\mathbb{E}}(W_{jik}) &= E_{\xi_{ji}}\{E(W_{jik}|O_{ji}, \xi_{ji})|O_{ji}\} \\
&= 0
\end{aligned}
$$

We iterate between the E- and M-steps until convergence, for example, stopping when the maximum of the absolute differences of the estimates at two successive iterations is less than 10^{-4}. Denote the final estimator of regression coefficients as $(\hat{\beta}, \hat{\gamma})$. The final estimator of baseline hazard for the j-th stratum is obtained as $\hat{\Lambda}_j(t) = \sum_{k=1}^{m_j} \mathbb{I}(t \leq t_{jk})\hat{\lambda}_{jk}$ where $\hat{\lambda}_{jk}$ is the final estimates of λ_{jk}. The high dimensional parameter λ_{jk} are calculated explicitly in the M-step. We proved the weighted log-likelihood our weighted EM is non-decreasing from Lemma 4.1 in Appendix.

2.4 Variance Estimation

Our variance estimator is derived from a weighted bootstrap approach. Unlike the conventional bootstrap method, which involves resampling the data with replacement, the weighted bootstrap technique applies different weights to the log-likelihood function without actual resampling of the data. Typically, these weights are generated from a random sample with a mean and variance equal to 1. However, the conventional weighted bootstrap method often fails to provide accurate results because it overlooks negative between-subject correlations induced by sampling, resulting in an inflated variance estimator (Cai and Zheng (2013), Payne et al. (2016)). Particularly, if the missingness of covariate depends on parameters, the weights are not independent, meaning the weight of one observation influences others. In such cases, the usual weighted bootstrap method proves ineffective. To obtain a valid variance estimator, we address this issue by perturbing both the inverse probability weights and the log-likelihood function. As outlined in Cai et al., this allows us to recover the effect of correlation on the variance by estimating the selection probabilities for each perturbed sample. The procedure is described as follows:

- Generate a sequence IID random variables u_{ji} from $Exp(1)$ and let $U = \{u_{ji}, j = 1 \cdots J, i = 1 \cdots n_j\}$
- Use U to obtain perturbed weights $\omega^*_{ji} = \frac{\eta_{ji}}{\hat{q}^*_{ji}}$ where $\hat{q}^*_{ji}$ is obtained from the following procedures:
  - Fit a logistic regression model using the weighted log-likelihood with weight $\{u_{j1}, u_{j2}, \ldots, u_{jn_j}\}$. Let $\hat{\pi}_j = (\hat{\pi}_{0j}, \hat{\pi}_{1j})$ be the coefficients of the fitted logistic regression model, i.e.

$$\hat{\pi}_j = \arg\max_{\pi_j} \sum_{i=1}^{n_j} u_{ji} \left\{ \eta_{ji}(\pi_{0j} + \pi_{1j} z_{ji}) - \log[1 + \exp(\pi_{0j} + \pi_{1j} z_{ji})] \right\}$$

  - Then $\hat{q}^*_{ji} = \frac{\exp(\hat{\pi}_{0j} + \hat{\pi}_{1j} z_{ji})}{1 + \exp(\hat{\pi}_{0j} + \hat{\pi}_{1j} z_{ji})}$ (Notice that when the missingness of X does not depend on Z, then $\hat{q}^*_{ji} = \frac{\exp(\hat{\pi}_{0j})}{1 + \exp(\hat{\pi}_{0j})} = \frac{\sum_{i=1}^{n_j} u_{ji}\eta_{ji}}{\sum_{i=1}^{n_j} u_{ji}}$)

- With the perturbed inverse probability weight, we can set up our weighted complete data log-likelihood (perturbed)

$$l_n^*(\beta,\Lambda)=\sum_{j=1}^{J}\sum_{i=1}^{n_j}u_{ji}\omega_{ji}^*\bigg(\sum_{k=1}^{m_j}I(t_{jk}\le R_{ji}^*)[W_{jik}\log\{\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma Z_{ik}}\}$$
$$-\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma Z_{jik}}-\log(W_{jik}!)]+\log f(\xi_{ji})\bigg)$$

  where $R_{ji}^*=\Delta_{1ji}T_{ji}+\Delta_{2ji}R_{ji}+(1-\Delta_{1ji}-\Delta_{2ji})L_{ji}$

- Use the EM procedure introduced before with

$$\lambda_{jk}^*=\frac{\sum_{i=1}^{n_j}u_{ji}\omega_{ji}^*I(t_{jk}\le R_{ji}^*)\hat{\mathbb{E}}(W_{jik})}{\sum_{i=1}^{n_j}u_{ji}\omega_{ji}^*I(t_{jk}\le R_{ji}^*)\hat{\mathbb{E}}(\xi_{ji})e^{\beta^T X_{jik}+\gamma Z_{jik}}}$$

-

$$\begin{aligned}\sum_{j=1}^{J}\sum_{i=1}^{n_j}u_{ji}\omega_{ji}^*\bigg(\sum_{k=1}^{m_j}I(t_{jk}\le R_{ji}^*)\hat{\mathbb{E}}(W_{jik})[X_{jik}-\frac{\sum_{l=1}^{n_j}u_{jl}\omega_{jl}^*I(t_{jk}\le R_{jl}^*)\hat{\mathbb{E}}(\xi_{jl})e^{\beta^\top X_{jlk}+\gamma Z_{jlk}}X_{jlk}}{\sum_{l=1}^{n_j}u_{jl}\omega_{jl}^*I(t_{jk}\le R_{jl}^*)\hat{\mathbb{E}}(\xi_{jl})e^{\beta^\top X_{jlk}+\gamma Z_{jlk}}}]\bigg)\triangleq 0\\ \sum_{j=1}^{J}\sum_{i=1}^{n_j}u_{ji}\omega_{ji}^*\bigg(\sum_{k=1}^{m_j}I(t_{jk}\le R_{ji}^*)\hat{\mathbb{E}}(W_{jik})[Z_{jik}-\frac{\sum_{l=1}^{n_j}u_{jl}\omega_{jl}^*I(t_{jk}\le R_{jl}^*)\hat{\mathbb{E}}(\xi_{jl})e^{\beta^\top X_{jlk}+\gamma Z_{jlk}}Z_{jlk}}{\sum_{l=1}^{n_j}u_{jl}\omega_{jl}^*I(t_{jk}\le R_{jl}^*)\hat{\mathbb{E}}(\xi_{jl})e^{\beta^\top X_{jlk}+\gamma Z_{jlk}}}]\bigg)\triangleq 0\end{aligned} \tag{2.12}$$

- Repeat the above procedure B times to get an B different $\hat{\beta}^*,\hat{\gamma}^*,\hat{\Lambda}^*$s.
  - The standard error estimator of $\hat{\beta},\hat{\gamma}$ are the sample standard deviation of those B $\hat{\beta}^*,\hat{\gamma}^*$s.
  - In each iteration of the procedure, the function $\hat{\Lambda}^*$ is interpolated over the interval $[0,\tau]$ with a spacing of 0.01. This results in a sequence of time points $0=t_1,t_2,\ldots,t_M=\tau$, where the difference between two consecutive time points is 0.01. This interpolation process yields a new function, denoted as $\hat{\Lambda}^{*'}$. The standard error estimator for $\hat{\Lambda}$ is then calculated as the sample standard deviation of these B sets of $\hat{\Lambda}^{*'}$ values at the specified time points $t_1,t_2,\ldots,t_M$.

2.5 Asymptotic Results

Without of losing generality, we prove the situation with $J=1$ which means single baseline. Define $\vartheta=(\beta,\gamma)$, $\theta=(\vartheta,\Lambda)$. We establish the asymptotic properties of $\hat{\theta}=(\hat{\vartheta},\hat{\Lambda})$ under the following regularity conditions.

1. The true value of ϑ, denoted by ϑ_0, lies in the interior of a known compact set $\mathcal{B}$ in R^d, and the true value of $\Lambda(\cdot)$, denoted by $\Lambda_0(\cdot)$, is continuously with positive derivatives in $[\zeta, \tau]$, where $[\zeta, \tau]$ is the union of the supports of $\{\Delta_1 T, (1-\Delta_1)L, (1-\Delta_1)R, (1-\Delta_1-\Delta_2)L\}$.

2. The vector $X(t)$, $Z(t)$ are uniformly bounded with uniformly bounded total variation over $[\zeta, \tau]$, and its left limit exists for any t. ($|X(t)| \leq K_x, |Z(t)| \leq K_z$ with $0 < K_x, K_z < \infty$) In addition, for any continuously differentiable function $g(\cdot)$, the expectations $E[g\{X_{(j)}(t)\}]$ for $j = 1, 2$ are continuously differentiable in $[\zeta, \tau]$, where $X_{(1)}(t)$ and $X_{(2)}(t)$ are increasing functions in the decomposition $X(t) = X_{(1)}(t) - X_{(2)}(t)$.

3. if $h(t) + \beta^\top X(t) = 0$ for all $t \in [\zeta, \tau]$ with positive probability, then $h(t) = 0$ for $t \in [\zeta, \tau]$ and $\beta = 0$.

4. $0 < P(\Delta = 0) < 1$, $P(L = \tau, R = \infty | \Delta = 0, \bar{X}) \geq c$ and $P(R - L > \eta_0 | \delta = 0, \bar{X}) = 1$ for some positive constants c and η_0. The conditional density of (L, R) given $\bar{X}$, denoted by $g(u, v)$, has continuous second-order partial derivatives with respect to u and v when $v - u > \eta_0$ and are continuously differentiable with respect to $\bar{X}$.

5. The transformation function G is three times continuously differentiable on $[0, \infty)$ with $G(0) = 0, G^{'}(x) > 0$ and $G(\infty) = \infty$.

6. The sampling probabilities q_1, q_2, q_3 are assumed strictly positive which means there exists a constant σ such that $0 < \sigma < q_i \leq 1$ for $i = 1, 2, 3$.

Theorem 1. *Assume Condition (1)-(6) hold. Then* $||\hat{\vartheta}^{\omega_{\hat{\alpha}_n}} - \vartheta_0||_d \xrightarrow{a.s.} 0$ *and* $\sup_{t \in [\zeta, \tau]} |\hat{\Lambda}(t) - \Lambda_0(t)| \xrightarrow{a.s.} 0$ *as* $n \to \infty$, *where* $|| \cdot ||_d$ *is the Euclidean norm.*

Theorem 2. *Under conditions (1) - (6), Then* $n^{1/2}(\hat{\beta}^{\omega_{\hat{\alpha}_n}} - \beta_0, \hat{\gamma}^{\omega_{\hat{\alpha}_n}} - \gamma_0, \hat{\Lambda}_0^{\omega_{\hat{\alpha}_n}} - \Lambda_0(t))$ *converges in distribution to a mean zero Gaussian process for* $t \in [\zeta, \tau]$.

Theorem 3. *Under conditions (1)-(6), the conditional distribution of* $\sqrt{n}(\tilde{\theta}^{\omega_{\hat{\alpha}_n}} - \hat{\theta}^{\omega_{\hat{\alpha}_n}})$ *given the data converges weakly to the asymptotic distribution of* $\sqrt{n}(\hat{\theta}^{\omega_{\hat{\alpha}_n}} - \theta_0)$.

2.6 Simulation Studies

In our simulation study, we considered two strata, each with its own unique baseline hazard function. The population was divided into these two strata, with the sizes of each stratum being approximately equal. Stratum membership was determined using a Bernoulli sampling process, ensuring an equal probability of assignment to either stratum.The simulations are conducted for the following scenarios:

1. $X = U_1+2U_3$ and $Z = -U_1+2U_2$ where U_1, U_2, U_3 are independent $U(0, 1)$. $\beta = 0.5, \gamma = -0.5$, $\Lambda_1(t) = \log(1 + \frac{t}{2})$, $\Lambda_2(t) = 0.2t$. For noncase: $\eta \sim Ber(0.3)$, means each observation has 30% chance to be selected. The missing rate for noncase is 70%; For case, $\eta \sim Ber(0.9)$, each observation has 90% chance to be selected. The overall missing rate is about 29% for $r = 0$, 32% for $r = 0.5$ and 36% for $r = 1$. Regression coefficient estimations results shown in Table 4.1 and baseline estimation results shown in Figures [4.1-4.9].

2. $X = U_1 + 2U_2$ and $Z \sim Ber(0.5)$ where U_1, U_2 are independent $U(0, 1)$. $\beta = -0.5, \gamma = 0.5$, $\Lambda_1(t) = \log(1 + \frac{t}{2})$, $\Lambda_2(t) = 0.2t$. For noncase, the selection probability for the first stratum is defined as $q_1(\eta = 1|Z) = \frac{\exp(0.1-Z)}{1+\exp(0.1-Z)}$ and for the second stratum, the selection probability defined as $q_2(\eta = 1|Z) = \frac{\exp(0.3-1.2Z)}{1+\exp(0.3-1.2Z)}$. Approximately 60% percent missing for noncase observations. For case, $\eta \sim Ber(0.9)$, each observation has 90% chance to be selected. There are approximately 40% observations in stratum 1 and 60% observations in stratum 2. The overall missing rate is 31% for $r = 0$, 35% for $r = 0.5$ and 38% for $r = 1$. Regression coefficient estimations results shown in Table 4.2 and baseline estimation results shown in Figures [4.10-4.18].

3. $X(t) = A + Bt$ and $Z \sim U(0, 1)$ where $A \sim U(3, 5)$ and $B \sim (-0.5, -1)$. $\beta = -0.5, \gamma = 0.5$, $\Lambda_1(t) = 1.5t$, $\Lambda_2(t) = t$. For noncase, the selection probability is defined as $q(\eta = 1|Z) = \frac{\exp(0.3-2Z)}{1+\exp(0.3-2Z)}$. Approximately 65% missing for noncase. For case, $\eta \sim Ber(0.9)$, each observation has 90% chance to be selected. The overall missing rate is around 28% for $r = 0$, 32% for $r = 0.5$ and 35% for $r = 1$. Regression coefficient estimations results shown in Table

4.3 and baseline estimation results shown in Figures [4.19-4.27].

We simulate partly interval-censored failure time data which encompassing a combination of exact observations, left-censored, interval-censored, and right-censored data. Specifically, for each study participant, we first generate the number of examination times $K \sim Ber(0.8) + 1$. If $K = 1$, we generated a single examination time $U_1 \sim Unif(0, 3\tau/4)$, where $(L, R]$ intervals were defined as $(0, U_1]$ if $T \leq U_1$ and (U_1, ∞) if $T > U_1$. For $K = 2$, we generated two examination times U_1 and U_2, with U_2 being $\min\{0.1 + U_1 + \exp(1)\tau/2, \tau\}$. The corresponding intervals were defined as $(0, U_1]$, $(U_1, U_2]$, and (U_2, ∞) for the respective ranges of T. When generating exact observations, where Δ_1 was set to 0 when $R = \infty$, and Δ_1 was generated from a Bernoulli distribution with probabilities $pt = 0, 0.2, 0.5, 1$ when $R < \infty$. In cases where $\Delta_1 = 1$, the failure time T was assumed to be precisely observed. We conducted these simulations with varying proportions of exact observations (pt) and sample sizes of $n = 800$ or 1200, while maintaining a fixed study duration of $\tau = 5$ for all scenarios, resulting in 30% missing rate. All our findings and results are based on 500 replicate simulations. In the table 4.1, 4.2, 4.3, "Bias" is the average point estimate from 500 replicates minus the true parameter value, "SSD" is the sample standard deviation of the point estimates, "ESE" is the average of estimated standard errors, and "CP" is the coverage proportion of 95% confidence intervals. The results presented in Tables 4.1 to 4.3 demonstrate the following findings across all simulation setups: (i) The proposed estimators show virtually no bias; (ii) The weighted bootstrap method provides standard error estimates that reliably represent the true variability of the estimators; (iii) The empirical coverage rates of the 95% confidence intervals, constructed using the normal approximation, are close to the nominal 95% level; (iv) As the sample size grows, both the bias and variability of the estimators diminish; (v) An increase in the proportion of exact observations p_t leads to a reduction in the standard deviation of the estimators, consistent with theoretical predictions.

2.7 Real Data Application

In real-world applications, we applied our proposed model to two randomized HIV trials: the HIV Vaccine Trials Network HVTN-704 (HPTN-085) and HVTN-703 (HPTN -081) (Corey et al.

(2021)). These trials were designed to evaluate the efficacy of a broadly neutralizing antibody, VRC, in preventing the acquisition of human immunodeficiency virus type 1 (HIV-1). HVTN 704/HPTN 085 enrolled and monitored approximately 2,700 men and transgender individuals across North America, South America, and Switzerland. In parallel, HVTN 703/HPTN 081 followed around 1,900 adult women in sub-Saharan Africa. Among the participants, 196 received VRC antibody injections, with antibody concentrations measured monthly from enrollment through 80 weeks. The primary efficacy endpoint was the diagnosis of HIV-1 infection by the week 80 trial visit, with HIV-1 testing conducted at 4-week intervals following enrollment. Due to the nature of the data, the exact timing of HIV-1 infection was not always observable; instead, only the interval during which HIV-1 seroconversion occurred was recorded. Participants who acquired HIV-1 between two visits contributed interval-censored cases, while those who did not acquire HIV-1 by the study's end contributed right-censored observations. Among the 104 participants who acquired HIV-1 across both trials, all had received VRC antibody injections. In our analysis, we applied mid-point imputation to smooth VRC concentration measurements at each time point. This approach involved calculating the average of two consecutive concentration measurements at the midpoint of their respective time intervals. By incorporating this technique, we achieved a more refined and stable representation of VRC concentration dynamics over time, improving data smoothness and mitigating the influence of extreme values. For participants who acquired HIV-1 infection, diagnosis dates were determined using adjudicated diagnosis dates based on validated assays (Corey et al. (2021)). Participant follow-up was right-censored at the earlier of their last negative HIV-1 sample collection date or $\tau = 86$ weeks. Among the 4,611 participants in HVTN-704/HPTN-085 and HVTN-703/HPTN-081 trials, 1,413 were in the USA or Switzerland, 1,274 in Brazil or Peru, 1,019 in South Africa, and 805 in sub-Saharan Africa (including Tanzania, Malawi, Zimbabwe, Botswana, Kenya, and Mozambique). Across both trials, there were 174 HIV-1 infections. Participants were categorized into three age groups: < 20, $[20, 30]$, and > 30, with 328, 2,903, and 1,380 members, respectively. In HVTN-704, 77 VRC concentration measurements were recorded among 846 participants in Brazil and Peru, while only 27 measurements were observed among 943 participants in the USA and Switzerland

(Table 4.4). This disparity underscores regional variations in VRC concentration data availability. Given the distinct social, economic, and health conditions across the four regions (USA/Switzerland, Brazil/Peru, South Africa, and sub-Saharan Africa), we considered employing region-specific baseline hazard functions in our analysis.

2.7.1 Model Fitting

We formulated a semiparametric transformation model of the cumulative hazard function of T given $X(\cdot), Z$ for each subgroup as:

$$\Lambda_j(t|X(\cdot), Z) = G\bigg(\int_0^t \exp\{\beta^\top X(s) + \gamma^\top Z(s)\} d\Lambda_j(s) \bigg) \tag{2.13}$$

where β and γ are unknown regression coefficients, $\Lambda_j(\cdot)$ is an unknown unique baseline function for the j-th stratum, and $G(\cdot)$ is a specified transformation function. In the first scenario, we evaluate the risk of contracting HIV over time with the time-dependent covariate $X = \log(\text{VRC})$ and the time-independent covariates $Z = (\text{High Dose}, \text{Age})$. The covariate 'High Dose' indicates whether the observation belongs to the high dose group, where High Dose $= 1$ denotes inclusion in the high dose group, and High Dose $= 0$ denotes otherwise. Let the first stratum be region USA/Switzerland, second stratum be region Brazil/Peru, third stratum be region South Africa and fourth stratum be region sub-Saharan Africa. To explore the differences in VRC concentration availability across different regions, we performed a logistic regression analysis to evaluate the potential relationship between VRC availability and region. The fitted logistic regression is represented by the equation (2.14):

$$\log \frac{Pr(\eta = 1)}{1 - Pr(\eta = 1)} = -0.335 - 0.9086\mathbb{I}(\text{USAS}) + 0.1353\mathbb{I}(\text{BP}) + 0.1243\mathbb{I}(\text{SA}) \tag{2.14}$$

Detailed information about this model is summarized in Table 4.5. The results indicate a statistically significant association between VRC availability and region, with a p-value of 0.00117 for the USA/Switzerland comparison. Then we apply the inverse probability weighting based on the fitted logistic regression (2.14) to fit the model (2.13). The standard errors of the model coefficients are estimated using a weighted bootstrap with 500 repetitions. Specifically, they are computed as the standard deviation of the results from 500 estimations with different weights, as described in Section

2.4.

We find log(VRC) concentration is negatively related to the risk of contracting HIV and elder people (age older than 20) are less likely to contract HIV. However, we find people in High Dose VRC group are more like to contract HIV compared with people in Low Dose group, which is counter-intuitive. We look at the log(VRC) plot (Fig. 2.1) over time with high/low dose group. The high-dose group

Table 2.1: Analysis results for HIV data with High Dose Indicator

Trials	Covariates	Proportional hazards			Proportional odds			Transformation model ($r = 0.5$)		
		Est.	SE	P.value	Est.	SE	P.value	Est.	SE	P.value
Combined	log(VRC)	-0.534	0.137	< 0.001	-0.573	0.228	0.012	-0.557	0.208	0.007
	High Dose	0.027	0.388	0.945	0.012	0.428	0.978	0.049	0.416	0.906
	Age $20-30$	-0.883	0.532	0.097	-0.914	0.544	0.093	-0.902	0.540	0.095
	Age > 30	-1.923	0.620	0.0019	-1.964	0.644	0.002	-1.947	0.640	0.002
HVTN-703	log(VRC)	-0.200	0.443	0.652	-0.213	0.456	0.640	-0.207	0.450	0.646
	High Dose	-0.187	0.783	0.811	-0.180	0.785	0.819	-0.184	0.784	0.814
	Age $20-30$	0.029	1.024	0.977	0.023	1.031	0.982	0.026	1.027	0.980
	Age > 30	0.110	0.985	0.911	0.105	1.002	0.917	0.107	0.993	0.914
HVTN-704	log(VRC)	-0.564	0.146	< 0.001	-0.707	0.304	0.020	-0.640	0.253	0.011
	High Dose	-0.046	0.493	0.926	0.097	0.602	0.872	0.030	0.560	0.957
	Age $20-30$	-1.104	0.626	0.078	-1.195	0.650	0.066	-1.153	0.640	0.072
	Age > 30	-2.994	0.813	< 0.001	-3.109	0.847	< 0.001	-3.055	0.832	< 0.001

consistently exhibits higher log(VRC) concentrations compared to the low-dose group, as illustrated in the log(VRC) versus dose group comparison, where the high-dose group (red) consistently shows elevated log(VRC) levels. This indicates the confounding between the high-dose indicator and log(VRC) concentrations. Next, we consider the interaction between high-dose indicator and log(VRC) concentrations. In this scenario, $X = (\log(\text{VRC}), \quad \log(\text{VRC}) * \text{High Dose})$. We apply the same inverse probability weighting as from (2.14). The fitted results listed in Table 2.2. We find the interaction term of high dose group and log(VRC) are negative meaning high dose group and log(VRC) are negatively correlated. However, from p.values, they are not significant and large. Other covariates exhibit similar effects as the first scenario (2.1). The estimated survival curves are presented in Figure 4.28. Across all regions, individuals with higher log(VRC) levels consistently show a greater likelihood of remaining HIV-free. Across all time points, in the USAS region, individuals at the 75th percentile of log(VRC) show a higher probability of remaining HIV-free compared to those at the 25th percentile within the same region. Furthermore, in the high-dose group, individuals at the 75th percentile of log(VRC) have a higher probability of remaining HIV-free compared to the low-dose group. This phenomenon is also evident in the relative hazard plot (Fig 2.2), where for log(VRC) values above the median, the high-dose group shows a lower risk of contracting HIV compared to the low-dose group. We also find the cumulative baseline hazards function for those four regions in Figure 4.30. We can tell that the region USA/Switzerland have lowest baseline hazard of getting HIV-1 while the region Brazil/Peru have the highest baseline hazard.

Table 2.2: Analysis results for HIV data with High Dose Indicator and Interaction

		Proportional hazards			Proportional odds			Transformation model ($r = 0.5$)		
Trials	Covariates	Est.	SE	P.value	Est.	SE	P.value	Est.	SE	P.value
Combined	logVRC	-0.320	0.396	0.418	-0.356	0.40	0.374	-0.340	0.397	0.393
	logVRC×High Dose	-0.267	0.397	0.501	-0.349	0.440	0.427	-0.311	0.416	0.455
	High Dose	0.544	0.827	0.510	0.809	1.041	0.437	0.685	0.942	0.467
	Age $20-30$	-0.879	0.499	0.078	-0.934	0.511	0.067	-0.91	0.507	0.073
	Age > 30	-1.920	0.629	0.002	-1.99	0.648	0.002	-1.957	0.640	0.002
HVTN-703	logVRC	-0.142	0.465	0.760	-0.155	0.477	0.746	-0.148	0.471	0.753
	logVRC×High Dose	-0.532	0.990	0.591	-0.536	0.979	0.584	-0.534	0.984	0.587
	High Dose	1.270	2.874	0.659	1.289	2.829	0.649	1.279	2.851	0.654
	Age $20-30$	-0.00372	1.041	0.997	-0.010	1.045	0.992	-0.007	1.043	0.995
	Age > 30	0.0305	1.001	0.976	0.0253	1.016	0.980	0.028	1.008	0.978
HVTN-704	logVRC	-0.646	0.769	0.393	-0.693	0.738	0.348	-0.675	0.748	0.367
	logVRC×High Dose	0.096	0.740	0.896	-0.017	0.695	0.981	0.039	0.708	0.956
	High Dose	-0.225	1.40	0.872	0.129	1.414	0.927	-0.045	1.392	0.974
	Age $20-30$	-1.096	0.635	0.084	-1.197	0.665	0.072	-1.149	0.653	0.079
	Age > 30	-2.990	0.825	0.0003	-3.111	0.865	0.0003	-3.052	0.848	0.0003

Based on the non-significance of the interaction term between the high-dose group and log(VRC) (Table 2.2), combined with the consistent observation that higher log(VRC) concentrations are associated with a reduced likelihood of HIV infection across both dose groups (Figures 4.28 and 2.2), we propose that excluding the high-dose group is a justified approach. In the subsequent analysis, we fit model (2.13) using $Z = \text{Age}$. The removal of the high-dose group is supported by its lack of statistical significance and its minimal relevance to understanding HIV infection mechanisms. The high-dose indicator consistently yielded non-significant and predominantly negative coefficients, indicating a negligible impact on the observed outcomes. Including such non-significant variables adds unnecessary complexity to the model without contributing meaningful insights. To enhance model parsimony and interpretability, we proceed by excluding the high-dose indicator. This refined model is expected to capture the key factors influencing HIV infection while avoiding unnecessary complexity, ultimately providing a clearer understanding of the data relationships. The results of this model are presented in Table 2.3, with the estimated survival curves shown in Figures 4.31 and 4.32. The results confirm that higher log(VRC) concentrations consistently correspond with a lower likelihood of HIV infection. The significance of log(VRC) in the model highlights a strong, inverse

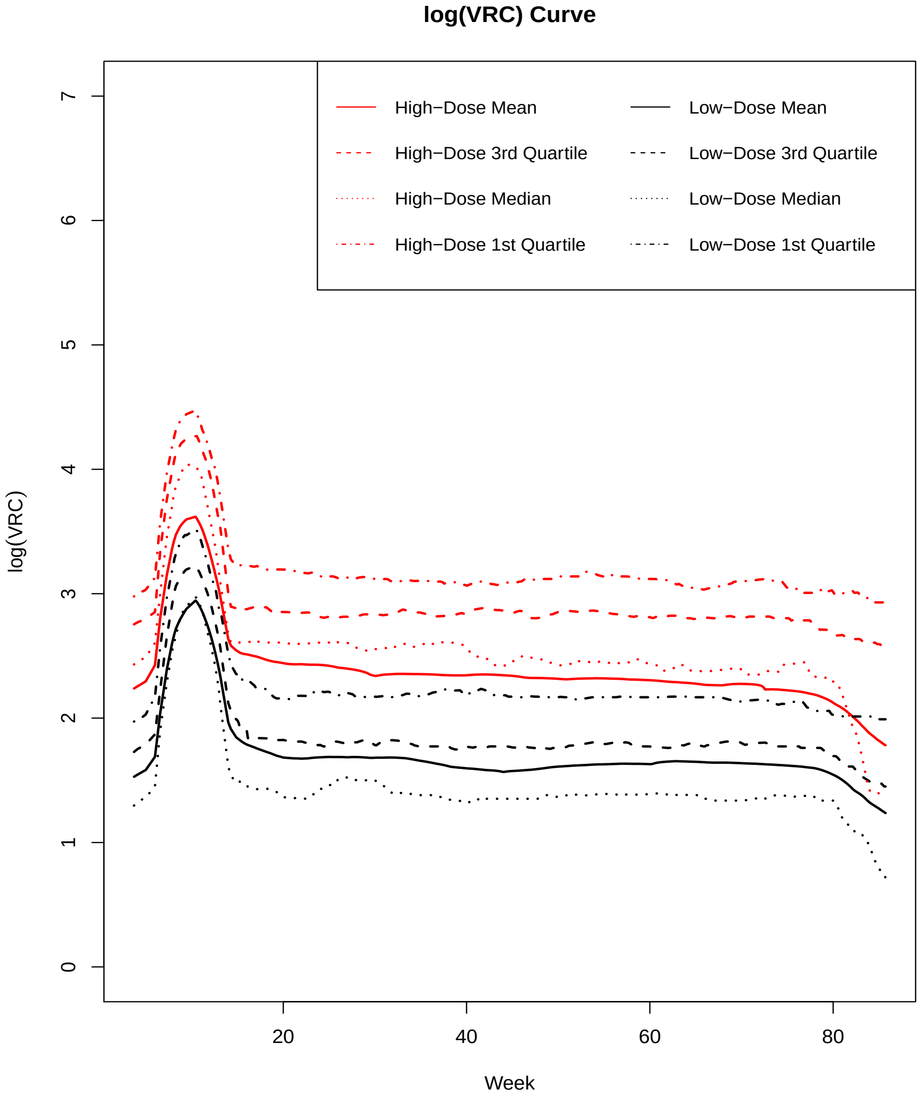

Figure 2.1: Observed log(VRC) concentration for High versus Low dose groups

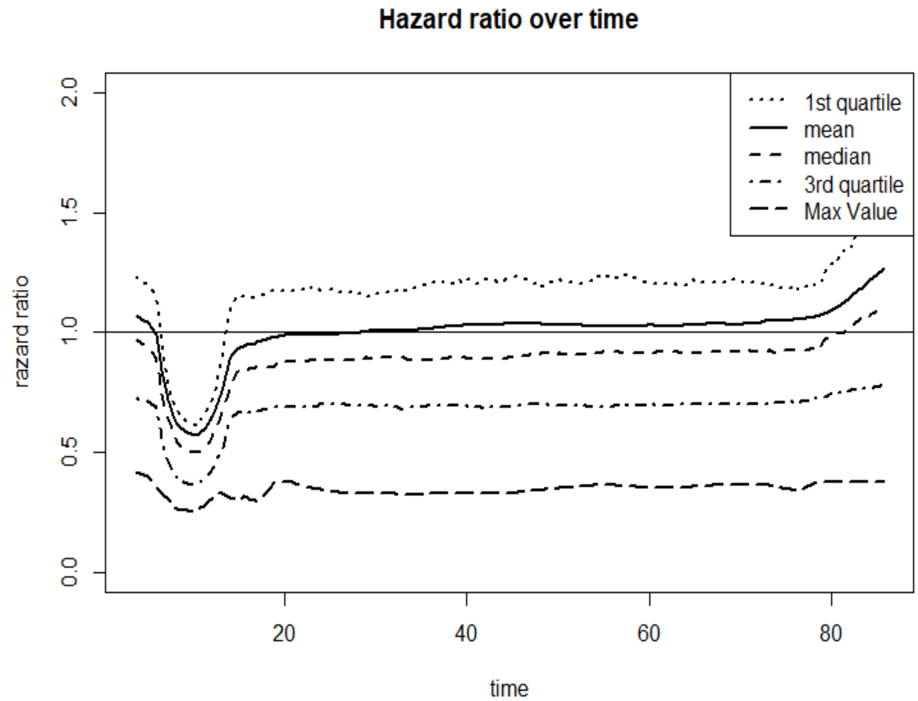

Figure 2.2: Relative hazard of high dose group against low dose group with Z=(High Dose Group, Age) and $X = (\log(\text{VRC}), \log(\text{VRC}) * \text{High Dose})$ at mean and quartiles of log(VRC)

relationship between log(VRC) levels and the risk of HIV infection. This finding underscores the protective role of elevated log(VRC) concentrations and provides valuable insight into its influence in mitigating HIV infection risk.

Table 2.3: Analysis results for HIV data without High Dose Indicator

		Proportional hazards			Proportional odds			Transformation model ($r = 0.5$)		
Trials	Covariates	Est.	SE	P.value	Est.	SE	P.value	Est.	SE	P.value
Combined	logVRC	-0.539	0.134	< 0.0001	-0.556	0.176	0.0016	-0.545	0.163	0.0008
	Age $20 - 30$	-0.893	0.501	0.075	-0.929	0.527	0.078	-0.915	0.517	0.077
	Age > 30	-1.933	0.608	0.0015	-1.979	0.634	0.0018	-1.957	0.640	0.002
HVTN-703	logVRC	-0.264	0.303	0.383	-0.275	0.312	0.378	-0.270	0.307	0.380
	Age $20 - 30$	0.105	0.931	0.910	0.095	0.947	0.920	0.100	0.939	0.915
	Age > 30	0.151	0.958	0.874	0.146	0.977	0.882	0.149	0.967	0.878
HVTN-704	logVRC	-0.567	0.147	0.0001	-0.681	0.246	0.0057	-0.634	0.218	0.0036
	Age $20 - 30$	-1.098	0.617	0.075	-1.200	0.654	0.066	-1.156	0.640	0.071
	Age > 30	-2.983	0.790	0.0001	-3.121	0.840	0.0002	-3.060	0.821	0.0002

2.7.2 Remarks

In the real data application, we conduct two analyses. The first analysis includes the high dose indicator in proposed model. The second analysis does not include the high dose indicator. Both results suggest that larger amount of VRC antibody injection lead to lower risks of getting HIV infection and elder people have lower risk of getting HIV infection. The second model provides better interpretability, given that a majority of its coefficients are statistically significant, whereas the first model does not exhibit the same level of significance. In the trial HVTN-703, both analyses fail to provide significant model fitting results. This is partly caused by the HIV sub-variant in Sub-Saharan has more resistant to VRC01 and higher transmissibility compared to the HIV subvariant in other regions. We also investigated the best transformation model for the dataset. We conduct our analysis with model (2.13) with $G(x) = \frac{\log(1+rx)}{r}$. We tested r values in the interval $[0, 3]$ with a step size of 0.1 and chose the value of r that yielded the maximum weighted log-likelihood at the final parameter estimates. We found that the weighted log-likelihood was maximized at $r = 0$. Our simulation studies in Tables 2.2, 2.3 supported this finding by demonstrating that the fitting regression values did not change significantly for different values of r due to the high censoring rate.

2.8 Concluding Remarks

This paper introduces a novel and unified methodology designed to address the challenges posed by partly interval-censored data within the framework of a linear transformation model, while also accounting for missing covariate information. The proposed approach offers a versatile solution that tackles the complexities associated with both left/right/interval censoring, providing a comprehensive tool for statistical modeling in various research settings. Moreover, the application of our method to HIV antibody research showcases its practical relevance and effectiveness in a real-world context. By incorporating the unique features of HIV antibody data, such as interval-censoring and missing VRC01 information, our approach proves to be a valuable asset in understanding the dynamics of HIV progression. The ability to handle these complexities allows for more accurate and nuanced analyses, contributing to the advancement of our understanding of HIV infection dynamics. The presented results, supported by empirical evidence and simulations, underscore the efficacy and reliability of the proposed method. Its successful application in the field of HIV antibody research reinforces its potential for broader application in epidemiological and biomedical studies where similar data intricacies exist. In conclusion, this paper not only introduces a cutting-edge statistical methodology but also demonstrates its practical utility through its application in the critical domain of HIV research. The proposed unified method has the potential to significantly impact the way researchers analyze and interpret complex datasets, ultimately contributing to advancements in our understanding of various diseases and informing more effective.

CHAPTER 3: THE STRATIFIED COX MODEL WITH LONGITUDINAL COVARIATES MEASURED WITH ERRORS AND SUBJECT TO MISSINGNESS

3.1 Introduction and Literature Review

This study is motivated by the HIV Antibody Trials HVTN-703/HVTN-704, which aim to evaluate the risk of HIV infection associated with the administration of the VRC antibody across diverse geographic regions. In these trials, a subset of participants undergoes VRC antibody concentration measurements at regular 4-week intervals, while others do not receive such monitoring. A primary objective is to assess the associations between age, time-dependent VRC concentration, and the clinical endpoint of HIV infection.

A significant methodological challenge arises from the intermittent and error-prone nature of VRC concentration measurements. Naive approaches that either ignore measurement errors or replace them with imputed values can lead to biased estimates. Additionally, the time-to-event data for the clinical endpoint is partly interval-censored, meaning the exact time of HIV infection is not observed but is known to occur within an interval defined by two consecutive visit dates.

Currently, there is a lack of established statistical methodologies to address the complexities of proportional hazards models in the presence of partly interval-censored failure times, longitudinal covariates subject to measurement errors, and missing covariate information. This study seeks to fill this methodological gap by developing robust inference procedures to ensure valid statistical analysis under these challenging conditions.

Estimation of semiparametric transformation with interval/partly interval censored data have been studied thoroughly. Zeng et al. (2016) proposed an efficient algorithm using Expectation-Maximization (EM) for estimating transformation models with interval-censored data. Zhang et al. (2010) introduced a spline based sieve semiparametric maximum likelihood method to estimate the proportional hazards with interval censored data. Kim (2003) studied the maximum likelihood estimation for the proportional hazards model with partly interval censored data using generalized Gauss-Seidel algorithm with midpoint imputation for the interval censored data. Gao et al. (2017) studied generalized Buckley-James estimator for partly interval censored data with failure time under accelerate failure time model. Gao et al. (2019) proposed an EM algorithm for the partly interval censored data under the asymptomatic disease and symptomatic disease and random effects.

Censored data with covariates measurement error under semiparametric regression model have been studied extensively. Tsiatis and Davidian (2001) studied Cox proportional hazards model with right censored data and longitudinal covariates with measurement error using conditional score approach. Song and Ma (2008) studied interval-censored data with covariate measurement error under proportional hazard model using multiple imputation approach to converting interval censored data into right censored data and then applying conditional score approach. Wen and Chen (2014) studied interval censored data with covariate measurement error using working independence strategy and conditional score approach. Mandal et al. (2019) studied studied interval-censored data with covariate measurement error under linear transformation model by multiple imputation of both event time and covariates. Sun et al. (2023) studied partly interval censored data with covariate measurement error under proportional hazard model by induced hazard approach and EM algorithm.

The existing literature lacks coverage on the analysis of partly interval censored data incorporating covariate measurement errors and covariate missingness within a proportional hazard model framework. This article aims to bridge this gap by introducing a weighted EM procedure for estimating proportional hazards models in the context of partly interval censored failure times. Additionally, we employ an induced hazard approach to address the impact of measurement errors in longitudinal covariates. Assuming an additive measurement error model for longitudinal covariates, we propose a nonparametric maximum likelihood estimation(NPMLE) approach. This involves deriving a measurement error induced hazard model, which illustrates the attenuating effects of ignoring measurement errors. To enable maximum likelihood estimation for partly interval-censored failure times, we devise an EM algorithm. Simulation studies demonstrate the effectiveness of our methods, highlighting their satisfactory finite-sample performance and revealing significant biases in naive approaches that either ignore measurement errors or rely on plug-in estimates. Furthermore, our simulations highlight the attenuating bias introduced by using plug-in estimates for the true underlying longitudinal covariate. While the additive measurement error model is often practical and verifiable for time-independent covariates, caution is advised when applying a measurement error model to time-varying covariates. Commonly used additive random effects models with known time-dependent basis functions may lead to bias if misspecified. The subsequent sections of this chapter are organized as follows: Section 3.2 introduces the data structure, models, and model assumptions. In Section 3.3, we derive the measurement error induced hazard model, while Section 3.4 presents a nonparametric maximum likelihood estimation approach, including the devised EM algorithm for partly interval censored failure times. Section 3.5 introduces a weighted bootstrap variance estimating procedure. The proposed methods' finite-sample performance is assessed through simulation studies in Section 3.7. Additionally, the methods are applied to the HVTN-703/704 data in Section 3.8. Finally, Section 3.9 provides concluding remarks.

3.2 Preliminaries

Let T_{ji} denote the failure time for the i-th subject in the j-th stratum, where $j = 1, \ldots, J$ and $i = 1, \ldots, n_j$. Let Z_{ji} be the $d \times 1$ vector of time-independent covariates that includes baseline covariates and treatment assignment, and $X_{ji}(t)$ the time-dependent covariate of interest. Let $\bar{X}_{ji}(t) = \{X_{ji}(u) : 0 \leq u \leq t\}$ denote the history of $X_{ji}(\cdot)$ up to time $t \in [0, \tau]$, where τ is the end of follow-up time. We assume that the conditional hazard function of T_{ji} given $\bar{X}_{ji}(t)$ and Z_{ji} only depends on Z_{ji} and the current value of $X_{ji}(t)$. Let $\lambda_j(t|\bar{X}_{ji}(t), Z_{ji})$ be the conditional hazard function of T_{ji} given $\bar{X}_{ji}(t)$ and Z_{ji}. We consider the stratified proportional hazards model

$$\lambda_j(t|\bar{X}_{ji}(t), Z_{ji}) = \lambda_j(t) \exp\{\beta X_{ji}(t) + \gamma^\top Z_{ji}\} \tag{3.1}$$

for $0 \leq t \leq \tau$, where $\lambda_j(t)$ is an unspecified baseline function for the j-th stratum, and β and γ are 1- and d- dimensional vectors of regression parameters, respectively. We investigate model (3.1) under partly interval censored failure time data and when the time-dependent covariate $X_{ji}(t)$ is subject to missing and measurement error. Partly interval censored data include observations of failure times that are precisely observed and failure times that are left, interval and/or right censored. Let Δ_{1ji} indicate whether the failure time T_{ji} is observed exactly, i.e. $\Delta_{1ji} = 1$ if T_{ji} is exactly observed and 0 otherwise. If $\Delta_{2ji} = 1$, let $(L_{ji}, R_{ji}]$ denote the smallest observed interval that brackets T_{ji}, where $L_{ji} \geq 0$ is the last monitoring time at which failure time has not occurred and $0 < R_{ji} < \infty$ is the first monitoring time at which failure time has occurred. $R_{ji} = \infty$ represent the situation where failure has not occurred by the last monitoring time which means the failure time being right censored by the last monitoring time. Thus, if $R_{ji} = \infty$, T_{ji} is right censored; if $0 \leq L_{ji} < R_{ji} < \infty$, T_{ji} is interval censored. The partly interval censored failure time for the i-th individual at j-th stratum can be represented as $\{\Delta_{1ji}, \Delta_{1ji}T_{ji}, \Delta_{2ji}, (1-\Delta_{1ji})L_{ji}, (1-\Delta_{1ji})R_{ji}\}$. The notations $\Delta_{1ji}, (1-\Delta_{1ji})L_{ji}, (1-\Delta_{1ji})R_i$ mean we observed T_{ji} if $\Delta_{1ji} = 1$ and we observed (L_{ji}, R_{ji}) if $\Delta_{1ji} = 0$. In the HVTN-703/HVTN-704 study, the failure time of interest is the time to HIV infection. Linear mixed effects models are commonly used to model longitudinal covariates measured with errors. Suppose that $X_{ji}(t)$ is measured at times $\nu_{ji,1} < \cdots < \nu_{ji,M_{ji}}$ before τ with errors and there are B_{jim} repeated measurements or replicates of $X_{ji}(\nu_{jim})$, where we let $B_{jim} = 1$ if there are no replicates. Let $W_{ji,b}(\nu_{jim})$ denote the bth measurement of $X_{ji}(\cdot)$ at time ν_{jim} with $j = 1, ..., J$, $i = 1, \ldots, n_j$, $m = 1, ..., M_{ji}$ and $b = 1, ..., B_{jim}$. We consider the linear mixed effects model for longitudinal covariates with measurement errors:

$$W_{ji,b}(\nu_{jim}) = X_{ji}(\nu_{jim}) + e_{jim,b} = \theta_{ji}^{\top} f(\nu_{jim}) + e_{jim,b} \tag{3.2}$$

where $f(\nu_{jim})$ is an $r \times 1$ vector of known design functions, θ_{ji} is an $r \times 1$ vector of unobserved random effects, and $e_{jim,b}$ is the measurement error at time ν_{jim}. We assume $\theta_{ji} = \vartheta_j + \nu_{ji}$, where ϑ_j is a vector of fixed parameters, and ν_{ji} (for $i = 1, \ldots, n_j$) are independent and identically distributed (iid) as $N(0, G_j)$, with G_j being an $r \times r$ nonnegative definite matrix and $j = 1, \ldots, J$. We also assume that $e_{jim,b}$ (for $i = 1, \ldots, n_j$; $m = 1, \ldots, M_{ji}$; and $b = 1, \ldots, B_{jim}$) are iid $N(0, \sigma_j^2)$ and independent of ν_{ji}, with $j = 1, \ldots, J$. The unknown parameters for the measurement error model are $\theta_{jW} = (\vartheta_j, G_j, \sigma_j^2)$. Also, note that the design function $f(\cdot)$ is usually chosen as a vector of basis functions, such as polynomials. In simulation and real data application, we consider $f(t) = (1, t)$ or

$f(t) = (1, t, t^2)$, orthogonal polynomials, or b-spline basis functions.

Define $W_{jik} = (W_{ji,1}(\nu_{jik}), ..., W_{ji,B_{jik}}(\nu_{jik}))$ and $e_{jik} = (e_{jik,1}, ..., e_{jik,B_{jik}})$. Let $\tilde{\nu} = (\nu_{ji1}, \nu_{ji2}, ..., \nu_{ji,M_{ji}})^\top$. The observed data for the i-th individual at j-th stratum can be formed as

$$\{\Delta_{1ji}, \Delta_{1ji}T_{ji}, (1-\Delta_{1ji})L_{ji}, (1-\Delta_{1ji})R_{ji}, X_{ji}, Z_{ji}, \tilde{\nu}_{ji}, \tilde{W}_{ji}\}, \quad i = 1, ..., n_j$$

We will employ individual-specific estimation of the longitudinal covariate $X_{ji}(t)$ via model (3.2). It does not require repeated measurements at each measurement time ν_{jim}, as long as the number of longitudinal measurements aover time is sufficient to estimate θ_{ji}, i.e. $M_{ji} \geq r$. The proposed estimation method allows $B_{jim} = 1$ for all j, i, m. If we have repeated measurements, then the efficiency in estimating β can be increased Sun et al., 2023.

3.3 Measurement Error Induced Hazard Model

The true longitudinal covariate $X_{ji}(t)$ is not observed. We obtain an individual-specific estimate $\hat{X}_{ji}(t)$ of $X_{ji}(t)$ using ordinary least squares method based on the observed data $(\tilde{\nu}_{ji}, \tilde{W}_{ji})$ and propose an approach by deriving the conditional hazard function of T_{ji} at time at time t conditional on Z_{ji} and $\hat{X}_{ji}(t)$. Only the longitudinal covariates in the past can be used to model current or future risk of failure. For example, in assessing the association of time-dependent HIV antibody VRC concentration with the endpoint HIV infection, only the VRC concentration measurements before HIV infection are meaningfully associated with the endpoint. Thus, we estimate $X_{ji}(t)$ based on the data before t to preserve the predictability.

Let $M_{ji}(t)$ denote the index of the last measurement time before t such that $\nu_{i,M_{ji}(t)} < t \leq \nu_{i,M_{ji}(t)+1}$. Since θ_i is r-dimensional, at least r longitudinal measurements from individual i in j-th before t are required, i.e., $M_{ji}(t) \geq r$. Let $\tilde{\nu}_{ji}(t) = (\nu_{ji1}, ..., \nu_{ji,M_{ji}(t)})^\top$, $\tilde{W}_{ji}(t) = (W_{ji1}, ..., W_{ji,M_{ji}(t)})^\top$ and $\tilde{e}_{ji}(t) = (e_{ji1}, e_{ji2}, ..., e_{ji,M_{ji}(t)})^\top$. Under model (3.2), $\tilde{W}_{ji}(t) = \tilde{F}_{ji}(t)\theta_{ji} + \tilde{e}_{ji}(t)$, where $\tilde{F}_{ji}(t) = \mathbb{B}_{ji}(t)\tilde{f}_{ji}(t)^\top$ with $\mathbb{B}_{ji} = \text{diag}(\mathbf{1}_{B_{ji1}}, \mathbf{1}_{B_{ji2}}, ..., \mathbf{1}_{B_{ji,M_{ji}(t)}})$, $\mathbf{1}_m$ is a $m \times 1$-vector of ones, and $\tilde{f}_{ji}(t) = (f(\nu_{ji1}), ..., f(\nu_{ji,M_{ji}(t)}))$. Hence the ordinary least squares estimator of θ_{ji} based on

$(\tilde{\nu}_{ji}(t), \tilde{W}_{ji}(t))$ for the i-th individual at j-stratum equals

$$\hat{\theta}_{ji}(t) = \left(\tilde{F}_{ji}^{\top}(t)\tilde{F}_{ji}(t)\right)^{-1}\tilde{F}_{ji}^{\top}(t)\tilde{W}_{ji}(t) \tag{3.3}$$

We estimate θ_{ji} based on the observations from subject i-th in the j-th stratum without pulling information from other individuals. The longitudinal covariate $X_{ji}(t)$ is estimated by $\hat{X}_{ji}(t) = f^{\top}(t)\hat{\theta}_{ji}(t)$ based on the observed error-prone covariate information. Since $\hat{\theta}_{ji}(t) = \theta_{ji} + \{\tilde{F}_{ji}^{\top}(t)\tilde{F}_{ji}(t)\}^{-1}\tilde{F}_{ji}^{\top}(t)\tilde{e}_{ji}(t)$, then we have

$$\begin{aligned}
\hat{X}_{ji}(t) &= f^{\top}(t)\hat{\theta}_{ji}(t) \\
&= f^{\top}(t)\theta_{ji}(t) + f^{\top}(t)\{\tilde{F}_{ji}^{\top}(t)\tilde{F}_{ji}(t)\}^{-1}\tilde{F}_{ji}^{\top}(t)\tilde{e}_{ji}(t) \\
&= X_{ji}(t) + f^{\top}(t)\{\tilde{F}_{ji}^{\top}(t)\tilde{F}_{ji}(t)\}^{-1}\tilde{F}_{ji}^{\top}(t)\tilde{e}_{ji}(t)
\end{aligned}$$

The two terms $X_{ji}(t)$ and $\tilde{e}_{ji}(t)$ are independent. Then $\hat{X}_{ji}(t)$ is normally distributed with mean $X_{ji}(t)$ and variance $d_{ji}(t, \sigma_j^2) = \sigma_j^2 f^{\top}(t)\{\tilde{F}_{ji}^{\top}(t)\tilde{F}_{ji}(t)\}^{-1}f(t)$. An estimator of σ_j^2 can be constructed using the residuals:

$$\hat{\sigma}_j^2 = n_j^{-1}\sum_{i=1}^{n_j} M_{ji}^{-1}\sum_{k=1}^{M_{ji}} B_{jik}^{-1}\sum_{b=1}^{B_{jik}}\left(W_{jib}(v_{jik}) - \hat{X}_{ji}(v_{jik})\right)^2 \tag{3.4}$$

Now we derive the induced hazard model of T_{ji} conditional on Z_{ji} and $\hat{X}_{ji}(t)$ under the measurement error model (3.2). Define the counting process increment $dN_{ji}(t) = \mathbb{I}(t \leq T_{ji} < t + dt, v_{jir} \leq t)$ and the at-risk process $Y_{ji}(t) = \mathbb{I}(T_{ji} \geq t, v_{jir} \leq t)$. $dN_{ji}(t) = 1$ means the failure time occurs at time t and after r longitudinal measurements. Motivated by the induced hazard approach in Sun et al. (2023), we can derive the conditional hazard function of T_{ji} at time t given $(\hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t), v_{jir} \leq t)$.

We present the regularity conditions needed for Proposition 1. Let $\tilde{U}_{ji} = (U_{ji1}, U_{ji2}, ..., U_{ji,K_{ji}})$ denote the monitoring times for the failure event for i-th individual for j-th stratum, where $0 = U_{ji0} < U_{ji1} < \cdots < U_{ji,K_{ji}} < U_{ji,K_{ji}+1} = \infty$. We use the monitoring times to generate interval censored failure times $[L_{ji}, R_{ji}]$, where $L_{ji} = \max\{U_{jik} : T_{ji} > U_{jik}, k = 0, ..., K_{ji}\}$ and $R_{ji} = \min\{U_{jik} : T_{ji} \leq U_{jik}, k = 1, ..., K_{ji}\}$ with $U_{ji0} = 0$ and $U_{ji,K_{ji}+1} = \infty$. We assume the following conditions that require non-informative monitoring times and a non-differential measurement error mechanism for the time-dependent covatiates.

1. The monitoring times, measurement times and measurement errors are non-informative given the information already provided by Z_{ji} and θ_{ji}, i.e.,T_{ji} is independent of $(\Delta_{1ji}, \Delta_{2ji}, \tilde{\nu}_{ji}, \tilde{e}_{ji})$ given (Z_{ji}, θ_{ji})

2. Measurement error $\tilde{e}_{ji}$ is independent of $(\Delta_{1ji}, \Delta_{2ji}, \tilde{U}_{ji}, \tilde{\nu}_{ji}, \tilde{e}_{ji})$ given (Z_{ji}, θ_{ji})

3. $X_{ji}(t)$, $0 \leq t \leq \tau$, is a left continuous process

Proposition 1. *Under conditions 1-3, for $j = 1, ..., J$ where J is the number of stratums, the induced hazard function is*

$$\lambda_j^*(t|\hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t)) = \lambda_{0j}(t) \exp\left\{\beta\zeta_{ji}(t)\hat{X}_{ji}(t) + \gamma^\top Z_{ji} + O_{ji}(\beta, t, \theta_{jW})\right\}, \quad \text{for} \quad t \geq v_{jir} \tag{3.5}$$

where $\sigma^2_{ji,rel}(t) = \frac{d_{ji}(t,\sigma_j^2)}{f(t)^\top G_j f(t) + d_{ji}(t,\sigma^2)}$, $\zeta_{ji}(t) = 1 - \sigma^2_{ji,rel}(t) = \frac{f(t)^\top G_j f(t)}{f(t)^\top G_j f(t) + d_{ji}(t,\sigma^2)}$, $O_{ji}(\beta, t, \theta_{jW}) = \beta\{\vartheta_j^\top f(t) + \frac{1}{2}\beta f(t)^\top G_j f(t)\}\sigma^2_{ji,rel}(t)$, $\theta_{jW} = (\vartheta_j, G_j, \sigma_j^2)$.

Proof. By definition of the counting process $dN_{ji}(t)$ and at risk process $Y_{ji}(t)$, we have

$P(dN_{ji}(t) = 1|\hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t), Y_{ji}(t) = 1)$

$= E\{E(I(dN_{ji}(t) = 1)|\tilde{W}_{ji}(t), \tilde{e}_{ji}(t), \hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t), Y_{ji}(t) = 1, \tilde{U}_{ji})|\hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t), Y_{ji}(t) = 1\}$

$= E\{E(I(dN_{ji}(t) = 1)|\tilde{W}_{ji}(t), \tilde{e}_{ji}(t), \hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t), Y_{ji}(t) = 1, \tilde{U}_{ji}I(\tilde{U}_{ji} \leq t), \tilde{U}_{ji}I(\tilde{U}_{ji} > t))|\hat{X}_{ji}(t),$

$Z_{ji}, \tilde{v}_{ji}(t), Y_{ji}(t) = 1\}$

$= E\{\lambda_j(t|X_{ji}(t), Z_{ji})dt|\hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t), Y_{ji}(t) = 1\}$

$= \lambda_{0j}(t) \exp(\gamma^\top Z_{ji}) E\{\exp\{\beta X_{ji}(t)\}|\hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t), Y_{ji}(t) = 1\}dt$

The conditional hazard function of T_{ji} at time t given $(\hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t), v_{jir} \leq t)$ equals

$$\lambda_j^*(t|\hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t)) = \lambda_{0j}(t) \exp(\gamma^\top Z_{ji}) E\{\exp\{\beta X_{ji}(t)\}|\hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t), Y_{ji}(t) = 1\}, \quad \text{for} \quad t \geq v_{jir}$$

By equation (3.3), $X_{ji}(t)$ is normally distributed conditional on $\hat{X}_{ji}(t)$ and $\tilde{v}_{ji}(t)$. Thus $\exp(\beta X_{ji}(t)|\hat{X}_{ji}(t), \tilde{v}_{ji}(t))$

follows log-normal distribution and

$$
\begin{aligned}
&E\{\exp(\beta X_{ji}(t)|\hat{X}_{ji}(t),\tilde{v}_{ji}(t))\}\\
&=\exp\{E(\beta X_{ji}(t)|\hat{X}_{ji}(t),\tilde{v}_{ji}(t))+\frac{1}{2}\mathrm{Var}(\beta X_{ji}(t)|\hat{X}_{ji}(t),\tilde{v}_{ji}(t))\}\\
&=\exp\{\beta E(X_{ji}(t)|\hat{X}_{ji}(t),\tilde{v}_{ji}(t))+\frac{1}{2}\beta^2\mathrm{Var}(X_{ji}(t)|\hat{X}_{ji}(t),\tilde{v}_{ji}(t))\}\\
&=\exp\left\{\beta\frac{E(X_{ji}(t))d_{ji}(t,\sigma^2)+\hat{X}_{ji}(t)f^\top G_j f(t)}{f^\top G_j f(t)+d_{ji}(t,\sigma^2)}+\frac{1}{2}\beta^2\frac{f^\top G_j f(t)d_{ji}(t,\sigma^2)}{f^\top G_j f(t)+d_{ji}(t,\sigma^2)}\right\}
\end{aligned}
$$

The conditional hazard function of T_{ji} at time t given $(\hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t), v_{ji} \le t)$ equals

$$
\begin{aligned}
&\lambda_j^*(t|\hat{X}_{ji}(t),Z_{ji},\tilde{v}_{ji}(t))\\
&=\lambda_{0j}(t)\exp\{\gamma^\top Z_{ji}\}E\{\exp(\beta X_{ji}(t))|\hat{X}_{ji}(t),\tilde{v}_{ji}(t)\},\quad t\ge v_{jir}\\
&=\lambda_{0j}(t)\exp\left\{\gamma^\top Z_{ji}+\beta\frac{E(X_{ji}(t))d_{ji}(t,\sigma^2)+\hat{X}_{ji}(t)f(t)^\top G_j f(t)}{f(t)^\top G_j f(t)+d_{ji}(t,\sigma^2)}+\frac{1}{2}\beta^2\frac{f(t)^\top G_j f(t)d_{ji}(t,\sigma^2)}{f(t)^\top G_j f(t)+d_{ji}(t,\sigma^2)}\right\}\\
&=\lambda_{0j}(t)\exp\left\{\beta\hat{X}_{ji}(t)\frac{f(t)^\top G_j f(t)}{f(t)^\top G_j f(t)+d_{ji}(t,\sigma^2)}+\gamma^\top Z_{ji}+\beta\{\vartheta_j^\top f(t)+\frac{1}{2}\beta f(t)^\top G_j f(t)\}\frac{d_{ji}(t,\sigma^2)}{f(t)^\top G_j f(t)+d_{ji}(t,\sigma^2)}\right\}\\
&=\lambda_{0j}(t)\exp\left\{\beta\zeta_{ji}(t)\hat{X}_{ji}(t)+\gamma^\top Z_{ji}+O_{ji}(\beta,t,\theta_{jW})\right\}
\end{aligned}
$$

□

3.4 Model Estimation

We assume the covariate $\hat{X}_{ji}(t)$ is missing at random (MAR), meaning the probability that $\hat{X}_{ji}(t)$ is missing depends only on observed data and not on the unobserved value of $\hat{X}_{ji}(t)$. Let η_{ji} denote the missingness indicator ($\eta_{ji} = 1$ if $\hat{X}_{ji}(t)$ is observed, and 0 otherwise). Under our design, the observation probability for $\hat{X}_{ji}(t)$ is modeled as:

$$
P(\eta_{ji}=1\mid Z_{ji})=\Delta_{1ji}q_1(Z_{ji})+\Delta_{2ji}q_2(Z_{ji})+(1-\Delta_{1ji}-\Delta_{2ji})q_3(Z_{ji}), \tag{3.6}
$$

for $i = 1,\ldots,n_j$ and $j = 1,\ldots,J$. Here, Δ_{1ji} and Δ_{2ji} are binary indicators for exact observation and interval censoring ($\Delta_{1ji} + \Delta_{2ji} \le 1$), with $1 - \Delta_{1ji} - \Delta_{2ji}$ indicating right censoring. The functions $q_1(Z_{ji})$, $q_2(Z_{ji})$, and $q_3(Z_{ji})$ represent selection probabilities for exact, interval, and right-censored observations, respectively, dependent on the covariate Z_{ji}. To address potential bias from missing data, we employ inverse probability weighting (IPW). The IPW weight is defined as:

$$\omega_{ji} = \frac{\eta_{ji}}{P(\eta_{ji} = 1 \mid Z_{ji})} = \frac{\eta_{ji}}{\Delta_{1ji} q_1(Z_{ji}) + \Delta_{2ji} q_2(Z_{ji}) + (1 - \Delta_{1ji} - \Delta_{2ji}) q_3(Z_{ji})}, \tag{3.7}$$

where the denominator corresponds to the observation probability in Equation (3.6). These weights are incorporated into the likelihood function to adjust for missingness under the MAR assumption. Now we derive an estimator of the induced hazard model based on partly interval censored data. The observed data from a random sample of study participants consist of $\{\Delta_{1ji}, \Delta_{1ji}T_{ji}, \Delta_{2ji}, (1-\Delta_{1ji})L_{ji}, (1-\Delta_{1ji})R_{ji}, Z_{ji}, \tilde{v}_{ji}, \tilde{W}_{ji}\}, i = 1, ..., n_j, j = 1, ..., J$. We revised methods in Zhou et al. (2021), Sun et al. (2023) to estimate the measurement error induced hazard model (3.5) with partly interval censored data. The conditional survival function of T_{ji} given $T_{ji} \geq v_{jir}$ equals $\exp\big(-\int_{v_{jir}}^{t} \lambda^*(x|\hat{X}_i(x), Z_{ji}, \tilde{v}_{ji}(x))dx\big)$. Let $\Lambda_{0j} = \int_0^t = \lambda_j(s)ds$. Let $h_{ji}(t, \beta, \gamma) = \beta\zeta_{ji}(t)\hat{X}_{ji}(t) + \gamma^\top Z_{ji} + O_{ji}(\beta, \gamma; \theta_{jW})$. The observed data weighted likelihood with the induced hazard (3.5) is

$$\begin{aligned} L_n(\beta,\gamma,\Lambda;\theta_W) = \prod_{j=1}^{J}\prod_{i=1}^{n_j} & \left\{ [\Lambda_j^{'}(T_{ji}) \exp\{h_{ji}(T_{ji},\beta,\gamma)\}]^{I(v_{jir} \le T_{ji})} \exp\left(- \int_{v_{jir}}^{T_{ji}} \exp\{h_{ji}(t,\beta,\gamma) d\Lambda_j(t)\} \right) \right\}^{\Delta_{1ji}\omega_{ji}} \\ & \times \left\{ \exp\left(- \int_{v_{jir}}^{L_{ji}} \exp\{h_{ji}(t,\beta,\gamma)\} d\Lambda_j(t) \right) - \exp\left(- \int_{v_{jir}}^{R_{ji}} \exp\{h_{ji}(t,\beta,\gamma)\} d\Lambda_j(t) \right) \right\}^{\Delta_{2ji}\omega_{ji}} \\ & \times \left\{ \exp\left(- \int_{v_{jir}}^{L_{ji}} \exp\{h_{ji}(t,\beta,\gamma)\} d\Lambda_j(t) \right) \right\}^{(1-\Delta_{1ji}-\Delta_{2ji})\omega_{ji}} \end{aligned} \tag{3.8}$$

Following the approach introduced by Zeng et al. (2016), Zhou et al. (2021), we treat $\Lambda_j(t)$ as a step function with non-negative jumps at ordered unique time points T_{ji} and $(L_{ji}, R_{ji}], i = 1, ..., n_j, j = 1, ..., J$. Let $0 = t_{j0} < t_{j1} < \cdots < t_{jm_j}$ be ordered unique values of the sets $\{\Delta_{1ji}T_{ji}, (1-\Delta_{1ji})L_{ji}, \Delta_{2ji}R_{ji} : i = 1, ..., n_j\}$ at j-th stratum. Let λ_{jk} be the jump size of the estimator for $\Lambda_j(t)$ at t_{jk} for $k = 1, ..., m_j$ and let $\lambda_{j0} = 0$. Let $h_{ji}(t_{jk}, \beta, \gamma) = \beta\zeta_{jik}\hat{X}_{jik} + \gamma^\top Z_{ji} + O_{ji}(\beta, t_{jk}, \theta_{jW})$, where $\hat{X}_{jik} = \hat{X}_{ji}(t_{jk})$ and $\zeta_{jik} = \zeta_{ji}(t_{jk})$. The likelihood function of (3.8) becomes

$$\begin{aligned} L_n(\beta, \gamma, \Lambda; \theta_W) = \prod_{j=1}^{J} \prod_{i=1}^{n_j} \Bigg\{ & [\Lambda_j\{T_{ji}\} \exp\{h_{ji}(T_{ji}, \beta, \gamma)\}]^{I(v_{jir} \leq T_{ji})} \\ & \exp\Bigg(- \sum_{t_{jk} \leq T_{ji}} I(v_{jir} \leq t_{jk}) \lambda_{jk} \exp\{h_{ji}(t_{jk}, \beta, \gamma) d\Lambda_j(t)\} \Bigg) \Bigg\}^{\Delta_{1ji}\omega_{ji}} \\ \times \Bigg\{ & \exp\Bigg(- \sum_{t_{jk} \leq L_{ji}} I(v_{jir} \leq t_{jk}) \lambda_{jk} \exp\{h_{ji}(t_{jk}, \beta, \gamma)\} \Bigg) \\ & \Bigg[1 - \exp\Bigg(- \sum_{L_{ji} < t_{jk} \leq R_{ji}} I(v_{jir} \leq t_{jk}) \lambda_{jk} \exp\{h_{ji}(t_{jk}, \beta, \gamma)\} \Bigg) \Bigg] \Bigg\}^{\Delta_{2ji}\omega_{ji}} \\ \times \Bigg\{ & \exp\Bigg(- \sum_{t_{jk} \leq L_{ji}} I(v_{jir} \leq t_{jk}) \lambda_{jk} \exp\{h_{ji}(t_{jk}, \beta, \gamma)\} \Bigg) \Bigg\}^{(1-\Delta_{1ji}-\Delta_{2ji})\omega_{ji}} \end{aligned} \tag{3.9}$$

where $\Lambda_j\{T_{ji}\}$ denote the jump size of $\Lambda_j(t)$ at T_{ji}. We deploy EM algorithm to maximum the likelihood function in equation (3.9). Let ρ_{jik} be independent Poisson random variables with means $\mu_{jik} = \lambda_{jik} \exp\big\{h_{ji}(t_{jk}, \beta, \gamma)\big\}, i = 1, ..., n_j, k = 1, ..., m_j, j = 1, ..., J$. Following Zhou et al. (2021), for $i = 1, ..., n_j$ and $j = 1, ..., J$, we define

$$\begin{cases} A_{ji} = \Delta_{1ji} \sum_{t_{jk} < T_{ji}} I(v_{jir} \leq t_{jk}) \rho_{jik} \\ B_{ji} = \Delta_{1ji} \sum_{t_{jk} = T_{ji}} I(v_{jir} \leq t_{jk}) \rho_{jik} \\ C_{ji} = \Delta_{2ji} \sum_{t_{jk} \leq L_{ji}} I(v_{jir} \leq t_{jk}) \rho_{jik} \\ D_{ji} = \Delta_{2ji} \sum_{L_{ji} < t_{jk} \leq R_{ji}} I(v_{jir} \leq t_{jk}) \eta_{jik} \\ E_{ji} = (1 - \Delta_{1ji} - \Delta_{2ji}) \sum_{t_{jk} \leq L_{ji}} I(v_{jir} \leq t_{jk}) \rho_{jik} \end{cases}$$

Let $\hat{X}_{ji\cdot} = \{\hat{X}_{jik}, k = 1, ..., m_j\}$. The observed data consists of

$$\begin{cases} (\tilde{v}_{ji},\, T_{ji},\, \eta_{ji}\hat{X}_{ji\cdot},\, Z_{ji},\, A_{ji} = 0,\, B_{ji} = 1) & \text{if } \Delta_{1ji} = 1 \\ (\tilde{v}_{ji},\, L_{ji},\, R_{ji},\, \eta_{ji}\hat{X}_{ji\cdot},\, Z_{ji},\, C_{ji} = 0,\, D_{ji} > 0) & \text{if } \Delta_{2ji} = 1 \\ (\tilde{v}_{ji},\, L_{ji},\, \eta_{ji}\hat{X}_{ji\cdot},\, Z_{ji},\, E_{ji} = 0) & \text{if } 1 - \Delta_{1ji} - \Delta_{2ji} = 1 \end{cases} \tag{3.10}$$

The likelihood function of the observed data in (3.10) is

$$L_n^* = \prod_{j=1}^{J} \prod_{i=1}^{n_j} \big\{P(A_{ji} = 0, B_{ji} = 1)\big\}^{\Delta_{1ji}\omega_{ji}} \big\{P(C_{ji} = 0, D_{ji} > 0)\big\}^{\Delta_{2ji}\omega_{ji}} P(E_{ji} = 0)^{(1-\Delta_{1ji}-\Delta_{2ji})\omega_{ji}}$$

Notice that $P(A_{ji} = 0, B_{ji} = 1)$ equivalent to the term in likelihood function (3.9) with $\Delta_{1ji} = 1$, $P(C_{ji} = 0, D_{ji} > 0)$ equivalent to the term in likelihood function (3.9) with $\Delta_{2ji} = 1$ and $P(E_{ji} = 0)$

corresponds to the term in likelihood function (3.9) with $1-\Delta_{1ji}-\Delta_{2ji}=1$. Then L_n^* can be written as

$$L_n(\beta,\gamma,\Lambda;\theta_W)=\prod_{j=1}^{J}\prod_{i=1}^{n_j}\left\{\prod_{t_{jk}<T_{ji}}P(\rho_{jik}=0)^{I(v_{jir}\le t_{jk})}\prod_{t_{jk}=T_{ji}}P(\rho_{jik}=1)^{I(v_{jir}\le t_{jk})}\right\}^{\Delta_{1ji}\omega_{ji}}$$
$$\times\left\{\prod_{t_{jk}\le L_{ji}}P(\rho_{jik}=0)^{I(v_{jir}\le t_{jk})}\left[1-\prod_{L_{ji}<t_{jk}\le R_{ji}}P(\rho_{jik}=0)^{I(v_{jir}\le t_{jk})}\right]\right\}^{\Delta_{2ji}\omega_{ji}}$$
$$\times\left\{\prod_{t_{jk}\le L_{ji}}P(\rho_{jik}=0)^{I(v_{jir}\le t_{j_k})}\right\}^{(1-\Delta_{1ji}-\Delta_{2ji})\omega_{ji}} \tag{3.11}$$

We maximize the likelihood function (3.11). Let $R_{ji}^*=\Delta_{1ji}T_{ji}+\Delta_{2ji}R_{ji}+(1-\Delta_{1ji}-\Delta_{2ji})L_{ji}$ and define $1_{jik}^*=\mathbb{I}(v_{ir}\le t_{jk}\le R_{ji}^*)$ and ρ_{jik} be the independent Poisson random variables with means $\mu_{jik}=\lambda_{jk}\exp\{h_{ji}(t_{jk},\beta,\gamma)\}$ for $j=1\ldots J; i=1,\ldots n_j, k=1,\ldots m_j$. Treating ρ_{jik} as missing data. The complete weighted data log-likelihood is given by

$$l_n(\beta,\Lambda)=\sum_{j=1}^{J}\sum_{i=1}^{n_j}\omega_{ji}\left(\sum_{k=1}^{m_j}1_{jik}^*\left[\rho_{jik}\log\{\mu_{jik}\}-\log(\rho_{jik}!)-\mu_{jik}\right]\right) \tag{3.12}$$

where $R_{ji}^*=\Delta_{1ji}T_{ji}+\Delta_{2ji}R_{ji}+(1-\Delta_{1ji}-\Delta_{2ji})L_{ji}$. The expectation of the complete data log-likelihood is

$$\mathbb{E}(l_n(\beta,\Lambda))=\sum_{j=1}^{J}\sum_{i=1}^{n_j}\omega_{ji}\left(\sum_{k=1}^{m_j}1_{jik}^*\left[\hat{\mathbb{E}}\rho_{jik}\log\{\mu_{jik}\}-\hat{\mathbb{E}}\log(\rho_{jik}!)-\mu_{jik}\right]\right) \tag{3.13}$$

where $\hat{\mathbb{E}}(\cdot)$ is the posterior mean given the observed data. In the M-step, we estimate λ_{jk} and β,γ by maximizing (3.13) using the results from the previous iteration. To compute the optimal λ_{jk}, we solve the score equation:

$$\frac{\partial\mathbb{E}l_n(\beta,\Lambda)}{\partial\lambda_{jk}}=\sum_{i=1}^{n_j}\omega_{ji}\left(1_{jik}^*\left[\hat{\mathbb{E}}\rho_{jik}\frac{1}{\lambda_{jk}}-e^{\beta\zeta_{jik}\hat{X}_{jik}+\gamma^\top Z_{ji}+O_{ji}(\beta,t_{jk},\theta_{jW})}\right]\right)\triangleq 0.$$

The solution for λ_{jk} that maximizes (3.13) is given by

$$\lambda_{jk}=\frac{\sum_{i=1}^{n_j}\omega_{ji}1_{jik}^*\hat{\mathbb{E}}(\rho_{jik})}{\sum_{i=1}^{n_j}\omega_{ji}1_{jik}^*e^{\beta\zeta_{jik}\hat{X}_{jik}+\gamma^\top Z_{ji}+O_{ji}(\beta,t_{jk},\theta_{jW})}},\quad k=1,2,\ldots,m_j. \tag{3.14}$$

Using the previously obtained λ_{jk}, we define $Z^*_{jik} = \big((\zeta_{jik}\hat{X}_{jik} + \dot{O}_{ji}(\beta, t_{jk}, \theta_{jW}))^\top, Z_{ji}^\top\big)^\top$. We then update β and γ using the one-step Newton-Raphson method:

$$\sum_{j=1}^{J}\sum_{i=1}^{n_j}\omega_{ji}\Bigg(\sum_{k=1}^{m_j} 1^*_{jik}\hat{\mathbb{E}}(\rho_{jik})\Bigg[Z^*_{jik} - \frac{\sum_{l=1}^{n_j}\omega_{jl}1^*_{jlk}e^{\beta\zeta_{jlk}\hat{X}_{jlk}+\dot{O}_j(\beta,t_{jk},\theta_{jW})+\gamma^\top Z_{jlk}}Z^*_{jlk}}{\sum_{l=1}^{n_j}\omega_{jl}1^*_{jlk}e^{\beta\zeta_{jlk}\hat{X}_{jlk}+\dot{O}_j(\beta,t_{jk},\theta_{W_j})+\gamma^\top Z_{jlk}}}\Bigg]\Bigg) = 0.$$

In E-step, we find $\hat{\mathbb{E}}(\rho_{ji,k})$ which is the posterior mean of ρ_{ji} conditional on the observed data. If $\Delta_{1ji} = 1$, we have

$$\begin{aligned}\hat{\mathbb{E}}(\rho_{jik}) &= E(\rho_{jik}|A_{ji} = 0, B_{ji} = 1)\\ &= \begin{cases}1, & v_{jir} < t_{jk} = T_{ji}\\ 0, & v_{jik} < t_{jk} < T_{ji}\end{cases}\end{aligned}$$

For the observation with $\Delta_{2ji} = 1$, i.e. $v_{jir} \le t_{jk}$ and $L_{ji} < t_{jk} < R_{ji} < \infty$, then

$$\begin{aligned}\hat{\mathbb{E}}(\rho_{jik}) &= E(\rho_{jik}|\tilde{v}_{ji}, L_{ji}, R_{ji}, Z_{ji}, C_{ji} = 1, D_{ji} > 0)\\ &= \sum_{m=0}^{\infty} mP(\rho_{jik} = m|\tilde{v}_{ji}, L_{ji}, R_{ji}, Z_{ji}, C_{ji} = 1, D_{ji} > 0)\\ &= \sum_{m=0}^{\infty} m\frac{(\lambda_{jk}\exp\{h_{ji}(t_{jk},\beta,\gamma)\})^m\exp\{-\lambda_{jk}\exp\{h_{ji}(t_{jk},\beta,\gamma)\}\}/m!}{1-\exp\{-\sum_{L_{ji}<t_{jk}\le R_{ji}}1^*_{jik}\lambda_{jk}\exp\{h_{ji}(t_{jk},\beta,\gamma)\}\}}\\ &= \frac{\lambda_{jk}\exp\{h_{ji}(t_{jk},\beta,\gamma)\}}{1-\exp\{-\sum_{L_{ji}<t_{jk}\le R_{ji}}1^*_{jik}\lambda_{jk}\exp\{h_{ji}(t_{jk},\beta,\gamma)\}\}}\end{aligned} \tag{3.15}$$

If $v_{jir} \le t_{jk} \le L_{ji}$, it follows that $\hat{\mathbb{E}}(\rho_{jik}) = 0$. We obtain the estimator of $(\lambda_{jk}, j = 1, ..., J; i = 1, ..., n_j)$ and (β, γ) by iterating between the E steps and M steps until convergence and denote the final estimator of $(\hat{\lambda}_{jk}, j = 1, ..., J; i = 1, ..., n_j)$ and $(\hat{\beta}, \hat{\gamma})$. This EM procedure assumes that the measurement error model parameters θ_{jW} are known. These parameters can be estimated by existing methods for estimating linear mixed effects model. Similar to Sun et al. (2023), we use R package `lme4` to obtain the maximum likelihood estimates of $\hat{\theta}_{jW}$. The θ_{jW} is replaced by $\hat{\theta}_{jW}$ when we conducting the aforementioned EM algorithm.

3.5 Variance Estimation

The variance estimators of $\hat{\beta}, \hat{\gamma}, \hat{\Lambda}$ are obtained from weighted bootstrap which is similar to the procedure introduced in section 2.4. The weighted bootstrap applying different weights to the log-likelihood function with weights from a distribution with mean and variance equal to 1. The procedure of weighted bootstrap described as follows:

- Generate a sequence IID random variables u_{ji} from $Exp(1)$ and let $U = \{u_{ji}, j = 1 \cdots J, i = 1 \cdots n_j\}$
- Use U to obtain perturbed weights $\omega^*_{ji} = \frac{\eta_{ji}}{\hat{p}^*_{ji}}$ where $\hat{p}^*_{ji}$ is obtained from the following procedures:
  - Fit a logistic regression model using the weighted log-likelihood with weight $\{u_{j1}, u_{j2}, \ldots, u_{jn_j}\}$. Let $\hat{\pi}_j = (\hat{\pi}_{0j}, \hat{\pi}_{1j})$ be the coefficients of the fitted logistic regression model, i.e.

$$\hat{\pi}_j = \arg\max_{\pi_j} \sum_{i=1}^{n_j} u_{ji} \Big\{ \eta_{ji}(\pi_{0j} + \pi_{1j} z_{ji}) - \log[1 + \exp(\pi_{0j} + \pi_{1j} z_{ji})] \Big\}$$

  - Then $\hat{p}^*_{ji} = \frac{\exp(\hat{\pi}_{0j} + \hat{\pi}_{1j} z_{ji})}{1+\exp(\hat{\pi}_{0j} + \hat{\pi}_{1j} z_{ji})}$ (Notice that when the missingness of X does not depend on Z, then $\hat{p}^*_{ji} = \frac{\exp(\hat{\pi}_{0j})}{1+\exp(\hat{\pi}_{0j})} = \frac{\sum_{i=1}^{n_j} u_{ji}\eta_{ji}}{\sum_{i=1}^{n_j} u_{ji}}$)
- With the perturbed inverse probability weight, we can set up our weighted complete data log-likelihood (perturbed)

$$l^*_n(\beta, \Lambda|\theta) = \sum_{j=1}^{J} \sum_{i=1}^{n_j} u_{ji}\omega_{ji} \Bigg(\sum_{k=1}^{m_j} 1^*_{jik} \Big[\rho_{jik} \log\{\mu_{jik}\} - \log(\rho_{jik}!) - \mu_{jik} \Big] \Bigg)$$

  where $R^*_{ji} = \Delta_{1ji} T_{ji} + \Delta_{2ji} R_{ji} + (1 - \Delta_{1ji} - \Delta_{2ji}) L_{ji}$
- Use the EM procedure introduced before with

$$\lambda^*_{jk} = \frac{\sum_{i=1}^{n_j} u_{ji}\omega_{ji} 1^*_{jik} \hat{\mathbb{E}}(\rho_{jik})}{\sum_{i=1}^{n_j} u_{ji}\omega_{ji} 1^*_{jik} e^{\beta\zeta_{jik}\hat{X}_{jik} + \gamma^\top Z_{ji} + O_{ji}(\beta, t_{jk}, \theta_{jW})}} \quad k = 1, 2, ..., m_j$$

- $$\sum_{j=1}^{J} \sum_{i=1}^{n_j} u_{ji}\omega_{ji} \Bigg(\sum_{k=1}^{m_j} 1^*_{jik} \hat{\mathbb{E}}(\rho_{jik}) [Z^*_{jik} - \frac{\sum_{l=1}^{n_j} u_{jl}\omega_{jl} 1^*_{jlk} e^{\beta\zeta_{jlk}\hat{X}_{jlk} + \dot{O}_j(\beta, t_{jk}, \theta_{jW}) + \gamma Z_{jlk}} Z^*_{jlk}}{\sum_{l=1}^{n_j} u_{jl}\omega_{jl} 1^*_{jlk} e^{\beta\zeta_{jlk}\hat{X}_{jlk} + \dot{O}_j(\beta, t_{jk}, \theta_{W_j}) + \gamma Z_{jlk}}}] \Bigg) = 0 \tag{3.16}$$

- Repeat the above procedure B times to get an B different $\hat{\beta}^*, \hat{\gamma}^*, \hat{\lambda}^*$s.
  - The standard error estimator of $\hat{\beta}, \hat{\gamma}$ are the sample standard deviation of those B $\hat{\beta}^*, \hat{\gamma}^*$s.

– In each iteration of the procedure, the function $\hat{\lambda}^*$ is smoothed over the interval $[0, \tau]$ using Gaussian kernel with a bandwidth 0.1.This smooth process yields a smoothed function, denoted as $\hat{\lambda}^{*'}$. The standard error estimator for $\hat{\lambda}$ is then calculated as the sample standard deviation of these B $\hat{\lambda}^{*'}$ evaluated at time points $0 = t_1, t_2, \ldots, t_M = \tau$ where the difference between two consecutive time points is 0.01.

3.6 Estimation Based on the Entire Trajectory of Longitudinal Covariates

When estimating $X_{ji}(t)$, we rely on information available up to time t. This approach can result in significant estimation bias and high variance, particularly when t is small, due to the limited data available at early time points. By utilizing the entire time trajectory of X_{ji} up to the final time τ, we can draw on a more complete set of information, leading to more accurate and reliable estimates. The following section outlines the rationale behind this approach. Recall the ordinary least square estimation of in equation (3.3), since we use whole information of ith individual in the jth stratum, it becomes

$$\hat{\theta}_{ji} = \left(\tilde{F}_{ji}^\top(\tau)\tilde{F}_{ji}(\tau)\right)^{-1}\tilde{F}_{ji}^\top(\tau)\tilde{W}_{ji}(\tau) \tag{3.17}$$

which is independent of time t. Let $\hat{X}_{ji}(t,\tau) = f^\top(t)\hat{\theta}_{ji}$ and $d_{ji}(t,\tau,\sigma_j^2) = \sigma_j^2 f^\top(t)\{\tilde{F}_{ji}^\top(\tau)\tilde{F}_{ji}(\tau)\}^{-1}f(t)$, then the proposition (1) still hold and the induced hazard function (3.5) becomes

$$\lambda_j^*(t|\hat{X}_{ji}(t,\tau), Z_{ji}, \tilde{v}_{ji}(t)) = \lambda_{0j}(t)\exp\left\{\beta\zeta_{ji}(t,\tau)\hat{X}_{ji}(t,\tau) + \gamma^\top Z_{ji} + O_{ji}(\beta, t, \tau, \theta_{jW})\right\} \tag{3.18}$$

with $\sigma^2_{ji,rel}(t,\tau) = \frac{d_{ji}(t,\tau,\sigma_j^2)}{f(t)^\top G_j f(t) + d_{ji}(t,\tau,\sigma_j^2)}$, $\zeta_{ji}(t,\tau) = 1 - \sigma^2_{ji,rel}(t,\tau) = \frac{f^\top(t)G_jf(t)}{f^\top(t)G_jf(t)+d_{ji}(t,\tau,\sigma_j^2)}$, $O_{ji}(\beta,t,\tau,\theta_W) = \beta\{\vartheta_j f(t) + \frac{1}{2}\beta f^\top(t)G_jf(t)\}\sigma^2_{ji,rel}(t,\tau)$, $\theta_{jW} = (\vartheta_j, G_j, \sigma_j^2)$. The proof is similar to the proof in proposition (1). Since $\hat{X}_{ji}(t,\tau) = X_{ji}(t) + f^\top(t)\{\tilde{F}_{ji}^\top(\tau)\tilde{F}_{ji}(\tau)\}^{-1}\tilde{F}_{ji}^\top(\tau)\tilde{e}_{ji}(t)$, from property of conditional mean and variance of normal distributed variable, we have

$$\mathrm{E}(X_{ji}(t)|\hat{X}_{ji}(t,\tau),\tilde{v}_{ji}(t)) = \frac{E(X_{ji}(t))d_{ji}(t,\tau,\sigma_j^2) + \hat{X}_{ji}(t,\tau)f^\top(t)G_jf^\top(t)}{f^\top(t)G_jf^\top(t) + d_{ji}(t,\tau,\sigma_j^2)}$$

$$\mathrm{Var}(X_{ji}(t)|\hat{X}_{ji}(t,\tau),\tilde{v}_{ji}(t)) = \frac{f^\top(t)G_jf^\top(t)d_{ji}(t,\tau,\sigma_j^2)}{f^\top(t)G_jf^\top(t) + d_{ji}(t,\tau,\sigma_j^2)}$$

The conditional hazard function of T_{ji} at time t given $(\hat{X}_{ji}(t), Z_{ji}, \tilde{v}_{ji}(t), \tau)$ equals

$$\begin{aligned}
&\lambda_j^*(t|\hat{X}_{ji}(t,\tau), Z_{ji}, \tilde{v}_{ji}(t)) \\
&= \lambda_{0j}(t)\exp\{\gamma^\top Z_{ji}\}E\{\exp(\beta X_{ji}(t))|\hat{X}_{ji}(t,\tau), \tilde{v}_{ji}(t)\} \\
&= \lambda_{0j}(t)\exp\left\{\gamma^\top Z_{ji} + \beta\frac{E(X_{ji}(t))d_{ji}(t,\tau,\sigma_j^2) + \hat{X}_{ji}(t)f(t)^\top G_j f(t)}{f(t)^\top G_j f(t) + d_{ji}(t,\tau,\sigma_j^2)} + \frac{1}{2}\beta^2\frac{f(t)^\top G_j f(t) d_{ji}(t,\tau,\sigma_j^2)}{f(t)^\top G_j f(t) + d_{ji}(t,\tau,\sigma_j^2)}\right\} \\
&= \lambda_{0j}(t)\exp\left\{\beta\hat{X}_{ji}(t)\frac{f(t)^\top G_j f(t)}{f(t)^\top G_j f(t) + d_{ji}(t,\tau,\sigma_j^2)} + \gamma^\top Z_{ji} + \beta\{\vartheta_j f(t)\right. \\
&\left. + \frac{1}{2}\beta f(t)^\top G_j f(t)\}\frac{d_{ji}(t,\tau,\sigma_j^2)}{f(t)^\top G_j f(t) + d_{ji}(t,\tau,\sigma_j^2)}\right\} \\
&= \lambda_{0j}(t)\exp\left\{\beta\zeta_{ji}(t,\tau)\hat{X}_{ji}(t) + \gamma^\top Z_{ji} + O_{ji}(\beta, t, \tau, \theta_{jW})\right\}
\end{aligned}$$

thus we obtain the induced hazard function (3.18). Model estimation and variance estimation procedure are the same as we described in section 3.4 and 3.5 except we use the measurement error induced hazard function (3.18).

3.7 Simulation Studies

We examine the finite sample properties of proposed method via simulation studies. Let n be the sample size. For $i = 1, ..., n$, the failure time T_{ji} is generated from the proportional hazards model

$$\lambda_j(t|X_{ji}(t), Z_{ji}) = \lambda_j(t)\exp\{\beta X_{ji}(t) + \gamma^\top Z_{ji}\} \tag{3.19}$$

Let $\beta = 0.5$, $\gamma = -\log(2)$, $Z_{ji} \sim Ber(0.3)$ and $X_{ji} = (\nu_0 + b_{0ji}) + (\nu_1 + b_{1ji})t$. The partly interval censored data for the i-th individual in the j-th stratum generated as follows. We first generate the number of examination times $K \sim Ber(0.8) + 1$. If $K = 1$, we generated a single examination time $U_1 \sim Unif(0, 3\tau/4)$, where $(L, R]$ intervals were defined as $(0, U_1]$ if $T \leq U_1$ and (U_1, ∞) if $T > U_1$. For $K = 2$, we generated two examination times U_1 and U_2, with U_2 being $\min\{0.1 + U_1 + \exp(1)\tau/2, \tau\}$. Define $(L_{ji}, R_{ji}] = (0, U_{ji1}]$ if $T_{ji} \leq U_{ji1}$, $(L_{ji}, R_{ji}] = (U_{1ji}, U_{2ji}]$ if $U_{1ji} < T_{1ji} \leq U_{2ji}$ and $(L_{1ji}, R_{1ji}] = (U_{2ji}, \infty)$ if $T_{1ji} > U_{2ji}$. If $R = \infty$, we have $\Delta_{1ji} = \Delta_{2ji} = 0$; If $R_{ji} < \infty$, we generate $\Delta_{1ji} \sim Ber(p)$ with $p = 0.25$ or $p = 0.75$. If $\Delta_{1ji} = 1$, then the failure time T_{1ji} is observed exactly. We set the length of study $\tau = 5$ which yielding about 40% percent right censoring. The error-prone measurements $W_{ji,b}(v_{jik})$ are generated from the model

$$W_{ji,b} = X_{ji}(v_{jik}) + e_{jik,b}, \quad b = 1, ..., B_{jik}, \tag{3.20}$$

where $X_{ji}(v_{jik}) = (\nu_{0j} + b_{0ji}) + (\nu_{1j} + b_{1ji})v_{jik}$ which is specified before. We let $e_{jik,b} \sim N(0, \sigma^2)$ with $\sigma = 0.1$ or $\sigma = 0.2$ and the number of repeated measurements of $X_{ji}(v_{jik})$ is $B_{jik} = B = 1$ for all j, i, k. The missing model and baseline hazards have the following settings:

1. Assume $J = 2$ and $\lambda_1(t) = 0.1t$ and $\lambda_2(t) = 0.3$. The fixed effects are taken as $\nu_{01} = \nu_{02} = 1$ and $\nu_{11} = \nu_{12} = 0.5$. The random effect $(b_{0ji}, b_{1ji}) \sim N(0, G)$. We set $G = [0.02, -0.01; -0.01, 0.02]$. The selection probability for non-case (right censored observation) is 0.3 and selection probability for case is 0.9. Overall missing rate about 33%. Results are shown in Table 5.1.

2. Assume $J = 2$ and $\lambda_1(t) = 0.1t$ and $\lambda_2(t) = 0.3$. The fixed effects are taken as $\nu_{01} = \nu_{02} = 1$ and $\nu_{11} = \nu_{12} = 0.5$. The random effect $(b_{0ji}, b_{1ji}) \sim N(0, G)$. We set $G = [0.02, -0.01; -0.01, 0.02]$. For noncase, the selection probability for the first stratum is defined as $q_1(\eta = 1|Z) = \frac{\exp(0.1-Z)}{1+\exp(0.1-Z)}$ and for the second stratum, the selection probability defined as $q_2(\eta = 1|Z) = \frac{\exp(0.3-1.2Z)}{1+\exp(0.3-1.2Z)}$. For case, $\eta \sim Ber(0.9)$, each observation has 90% chance to be selected. The overall missing rate is around 30%. Results are shown in Table 5.2.

3. Assume $J = 2$ and $\lambda_1(t) = 0.1t$ and $\lambda_2(t) = 0.3$. The fixed effects are taken as $(\nu_{01}, \nu_{11}) = (0.85, 0.4)$ and $(\nu_{02}, \nu_{12}) = (1.15, 0.5)$. The random effects $(b_{0ji}, b_{1ji}) \sim N(0, G)$ and $G = [0.02, -0.01; -0.01, 0.02]$ for $j = 1, 2$. For noncase, the selection probability for the first stratum is defined as $q_1(\eta = 1|Z) = \frac{\exp(0.1-Z)}{1+\exp(0.1-Z)}$ and for the second stratum, the selection probability defined as $q_2(\eta = 1|Z) = \frac{\exp(0.3-1.2Z)}{1+\exp(0.3-1.2Z)}$. For case, $\eta \sim Ber(0.9)$, each observation has 90% chance to be selected. The overall missing rate is around 30%. Results are shown in Table 5.3.

4. Assume $J = 2$ and $\lambda_1(t) = 0.1t$ and $\lambda_2(t) = 0.3$. The fixed effects are taken as $(\nu_{01}, \nu_{11}) = (0.85, 0.4)$ and $(\nu_{02}, \nu_{12}) = (1.15, 0.5)$. The random effects $(b_{0ji}, b_{1ji}) \sim N(0, G)$ and $G = [0.02, -0.01; -0.01, 0.02]$ for $j = 1, 2$. For noncase, the selection probability for the first stratum

is defined as $q_1(\eta = 1|Z) = \frac{\exp(0.1-Z)}{1+\exp(0.1-Z)}$ and for the second stratum, the selection probability defined as $q_2(\eta = 1|Z) = \frac{\exp(0.3-1.2Z)}{1+\exp(0.3-1.2Z)}$. For case, $\eta \sim Ber(0.9)$, each observation has 90% chance to be selected. The overall missing rate is around 30%. In this scenario, we estimate $X_{ji}(t)$ by the observed measurements before τ. Results are shown in Table 5.4.

We simulated with sample size $n = 800$ and 1200. The estimation results for (β, γ) based on 500 simulations. We perform 200 bootstrap for each simulation. The Bias is the average point estimate minus the true parameter value, SSD is the sample standard deviation of point estimates, ESE is the average of estimated standard errors and CP is the coverage proportion of the 95% confidence interval. CP is the coverage proportion of the 95% confidence interval of $\hat{\lambda}(t)$. The results presented in Tables 5.1 to 5.4 demonstrate the following key findings: (i) The proposed estimators exhibit virtually no bias, indicating their accuracy in capturing the true parameters. (ii) The weighted bootstrap method yields standard error estimates that consistently and reliably reflect the true variability of the estimators. (iii) The empirical coverage rates of the 95% confidence intervals, constructed using the normal approximation, are consistently close to the nominal 95% level, suggesting the validity of the proposed method. (iv) As the sample size increases, both the bias and variability of the estimators decrease. Furthermore, an increase in the proportion of exact observations, p_t, leads to a reduction in the standard deviation of the estimators. (v) Our proposed approach (Simulation Setup 4), which leverages the complete available information for each X_{ji} to estimate $\hat{X}_{ji}(t)$, demonstrates superior performance in model estimation. Compared to the method that estimates $\hat{X}_{ji}(t)$ using only time points prior to t (Simulation Setup 3), this approach achieves a smaller standard deviation, resulting in more stable estimation of model coefficients and enhanced precision in the results. Additionally, it provides an alternative and effective way of estimating $X_{ji}(t)$, further contributing to the robustness of the estimation process.

3.8 Real Data Application

We analyze data from the HVTN-703/704 trial, where the longitudinal covariate HIV-1 antibody VRC (measured as logVRC) is subject to measurement error. Let $X(t)$ denote the time-dependent logVRC value, Z represent the time-independent covariate Age group (categorized as described in

Chapter 2), and T be the time from enrollment to HIV onset. The model is stratified by geographic region, with four strata ($j = 1, 2, 3, 4$) corresponding to USAS, BP, SSA, and other SSA, respectively, as defined in Chapter 2. We assume the conditional hazard function for T given $X(t)$ and Z in the j-th stratum follows a Cox proportional hazards model:

$$\lambda_j(t|X(t), Z) = \lambda_j(t) \exp\left\{\beta X(t) + \gamma^\top Z\right\}, \tag{3.21}$$

where $\lambda_j(t)$ is the baseline hazard function for the j-th stratum, and β and γ are regression coefficients. These coefficients represent log hazard ratios, quantifying the association between HIV onset time and (i) logVRC (time-varying) and (ii) Age group (time-independent), respectively. The true HIV-1 antibody VRC trajectory, denoted $X(t)$, is unobservable; instead, we observe $W(t)$, an error-contaminated surrogate measurement of $X(t)$. For the observed concentration, we employed the last value carry forward (LVCF) approach to extend the observed measurement $W(t)$ to the primary endpoint for each participant, defined as either HIV infection for cases or the end of the study period for non-cases. We use full information about log(VRC) in estimating the $X(t)$, the true value of VRC. This estimation approach provides an alternative, especially when data is limited. While using available information about $X(t)$ up to time t in risk prediction is more interpretable, utilizing the full information of X can lead to a more stable model estimation. We justified that the induced hazard model remains valid when using full information about X (section 3.6), and simulations demonstrate that this approach improves estimation efficiency and works well (simulation setting 4). Next we describe the estimation of $X(t)$ using spline-based method. Suppose a and b are the lower and upper bounds of the observation times $\{(U_i, V_i) : i = 1, 2, ..., n\}$. Let $a = d_0 < d_1 < \cdots < d_K < d_{K+1} = b$ be a partition of $[a, b]$ into $K + 1$ sub-intervals $I_{K_t} = [d_t, d_{t+1}), \quad t = 0, ..., K$. Denote the set of partition points by $D_n = \{d_1, d_2, \cdots, d_K\}$. According to Schumaker (1981), (corollary 4.10) and Zhang et al. (2010), there exists a local basis $\mathcal{B}_n = \{b_t, 1 \leq t \leq q_n\}$, B-spline, and the number basis be $q_n \equiv K_n + m$. These basis functions are nonnegative and sum up to one at each point within $[a, b]$ and zero outside the interval $[d_t, d_{t+m}]$. From the log(VRC) concentration plot, Fig 2.1, we observe peak points before week 20, followed by a flat region. Therefore, we select 4 internal knots at weeks 5, 9, 13 and 17 to capture the key dynamics of the curve, ensuring an accurate representation of the rise and subsequent flattening. Therefore, the sequence of partitioning intervals are $[0,5],[5,9],[9,13],[13,17],[17,\tau]$. We choose the order of polynomial at each partition interval to be three. Therefore, we need seven B-spline basis as described in Schumaker (1981). In particular, let $b_1,...,b_8$ be a set of B-spline functions of order 3 on a knot sequence $\{d_0=0,d_1,d_2,d_3,d_4,d_5=\tau\}$. The B-spline basis functions are calculated recursively. For a B-spline of degree l and a knot sequence $0=d_0<d_1<d_2<\ldots<d_K<d_{K+1}=\tau$, the basis functions are calculated as follows:

- For $l=0$, the basis functions are piecewise constant and defined as follows:

$$b_i^0(t)=\begin{cases}1 & \text{if } d_i\le t<d_{i+1}\\ 0 & \text{otherwise}\end{cases}$$

- For $l>0$, the basis functions are recursively defined using the Cox-de Boor formula:

$$b_i^l(t)=\frac{t-d_i}{d_{i+l}-d_i}b_i^{l-1}(t)+\frac{d_{i+l+1}-t}{d_{i+l+1}-d_{i+1}}b_{i+1}^{l-1}(t)$$

The log(VRC) concentration, X, is the linear combination of these B-spline basis. Denote the estimated X from eight B-spline basis as $\hat{X}$ and we obtain it by regressing $W(t)$, the observed concentration of log (VRC), over B-spline basis, $b_1\cdots b_8$, using the least square approach. The fitting results are shown in Figs (5.1)-(5.4). Next we fit a linear mixed effect model with all regions and basis functions. The fixed effect is fitted using region indicators and B-spline basis while the random effect is obtained from the B-spline basis.

$$\begin{aligned}W(t)&\sim \text{USAS}+\text{BP}+\text{SSA}+\text{B-spline basis}\\ \textit{random}&\sim \text{B-spline basis}\end{aligned}\tag{3.22}$$

where *random* represent random effect in linear mixed effect model. However, the model did not converge and we find that the region indicators are not significant instead of region BP. Fig. 3.1 shows that the BP region, represented by the red line, consistently exhibits lower mean and median observed log(VRC) values, respectively, compared to the other regions (USAS, SSA, and Other SSA). This difference is particularly notable around the peak, where the region BP remains significantly below the others. This distinct behavior suggests that BP has unique VRC dynamics that differ from the rest, which warrants its separation in subsequent analyzes to avoid potential confounding and to better understand the underlying factors driving these regional differences. Then we refit the linear mixed effect model which stratified by region Brazil&Peru, meaning we fit linear mixed effect model separately for region Brazil&Peru and other three regions. The model has the following format:

$$\begin{aligned} W(t) &\sim \text{B-spline basis} \\ random &\sim \text{B-spline basis} \end{aligned} \tag{3.23}$$

The regions USAS, SSA and other SSA share a linear mixed effect model while region BP has a unique a linear mixed effect model. The results of fitting model (3.23) for region non-BP and BP contains the fixed effects, random effects and residuals. The fixed effects are presented in Table 3.1 where we can see that all predictors are statistically significant. The random effects shown in Table 3.2. The fitted residues for non-BP region is 0.126 and BP region is 0.096. Additionally, Figure 3.2 suggests that the fixed effect of fitted model in Table 3.1 effectively captures the pattern of log(VRC) in both BP and non-BP regions. Next, we fit model (3.21) using the linear mixed-effects model results. The linear mixed effect model parameter θ_{ji} is estimated using R package `lme4`. Then the estimated logVRC is $\hat{X}_{ji}(t) = f^\top(t)\hat{\theta}_{ji}$. As in Chapter 2, VRC concentrations are unavailable for some participants across four geographic regions (USAS, BP, SSA, and other SSA). The missingness of VRC measurements is modeled via logistic regression in Equation (2.14). To address potential bias from missing data, we employ inverse probability weighting (IPW), incorporating selection probabilities derived from Equation (2.14). The Cox model parameters—regression coefficients (β, γ) and stratum-specific baseline hazard functions $\lambda_j(t)$—are then estimated using the EM algorithm outlined in Section 3.4. Standard errors for model coefficients are estimated using a weighted bootstrap with 500 repetitions. The model estimation results are summarized in Table 3.3, and the corresponding estimated survival function is presented in Figure 5.5. The findings reveal that higher VRC concentrations are associated with a reduced risk of HIV infection, while older individuals exhibit a lower likelihood of contracting HIV. These results are consistent with the findings obtained in Chapter 2. Subsequently, we fitted the proportional hazard model described in Chapter 2 using the estimated $\hat{X}_{ji}(t)$. The estimation results for this model are summarized in Table 3.4, and the estimated survival function is illustrated in Figure 5.7. The results demonstrate a similar pattern to those observed in the previous models.

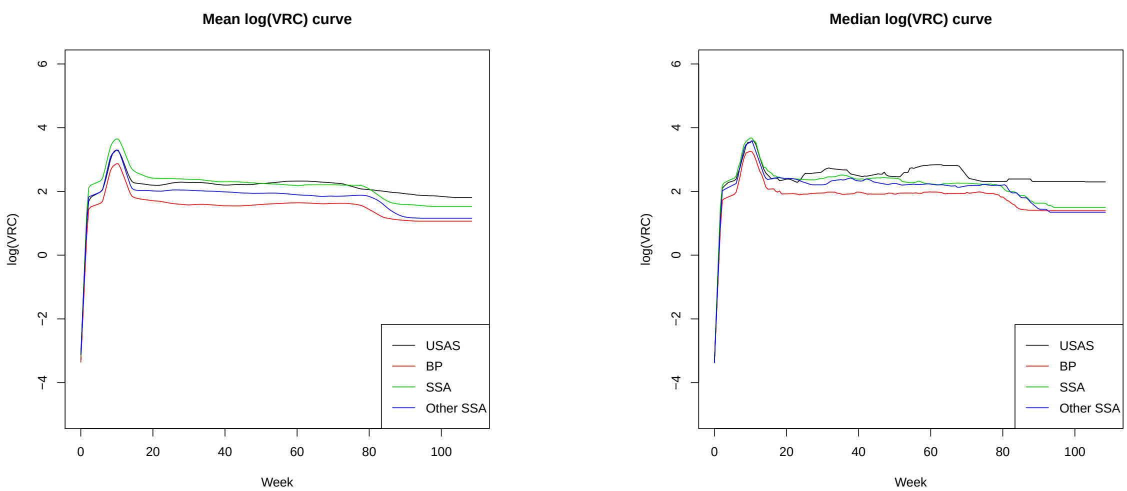

Figure 3.1: Mean and median of the observed log(VRC) curves for different regions. The left panel shows the mean of log(VRC) over time, while the right panel shows the median of log(VRC) over time.

Table 3.1: Estimation results of fixed effects in the linear mixed effects model for non-BP and BP regions.

	Estimate	Std. Error	t value
b_1	-3.18003	0.03247	-97.925
b_2	4.84639	0.13685	35.414
b_3	-0.16938	0.16439	-1.030
b_4	4.99466	0.08278	60.335
b_5	2.45836	0.08976	27.389
b_6	1.70035	0.19415	8.758
b_7	4.02896	0.34241	11.766
b_8	-1.43023	0.31835	-4.493

(a) Non-BP region

	Estimate	Std. Error	t value
b_1	-3.35331	0.03537	-94.808
b_2	4.13218	0.19136	21.594
b_3	-0.57333	0.22238	-2.578
b_4	4.33244	0.18376	23.577
b_5	1.97996	0.13946	14.197
b_6	0.64094	0.31334	2.045
b_7	4.01563	0.48645	8.255
b_8	-3.16073	0.47079	-6.714

(b) BP region

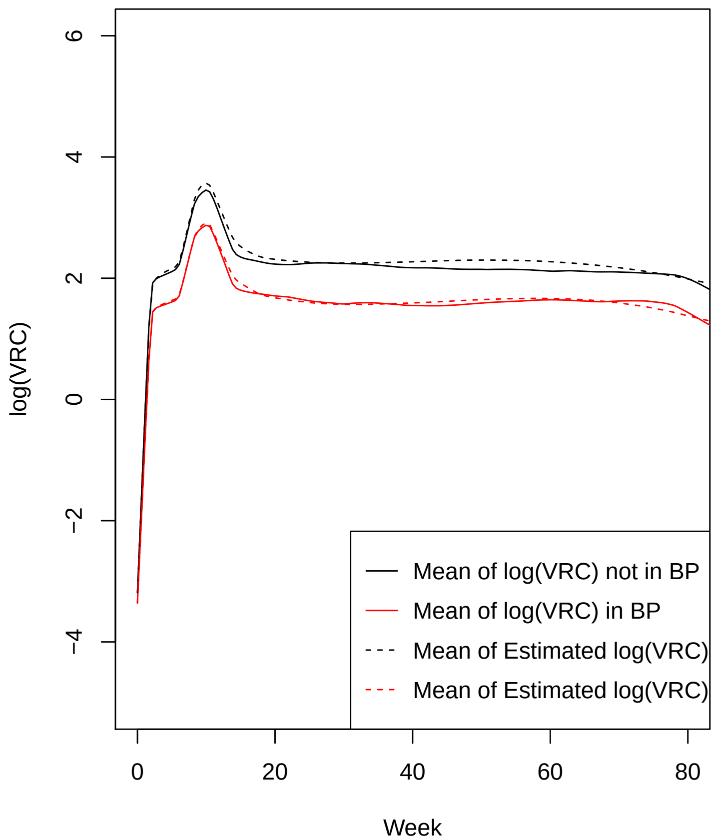

Figure 3.2: Mean log(VRC) curve estimated from the linear mixed effect model (3.23) v.s. true mean log(VRC) curve.

Table 3.2: Random effects matrices for non-BP and BP regions

(a) Non-BP region

	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BS8
BS1	0.00	-0.01	-0.01	0.00	0.00	-0.01	0.01	0.00
BS2	-0.01	1.64	1.91	0.30	0.76	1.14	-0.14	0.51
BS3	-0.01	1.91	2.37	0.41	0.75	1.68	-0.81	0.47
BS4	0.00	0.30	0.41	0.44	0.14	0.58	-0.10	0.12
BS5	0.00	0.76	0.75	0.14	0.79	0.19	1.22	-0.30
BS6	-0.01	1.14	1.68	0.58	0.19	3.10	-3.23	1.03
BS7	0.01	-0.14	-0.81	-0.10	1.22	-3.23	8.30	-2.07
BS8	0.00	0.51	0.47	0.12	-0.30	1.03	-2.07	3.43

(b) BP region

	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BS8
BS1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BS2	0.00	2.32	2.40	1.61	1.49	0.86	3.97	-1.24
BS3	0.00	2.40	3.07	1.44	1.88	1.24	5.00	-1.36
BS4	0.00	1.61	1.44	2.18	0.91	2.01	2.04	0.41
BS5	0.00	1.49	1.88	0.91	1.34	0.36	3.50	-0.91
BS6	0.00	0.86	1.24	2.01	0.36	5.70	-1.98	3.38
BS7	0.00	3.97	5.00	2.04	3.50	-1.98	12.48	-5.26
BS8	0.00	-1.24	-1.36	0.41	-0.91	3.38	-5.26	5.84

Table 3.3: Analysis of HIV Data

Trials	Covariates	Est.	SE	P.value
Combined	logVRC	-0.682	0.223	0.001
	Age $20-30$	-0.858	-0.511	0.093
	Age >30	-1.913	0.614	0.002
HVTN-703	logVRC	-0.434	0.290	0.135
	Age $20-30$	0.277	0.949	0.770
	Age >30	0.512	0.975	0.600
HVTN-704	logVRC	-0.728	0.245	0.003
	Age $20-30$	-0.942	0.643	0.142
	Age >30	-2.790	0.803	<0.001

Table 3.4: Analysis of HIV Data Using Estimated $\hat{X}_i(t)$ Based on the Method from Chapter 2

Trials	Covariates	Est.	SE	P.value
Combined	logVRC	-0.517	0.180	0.004
	Age $20-30$	-0.983	0.503	0.051
	Age >30	-2.154	0.639	<0.001
HVTN-703	logVRC	-0.520	0.313	0.099
	Age $20-30$	0.201	0.961	0.834
	Age >30	0.470	0.982	0.632
HVTN-704	logVRC	-1.274	0.479	0.008
	Age $20-30$	-1.594	0.983	0.105
	Age >30	-3.576	1.424	0.012

3.8.1 Remarks

We compare the data analysis results from three fitting methods: (1) Using the induced hazard model from Proposition 1; (2) Fitting the proportional hazard model of Chapter 2 using the observed $X_{ji}(t)$; (3) Fitting the proportional hazard model of Chapter 2 using the estimated $\hat{X}_i(t)$ The analysis involved fitting three distinct models to assess the relationship between covariates and the hazard of HIV infection. First, a model incorporating the proportion specified in Proposition 1 was fitted, with results presented in Table 3.3. Second, the proportional hazard model from Chapter 2 was fitted using the observed values $X_{ji}(t)$, yielding results in Table 2.3. Finally, the same proportional hazard model was fitted using the estimated values $\hat{X}_{ji}(t)$, with results detailed in Table 3.4. Across all models, the variable log(VRC) was consistently found to be statistically significant, underscoring its strong association with the hazard of HIV infection. Additionally, older age was associated with a lower risk of HIV infection in all three models. These consistent findings highlight the robustness of the results and reinforce the importance of log(VRC) and age as key predictors in the context of HIV risk.

3.9 Concluding Remarks

In this chapter, we study the Cox model with longitudinal covariates, say X, subject to measurement error and missingness. We applied linear mixed effects model to model longitudinal covariate measured with error and use inverse probability weighting approach to adjust for the bias introduced from missingness. For the estimated longitudinal covariate at time t, we proposed two ways of obtaining: (i) using the partial information before time t to get $\hat{X}(t)$; (ii)using complete information for each individual (before study end time τ). Both approaches are validated through theoretical justification and numerical studies. The measurement error-induced hazard was then applied to obtain the baseline hazard function. Then we devised a fast and stable EM approach, similar to Chapter 2, to obtain the model coefficient estimators and the baseline hazard estimators. The effectiveness of these approaches are supported by theoretical justification and numerous simulation studies. We then apply these techniques to HIV data, where the covariate log(VRC) is influenced by both missingness and measurement error. The estimated results closely align with those obtained under the assumption of no measurement error (as detailed in Chapter 2).

CHAPTER 4: SUPPLEMENTAL RESULTS FOR CHAPTER 2

4.1 EM Algorithm for Weighted Likelihood

For a single subject $Y = (Y_{obs}, Y_{miss})$. For the observed data, consider the weighted log-likelihood function

$$l^{\omega}(\theta) = \omega \log f(Y_{obs}, \theta),$$

where $f(Y_{obs}, \theta)$ is the observed-data likelihood and ω is a weight which depends only on Y_{obs}. Let $\hat{\theta}$ be the maximizer of $l^{\omega}(\theta)$, i.e. $\hat{\theta} = \arg\max_{\theta} l^{\omega}(\theta)$.

We develop an EM algorithm to find $\hat{\theta}$. Let Y_{miss} be the miss value of latent variable. Let $f(Y)$ be the complete data likelihood. At k-th iteration,

1. E-step : Compute $Q(\theta|\theta^{(k)}) = E(\omega \log f(Y;\theta)|Y_{obs}, \theta^{(k)})$
2. M-step : obtain $\theta^{(k+1)} = \arg\max_{\theta} Q(\theta|\theta^{(k)})$
3. iterative beween E- and M-steps until convergence

At each iteration of the EM algorithm, we have

Lemma 4.1. $l^{\omega}(\theta^{(k+1)}) = \omega \log f(Y_{obs}; \theta^{(k+1)}) \geq \omega \log f(Y_{obs}; \theta^{(k)}) = l^{\omega}(\theta^{(k)})$

Proof. According to the M-step, we have $Q(\theta^{(k+1)}|\theta^{(k)}) \geq Q(\theta^{(k)}|\theta^{(k)})$, that is

$$E(\omega \log f(Y; \theta^{(k+1)})|Y_{obs}, \theta^{(k)}) \geq E(\omega \log f(Y; \theta^{(k)})|Y_{obs}, \theta^{(k)})$$

Notice that $f(Y;\theta) = f(Y_{miss}|Y_{obs};\theta)f(Y_{obs};\theta)$, we have

$$\begin{aligned}
E(\omega \log f(Y;\theta^{(k+1)})|Y_{obs},\theta^{(k)}) &= E(\omega \log f(Y_{miss}|Y_{obs};\theta^{(k+1)})f(Y_{obs};\theta^{(k+1)})|Y_{obs},\theta^{(k)}) \\
&= E(\omega \log f(Y_{miss}|Y_{obs};\theta^{(k+1)})|Y_{obs},\theta^{(k)}) + \omega \log f(Y_{obs};\theta^{(k+1)}) \\
&\geq E(\omega \log f(Y_{miss}|Y_{obs};\theta^{(k)})|Y_{obs},\theta^{(k)}) + E(\omega \log f(Y_{obs};\theta^{(k)})|Y_{obs},\theta^{(k)}) \\
&= E(\omega \log f(Y_{miss}|Y_{obs};\theta^{(k)})|Y_{obs},\theta^{(k)}) + \omega \log f(Y_{obs};\theta^{(k)})
\end{aligned} \tag{4.1}$$

From the property of Kullback-Leibler divergence,

$$E(\log f(Y_{miss}|Y_{obs};\theta^{(k+1)})|Y_{obs},\theta^{(k)}) \leq E(\log f(Y_{miss}|Y_{obs};\theta^{(k)})|Y_{obs},\theta^{(k)})$$

Also notice that ω depends only on the Y_{obs} and is non-negative, we have

$$E(\omega \log f(Y_{miss}|Y_{obs};\theta^{(k+1)})|Y_{obs},\theta^{(k)}) \leq E(\omega \log f(Y_{miss}|Y_{obs};\theta^{(k)})|Y_{obs},\theta^{(k)}) \tag{4.2}$$

Combining inequalities (4.1), (4.2), we have

$$\omega \log f(Y_{obs};\theta^{(k+1)}) \geq \omega \log f(Y_{obs};\theta^{(k)})$$

which means the weighted data likelihood function are non-decreasing at each iteration and equality holds if and only if $\theta^{(k+1)} = \theta^{(k)}$ from identifiability. To avoid local maxima, we suggest using a set of different initial values of $\theta^{(0)}$. □

4.2 Some Derivation Details

$$\begin{aligned}
P(A_{ji}=0, B_{ji}=1|\xi_{ji}) &= P(\sum_{t_{j_k}<T_{ji}} W_{jik}=0|\xi_{ji})P(I(t_{j_k}=T_{ji})W_{jik}=1|\xi_{ji}) \\
&= \exp\{\xi_{ji}\sum_{t_{j_k}<T_{ji}} \lambda_k e^{-\beta^\top X_{jiT}-\gamma^\top Z_{jiT}}\}\exp\{\xi_{ji}\lambda_{T_{ji}}e^{-\beta^\top X_{jiT}-\gamma^\top Z_{jiT}}\}\xi_{ji}\lambda_{T_{ji}}e^{-\beta^\top X_{jiT}-\gamma^\top Z_{jiT}} \\
&= \xi_{ji}\exp\{-\xi_{ji}S_{jiT}\}\lambda_{T_{ji}}\exp\{-\beta^\top X_{jiT}-\gamma^\top Z_{jiT}\}
\end{aligned}$$

The unconditioned probability is

$$
\begin{aligned}
P(A_{ji}=0,B_{ji}=1) &= \int_{\xi_{ji}} P(A_{ji}=0,B_{ji}=1|\xi_{ji})f(\xi_{ji})d\xi_{ji} \\
&= \int_{\xi_{ji}} \xi_{ji}\exp\{-\xi_{ji}S_{jiT}\}\lambda_{T_{ji}}\exp\{-\beta^\top X_{jiT}-\gamma^\top Z_{jiT}\}f(\xi_{ji})d\xi_{ji} \\
&= \lambda_{T_{ji}}\exp\{-\beta^\top X_{jiT}-\gamma^\top Z_{jiT}\}\int_{\xi_{ji}} \xi_{ji}\exp\{-\xi_{ji}S_{jiT}\}f(\xi_{ji})d\xi_{ji} \\
&= \lambda_{T_{ji}}\exp\{-\beta^\top X_{jiT}-\gamma^\top Z_{jiT}\}G^{'}(S_{jiT})\exp\{-G(S_{jiT})\}
\end{aligned}
$$

The the conditional distribution of ξ_{ji} given observations is

$$
P(\xi_{ji}|A_{ji}=0,B_{ji}=1) = \frac{P(A_{ji}=0,B_{ji}=1|\xi_{ji})f(\xi_{ji})}{P(A_{ji}=0,B_{ji}=1)} = \frac{\xi_{ji}\exp\{-\xi_{ji}S_{jiT}\}f(\xi_{ji})}{G^{'}(S_{jiT})\exp\{-G(S_{jiT})\}}
$$

Thus the conditional expectation of ξ_{ji} given observations is (consider whether $\eta_{ji}\perp\!\!\!\perp W_{jik}|(X_{ji},Z_{ji})$)

$$
\begin{aligned}
\hat{\mathbb{E}}(\xi_{ji}) &= E(\xi_{ji}|A_{ji}=0,B_{ji}=1) \\
&= \int_{\xi_{ji}} \xi_{ji}P(\xi_{ji}|A_{ji}=0,B_{ji}=1)d\xi_{ji} \\
&= \int_{\xi_{ji}} \frac{\xi_{ji}^2\exp\{-\xi_{ji}S_{jiT}\}f(\xi_{ji})}{G^{'}(S_{jiT})\exp\{-G(S_{jiT})\}}d\xi_{ji} \\
&= \frac{1}{G^{'}(S_{jiT})\exp\{-G(S_{jiT})\}}\int_{\xi_{ji}} \xi_{ji}^2\exp\{-\xi_{ji}S_{jiT}\}f(\xi_{ji})d\xi_{ji} \\
&= \frac{\exp\{G(S_{jiT})\}}{G^{'}(S_{jiT})}\left[(G^{'}(S_{jiT}))^2\exp\{-G(S_{jiT})\} - G^{''}(S_{jiT})\exp\{-G(S_{jiT})\right] \\
&= G^{'}(S_{jiT}) - \frac{G^{''}(S_{jiT})}{G^{'}(S_{jiT})}
\end{aligned}
$$

If $\Delta_{2ji}=1$, **strictly interval censored**, the posterior distribution of the given observation, $f(\xi_{ji}|O_{ji},\Delta_{2ji}=1)=f(\xi_{ji}|C_{ji}=0,D_{ji}>0)$.

Notice that

$$
P(C_{ji}=0,D_{ji}>0) = \int_{\xi_{ji}} P(C_{ji}=0,D_{ji}>0|\xi_{ji})f(\xi_{ji})d\xi_{ji}
$$

and

$$
\begin{aligned}
P(C_{ji}=0,D_{ji}>0|\xi_{ji}) &= P(C_{ji}=0|\xi_{ji}) - P(C_{ji}=0,D_{ji}=0|\xi_{ji}) \\
&= P(\sum_{t_{j_k}\le L_{ji}} W_{jik}=0|\xi_{ji}) - P(\sum_{t_{j_k}\le R_{ji}} W_{jik}=0|\xi_{ji}) \\
&= \exp\{-\xi_{ji}\sum_{t_{j_k}\le L_{ji}} \lambda_{jk} e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\} - \exp\{-\xi_{ji}\sum_{t_{j_k}\le R_{ji}} \lambda_{jk} e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\} \\
&= \exp\{-\xi_{ji}S_{jiL}\} - \exp\{-\xi_{ji}S_{jiR}\}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
P(C_{ji}=0,D_{ji}>0) &= \int_{\xi_{ji}} P(C_{ji}=0,D_{ji}>0|\xi_{ji}) f(\xi_{ji}) d\xi_{ji} \\
&= \int_{\xi_{ji}} (\exp\{-\xi_{ji}S_{jiL}\} - \exp\{-\xi_{ji}S_{jiR}\}) f(\xi_{ji}) d\xi_{ji} \\
&= \exp\{-G(S_{jiL})\} - \exp\{-G(S_{jiR})\}
\end{aligned}
$$

Thus, the posterior distribution of the given observation,

$$
\begin{aligned}
f(\xi_{ji}|O_{ji},\Delta_{2ji}=1) &= f(\xi_{ji}|C_{ji}=0,D_{ji}>0) \\
&= \frac{f(C_{ji}=0,D_{ji}>0|\xi_{ji}) f(\xi_{ji})}{P(C_{ji}=0,D_{ji}>0)} \\
&= \frac{(\exp\{-\xi_{ji}S_{jiL}\} - \exp\{-\xi_{ji}S_{jiR}\}) f(\xi_{ji})}{\exp\{-G(S_{jiL})\} - \exp\{-G(S_{jiR})\}}
\end{aligned}
$$

So the conditional expectation is with strictly interval censored case is

$$
\begin{aligned}
\hat{\mathbb{E}}(\xi_{ji}) &= \int_{\xi_{ji}} \xi_{ji} f(\xi_{ji}|C_{ji}=0,D_{ji}>0) d\xi_{ji} \\
&= \int_{\xi_{ji}} \xi_{ji} \frac{(\exp\{-\xi_{ji}S_{jiL}\} - \exp\{-\xi_{ji}S_{jiR}\}) f(\xi_{ji})}{\exp\{-G(S_{jiL})\} - \exp\{-G(S_{jiR})\}} d\xi_{ji} \\
&= \frac{G^{'}(S_{jiL})\exp\{-G(S_{jiL})\} - G^{'}(S_{jiR})\exp\{-G(S_{jiR})\}}{\exp\{-G(S_{jiL})\} - \exp\{-G(S_{jiR})\}}
\end{aligned}
$$

Now we discuss $\hat{\mathbb{E}}(W_{jik})$. If $\Delta_{1ji} = 1$, which means the data is observed exactly, we have

$$\hat{\mathbb{E}}(W_{jik}) = E(W_{jik}|A_{ji} = 0, B_{ji} = 1)$$
$$= \begin{cases} 1, & t_{j_k} = T_{ji} \\ 0, & t_{j_k} < T_{ji} \end{cases}$$

If $\Delta_{2ji} = 1$ and $t_{j_k} \leq L_{ji}$, we have

$$\begin{aligned}\hat{\mathbb{E}}(W_{jik}) &= E(W_{jik}|C_{ji} = 0, D_{ji} > 0) \\ &= E(W_{jik}|\sum_{t_{j_k} \leq L_{ji}} W_{jik} = 0, \sum_{L_{ji} < t_{j_k} \leq R_{ji}} W_{jik} > 0) \\ &= 0, \quad t_{j_k} \leq L_{ji}\end{aligned}$$

When and $L_{ji} < t_k \leq R_{ji}$ with $R_{ji} < \infty$, we have

$$\hat{\mathbb{E}}(W_{jik}) = E_{\xi_{ji}}\{E(W_{jik}|\xi_{ji}, C_{ji} = 0, D_{ji} > 0)|C_{ji} = 0, D_{ji} > 0\}$$

We first need to compute $E(W_{jik}|\xi_{ji}, C_{ji} = 0, D_{ji} > 0)$. Notice that

$$\begin{aligned}P(\sum_{L_{ji} < t_{j_k} \leq R_{ji}} W_{jik} > 0|\xi_{ji}) &= 1 - P(\sum_{L_{ji} < t_{j_k} \leq R_{ji}} W_{jik} = 0|\xi_{ji}) \\ &= 1 - \exp\{-\xi_{ji}(S_{jiR} - S_{jiL})\}\end{aligned}$$

and thus, that

$$\begin{aligned}P(W_{jik} = m|\xi_{ji}, \sum_{L_{ji} < t_{j_k} \leq R_{ji}} W_{jik} > 0) &= \frac{P(W_{jik} = m, \sum_{L_{ji} < t_{j_k} \leq R_{ji}} W_{jik} > 0|\xi_{ji})}{P(\sum_{L_{ji} < t_{j_k} \leq R_{ji}} W_{jik} > 0|\xi_{ji})} \\ &= \frac{(\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik} + \gamma^\top Z_{jik}})^m \exp\{-\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik} + \gamma^\top Z_{jik}}\}/m!}{1 - \exp\{-\xi_{ji}(S_{jiR} - S_{jiL})\}}\end{aligned}$$

Then we have

$$
\begin{aligned}
&E_{\xi_{ji}}(W_{jik}|\xi_{ji}, \sum_{L_{ji}<t_{j_k}\leq R_{ji}} W_{jik}>0)\\
=&\sum_{m=1}^{\infty} m\frac{(\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}})^m \exp\{-\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\}/m!}{1-\exp\{-\xi_{ji}(S_{jiR}-S_{jiL})\}}\\
=&\frac{\sum_{m=1}^{\infty} m(\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}})^m \exp\{-\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}\}/m!}{1-\exp\{-\xi_{ji}(S_{jiR}-S_{jiL})\}}\\
=&\frac{\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}}{1-\exp\{-\xi_{ji}(S_{jiR}-S_{jiL})\}}
\end{aligned}
$$

Finally, the expectation of W_{jik} given the observations is

$$
\begin{aligned}
\hat{\mathbb{E}}(W_{jik}) &= E_{\xi_{ji}}\{E(W_{jik}|\xi_{ji}, C_{ji}=0, D_{ji}>0)|C_{ji}=0, D_{ji}>0\}\\
&= E_{\xi_{ji}}\{\frac{\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}}{1-\exp\{-\xi_{ji}(S_{jiR}-S_{jiL})\}}|C_{ji}=0, D_{ji}>0\}\\
&= \int_{\xi_{ji}}\frac{\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}}{1-\exp\{-\xi_{ji}(S_{jiR}-S_{jiL})\}} f(\xi_{ji}|C_{ji}=0, D_{ji}>0)d\xi_{ji}\\
&= \int_{\xi_{ji}}\frac{\xi_{ji}\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}}{1-\exp\{-\xi_{ji}(S_{jiR}-S_{jiL})\}} \frac{(\exp\{-\xi_{ji}S_{jiL}\}-\exp\{-\xi_{ji}S_{jiR}\})f(\xi_{ji})}{\exp\{-G(S_{jiL})\}-\exp\{-G(S_{jiR})\}}d\xi_{ji}\\
&= \frac{\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}}{\exp\{-G(S_{jiL})\}-\exp\{-G(S_{jiR})\}}\int_{\xi_{ji}}\frac{\xi_{ji}(\exp\{-\xi_{ji}S_{jiL}\}-\exp\{-\xi_{ji}S_{jiR}\})f(\xi_{ji})}{1-\exp\{-\xi_{ji}(S_{jiR}-S_{jiL})\}}d\xi_{ji}\\
&= \frac{\lambda_{jk}e^{\beta^\top X_{jik}+\gamma^\top Z_{jik}}}{\exp\{-G(S_{jiL})\}-\exp\{-G(S_{jiR})\}}\int_{\xi_{ji}}\xi_{ji}f(\xi_{ji})\exp\{-\xi_{ji}S_{jiL}\}d\xi_{ji}\\
&= \frac{\lambda_{jk}\exp\{\beta^\top X_{jik}+\gamma^\top Z_{jik}\}}{\exp\{-G(S_{jiL})\}-\exp\{-G(S_{jiR})\}}G^{'}(S_{jiL})\exp\{-G(S_{jiL})\}
\end{aligned}
$$

For the **right censored** case, namely, $\Delta_{1ji}=\Delta_{2ji}=0$; The observed data consists of $O_{ji}=(R_{ji}, X_{ji}, C_{ji}=0)$; Then the conditional expectation of ξ_{ji} given observations is $\hat{\mathbb{E}}(\xi_{ji})$. Notice that the

$$\begin{aligned}
P(\xi_{ji}|O_{ji}) &= \frac{P(O_{ji}|\xi_{ji})f(\xi_{ji})}{P(O_{ji})} \\
&= \frac{P(C_{ji}=0|\xi_{ji})f(\xi_{ji})}{P(C_{ji}=0, D_{ji}=0)} \\
&= \frac{P(\sum_{t_{j_k}\le R_{ji}} W_{jik}=0|\xi_{ji})f(\xi_{ji})}{P(\sum_{t_k\le R_{ji}} W_{jik}=0)} \\
&= \frac{\exp\{-\xi_{ji}S_{jiL}\}f(\xi_{ji})}{\int_{\xi_{ji}}\exp\{-\xi_{ji}S_{jiL}\}f(\xi_{ji})d\xi_{ji}} \\
&= \frac{\exp\{-\xi_{ji}S_{jiL}\}f(\xi_{ji})}{\exp\{-G(S_{jiL})\}}
\end{aligned}$$

Then we have

$$\begin{aligned}
\hat{\mathbb{E}}(\xi_{ji}) &= E(\xi_{ji}|O_{ji}) \\
&= \int_{\xi_{ji}} \xi_{ji} f(\xi_{ji}|O_{ji})d\xi_{ji} \\
&= \int_{\xi_{ji}} \xi_{ji}\frac{\exp\{-\xi_{ji}S_{jiL}\}f(\xi_{ji})}{\exp\{-G(S_{jiL})\}}d\xi_{ji} \\
&= \frac{G^{'}(S_{jiL})\exp\{-G(S_{jiL})\}}{\exp\{-G(S_{jiL})\}} \\
&= G^{'}(S_{jiL})
\end{aligned}$$

The conditional expectation of W_{jik} is

$$\begin{aligned}
\hat{\mathbb{E}}(W_{jik}) &= E_{\xi_{ji}}\{E(W_{jik}|O_{ji},\xi_{ji})|O_{ji}\} \\
&= E_{\xi_{ji}}\{\sum_{m=1}^{\infty} m\times P(W_{jik}=m|O_{ji},\xi_{ji})|O_{ji}\} \\
&= E_{\xi_{ji}}\{\sum_{m=1}^{\infty} m\times 0|O_{ji}\} \\
&= 0
\end{aligned}$$

4.3 Proofs of Theorems

4.3.1 Lemmas

Lemma 1. *For* $\beta,\gamma\in\mathcal{B},\alpha\in\mathcal{B}_2,\Lambda\in\mathcal{M}$, $L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)=O(1)$ *and* $L^{\omega_{\hat{\alpha}_n}}(\beta,\gamma,\Lambda)=O_p(1)$

Proof. Similar to consistency proof, let c_1 and c_2 be positive constants such $c_1\le\exp(\beta^\top X(t)+$

$\gamma^\top Z(t)) \le c_2$ for $\beta, \gamma \in \mathcal{B}$ and $\dot{G}(\cdot)$ is bounded $\Lambda\{T\} = O(1)$ with $\Lambda \in \mathcal{M}$. Then we have

$$
\begin{aligned}
&\exp\Bigg\{\omega_{\alpha_0}\Delta_1 \log\Bigg\{G^{'}\Bigg[\int_0^T \exp\{\beta^\top X(t) + \gamma^\top Z(t)\}d\Lambda(t)\Bigg]\exp\{\beta^\top X(T) + \gamma^\top Z(t)\}\Lambda\{T\}\Bigg\}\\
&\exp\Bigg(-G\Bigg[\int_0^T \exp\{\beta^\top X(t) + \gamma^\top Z(t)\}d\Lambda(t)\Bigg]\Bigg)\Bigg\}\\
&\le \exp\Bigg\{\omega_{\alpha_0}\log\Bigg\{G'[\int_0^T c_2 d\Lambda(t)]c_2 O(1)\exp\Bigg(-G[\int_0^T c_1 d\Lambda(t)]\Bigg)\Bigg\}\Bigg\}\\
&= \exp\Bigg\{\omega_{\alpha_0}\log\Bigg\{G'[c_2\Lambda(T)]c_2 O(1)\exp\Bigg(-G[c_1\Lambda(T)]\Bigg)\Bigg\}\Bigg\}\\
&\le \exp\{\omega_{\alpha_0}O(1)\}\\
&= O(1)\\
&\exp\Bigg\{\omega_{\alpha_0}\Delta_2 \log\Bigg\{\exp\Bigg(-G\Bigg[\int_0^L \exp\{\beta^\top X(t) + \gamma^\top Z(t)\}d\Lambda(t)\Bigg]\Bigg)\\
&-\exp\Bigg(-G\Bigg[\int_0^R \exp\{\beta^\top X(t) + \gamma^\top Z(t)\}d\Lambda(t)\Bigg]\Bigg)\Bigg\}\Bigg\}\\
&\le \exp\Bigg\{\omega_{\alpha_0}\Delta_2 \log\Bigg\{\exp\Bigg(-G\Bigg[\int_0^L \exp\{\beta^\top X(t) + \gamma^\top Z(t)\}d\Lambda(t)\Bigg]\Bigg)\Bigg\}\Bigg\}\\
&= \exp\Bigg\{-\omega_{\alpha_0}\Delta_2 G\Bigg[\int_0^L \exp\{\beta^\top X(t) + \gamma^\top Z(t)\}d\Lambda(t)\Bigg]\Bigg\}\\
&\le \exp\Bigg\{-\omega_{\alpha_0}\Delta_2 G\Bigg[\int_0^L c_1 d\Lambda(t)\Bigg]\Bigg\}\\
&\le \exp\{0\} = 1\\
&\exp\Bigg\{\omega_{\alpha_0}\Delta_3 \log\Bigg\{\exp\Bigg(-G\Bigg[\int_0^L \exp\{\beta^\top X(t) + \gamma Z(t)\}d\Lambda(t)\Bigg]\Bigg)\Bigg\}\Bigg\}\\
&= \exp\Bigg\{-\omega_{\alpha_0}\Delta_3 G\Bigg[\int_0^L \exp\{\beta^\top X(t) + \gamma Z(t)\}d\Lambda(t)\Bigg]\Bigg\}\\
&\le \exp\Bigg\{-\omega_{\alpha_0}\Delta_2 G\Bigg[\int_0^L c_1 d\Lambda(t)\Bigg]\Bigg\}\\
&\le 1
\end{aligned}
$$

Therefore, $L^{\omega_{\alpha_0}}(\beta, \gamma, \Lambda) = O(1)$ and and $L^{\omega_{\hat{\alpha}_n}}(\beta, \gamma, \Lambda) = O_p(1)$ with $\beta, \gamma \in \mathcal{B}, \alpha \in \mathcal{B}_2, \Lambda \in \mathcal{M}$. □

Thus we have

$$
\begin{aligned}
&\left|\log\left(1-a+a\frac{L^{\omega_{\hat{\alpha}_n}}(\beta,\gamma,\Lambda)}{L^{\omega_{\hat{\alpha}_n}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)}\right)-\log\left(1-a+a\frac{L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)}\right)\right| \\
&\leq O_p(1)\left|\frac{L^{\omega_{\hat{\alpha}_n}}(\beta,\gamma,\Lambda)}{L^{\omega_{\hat{\alpha}_n}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)}-\frac{L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)}\right| \\
&\leq O_p(1)\left|L^{\omega_{\hat{\alpha}_n}}(\beta,\gamma,\Lambda)L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)-L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)L^{\omega_{\hat{\alpha}_n}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)\right| \\
&= O_p(1)\Big|L^{\omega_{\hat{\alpha}_n}}(\beta,\gamma,\Lambda)L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)-L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde{\Lambda}_0) \\
&+L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)-L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)L^{\omega_{\hat{\alpha}_n}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)\Big| \\
&\leq O_p(1)\Big(L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)\left|L^{\omega_{\hat{\alpha}_n}}(\beta,\gamma,\Lambda)-L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)\right| \\
&+L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)\left|L^{\omega_{\hat{\alpha}_n}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)-L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)\right|\Big) \\
&\leq O_p(1)O(1) \\
&\leq O_p(1)
\end{aligned}
$$

Lemma 2.

$$
\mathbb{E}[\dot{\phi}^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})]=-\mathbb{E}[\phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})]^2
$$

Proof.

$$
\begin{aligned}
\mathbb{E}[\dot{\phi}^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})]&=\mathbb{E}\left[\frac{d}{d\epsilon}\left(\frac{d}{d\epsilon}\log L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)\right)\right]\Bigg|_{\epsilon=0}+\mathbb{E}\left[\frac{d}{d\epsilon}\left(\frac{d}{d\epsilon}\log L^{\omega_{\alpha_0}}(\beta,\gamma_\epsilon,\Lambda)\right)\right]\Bigg|_{\epsilon=0} \\
&+\mathbb{E}\left[\frac{d}{d\epsilon}\left(\frac{d}{d\epsilon}\log L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda_\epsilon)\right)\right]\Bigg|_{\epsilon=0}+\text{crossterm}
\end{aligned}
$$

For the first term, $\mathbb{E}\left[\frac{d}{d\epsilon}\left(\frac{d}{d\epsilon}\log L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)\right)\right]\Big|_{\epsilon=0}$, we can further compute it as

$$
\begin{aligned}
\mathbb{E}\left[\frac{d}{d\epsilon}\left(\frac{d}{d\epsilon}\log L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)\right)\right]\Bigg|_{\epsilon=0}&=\mathbb{E}\left[\frac{d}{d\epsilon}\left(\frac{\frac{d}{d\epsilon}L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}\right)\Bigg|_{\epsilon=0}\right] \\
&=\mathbb{E}\left[\left(\frac{\frac{d^2}{d\epsilon^2}L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}-\left[\frac{\frac{d}{d\epsilon}L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}\right]^2\right)\Bigg|_{\epsilon=0}\right]
\end{aligned}
$$

For $\mathbb{E}\left[\left(\frac{\frac{d^2}{d\epsilon^2}L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}\right)\Big|_{\epsilon=0}\right]$, we can further expand it as

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\frac{d^2}{d\epsilon^2}L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}\right)\Big|_{\epsilon=0}\right] &= \mathbb{E}\left[\frac{\frac{d}{d\epsilon}\frac{d}{d\epsilon}L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}\Big|_{\epsilon=0}\right] \\
&= \mathbb{E}\left[\frac{d}{d\epsilon}\left(\frac{\frac{d}{d\epsilon}L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)}\right)\Big|_{\epsilon=0}\right] \\
&= \frac{d}{d\epsilon}\mathbb{E}\left[\left(\frac{\frac{d}{d\epsilon}L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)}\right)\Big|_{\epsilon=0}\right] \\
&= \frac{d}{d\epsilon}\mathbb{E}\left[\frac{d}{d\epsilon}\omega_{\alpha_0}(\eta,Z)\log L(\beta_\epsilon,\gamma,\Lambda)\Big|_{\epsilon=0}\right] \\
&= \frac{d}{d\epsilon}\mathbb{E}\left[\mathbb{E}\left[\frac{d}{d\epsilon}\omega_{\alpha_0}(\eta,Z)\log L(\beta_\epsilon,\gamma,\Lambda)|_{\epsilon=0}\Big|Z,\Delta_1,\Delta_1,L,T,R\right]\right] \\
&= \frac{d}{d\epsilon}\mathbb{E}\left[\frac{d}{d\epsilon}\mathbb{E}\left[\omega_{\alpha_0}(\eta,Z)\log L(\beta_\epsilon,\gamma,\Lambda)\Big|Z,\Delta_1,\Delta_1,L,T,R\right]\Big|_{\epsilon=0}\right] \\
&= \frac{d}{d\epsilon}\mathbb{E}\left[\mathbb{E}\left[\omega_{\alpha_0}(\eta,Z)\Big|Z,\Delta_1,\Delta_2,L,T,R\right]\right. \\
&\times\left.\frac{d}{d\epsilon}\mathbb{E}\left[\log L(\beta_\epsilon,\gamma,\Lambda)\Big|Z,\Delta_1,\Delta_1,L,T,R\right]\Big|_{\epsilon=0}\right] \\
&= \frac{d}{d\epsilon}\mathbb{E}\left[\frac{d}{d\epsilon}\mathbb{E}\left[\log L(\beta_\epsilon,\gamma,\Lambda)\Big|Z,\Delta_1,\Delta_1,L,T,R\right]\Big|_{\epsilon=0}\right] \\
&= \frac{d}{d\epsilon}\mathbb{E}\left[\mathbb{E}\left[\frac{d}{d\epsilon}\log L(\beta_\epsilon,\gamma,\Lambda)\Big|Z,\Delta_1,\Delta_1,L,T,R\Big|_{\epsilon=0}\right]\right] \\
&= \frac{d}{d\epsilon}\underbrace{\mathbb{E}\left[\frac{d}{d\epsilon}\log L(\beta_\epsilon,\gamma,\Lambda)\Big|_{\epsilon=0}\right]}_{\text{independent of }\epsilon} \\
&= 0
\end{aligned}$$

Thus, we reach the conclusion

$$\mathbb{E}\left[\frac{d}{d\epsilon}\left(\frac{d}{d\epsilon}\log L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)\right)\right]\Big|_{\epsilon=0} = -\mathbb{E}\left[\left(\left[\frac{\frac{d}{d\epsilon}L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}\right]^2\right)\Big|_{\epsilon=0}\right]$$

Similarly, we have

$$\mathbb{E}\left[\frac{d}{d\epsilon}\left(\frac{d}{d\epsilon}\log L^{\omega_{\alpha_0}}(\beta,\gamma_\epsilon,\Lambda)\right)\right]\Big|_{\epsilon=0} = -\mathbb{E}\left[\left(\left[\frac{\frac{d}{d\epsilon}L^{\omega_{\alpha_0}}(\beta,\gamma_\epsilon,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}\right]^2\right)\Big|_{\epsilon=0}\right]$$

,

$$\mathbb{E}\left[\frac{d}{d\epsilon}\left(\frac{d}{d\epsilon}\log L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda_\epsilon)\right)\right]\Big|_{\epsilon=0} = -\mathbb{E}\left[\left(\left[\frac{\frac{d}{d\epsilon}L^{\omega_{\alpha_0}}(\beta_\epsilon,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda_\epsilon)}\right]^2\right)\Big|_{\epsilon=0}\right]$$

and cross terms can be proved similarly. Thus, we have

$$\mathbb{E}[\dot{\phi}^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})] = -\mathbb{E}[\phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})]^2$$

□

Lemma 3. $\sup_{\theta\in\Theta}\|\sqrt{n}(\Phi_{n,\theta}-\Phi_\theta)(\hat{\alpha}_n)-\sqrt{n}(\Phi_{n,\theta}-\Phi_\theta)(\alpha_0)\|_{\mathcal{H}} = o_p^*(1)$

Proof. We show the above equation holds by using the lemma 3.3.5 of van der Vaart and Wellner (1996). The proof consists of two steps:

1. First we show

$$\sup_{\boldsymbol{h}\in\mathcal{H}}\mathbb{P}(\phi^{\omega_{\hat{\alpha}_n}}(\theta)(\boldsymbol{h})-\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}))^2 \xrightarrow{\text{P}} 0$$

We expand the above equation and get

$$\begin{aligned}
&\mathbb{P}(\phi^{\omega_{\hat{\alpha}_n}}(\theta)(\boldsymbol{h})-\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}))^2\\
&=\mathbb{P}(\phi_1^{\omega_{\hat{\alpha}_n}}(\theta)(h_1)-\phi_1^{\omega_{\alpha_0}}(\theta)(h_1)+\phi_2^{\omega_{\hat{\alpha}_n}}(\theta)(h_2)-\phi_2^{\omega_{\alpha_0}}(\theta)(h_2)+\phi_3^{\omega_{\hat{\alpha}_n}}(\theta)(h_3)-\phi_3^{\omega_{\alpha_0}}(\theta)(h_3))^2\\
&\le 3\underbrace{\mathbb{P}(\phi_1^{\omega_{\hat{\alpha}_n}}(\theta)(h_1)-\phi_1^{\omega_{\alpha_0}}(\theta)(h_1))^2}_{A}+3\underbrace{\mathbb{P}(\phi_2^{\omega_{\hat{\alpha}_n}}(\theta)(h_2)-\phi_2^{\omega_{\alpha_0}}(\theta)(h_2))^2}_{B}+3\underbrace{\mathbb{P}(\phi_3^{\omega_{\hat{\alpha}_n}}(\theta)(h_3)-\phi_3^{\omega_{\alpha_0}}(\theta)(h_3))^2}_{C}
\end{aligned}$$

We prove that part A, B, C converges to 0 in probability.

$$
\begin{aligned}
A &= \mathbb{P}(\phi_1^{\omega_{\hat{\alpha}_n}}(\theta)(h_1) - \phi_1^{\omega_{\alpha_0}}(\theta)(h_1))^2 \\
&= \mathbb{P}\Bigg((\omega_{\hat{\alpha}_n} - \omega_{\alpha_0})\Delta_1 \Bigg\{ \frac{\ddot{G}[I_0(T;\theta)]I_1(T;\theta)}{\dot{G}[I_0(T;\theta)]} - \dot{G}[I_0(T;\theta)]I_1(T;\theta) + X(T) \Bigg\} h_1 \\
&+ (\omega_{\hat{\alpha}_n} - \omega_{\alpha_0})\Delta_2 \frac{\exp\{-G[I_0(R;\theta)]\}\dot{G}[I_0(R;\theta)]I_1(R;\theta) - \exp\{-G[I_0(L;\theta)]\}\dot{G}[I_0(L;\theta)]I_1(L;\theta)}{\exp\{-G[I_0(L;\theta)]\} - \exp\{-G[I_0(R;\theta)]\}} h_1 \\
&+ (\omega_{\alpha_0} - \omega_{\hat{\alpha}_n})(1 - \Delta_1 - \Delta_2)\dot{G}[I_0(L;\theta)]I_1(L;\theta)h_1 \Bigg)^2 \\
&\leq 3\mathbb{P}\left((\omega_{\hat{\alpha}_n} - \omega_{\alpha_0})\Delta_1 \Bigg\{ \frac{\ddot{G}[I_0(T;\theta)]I_1(T;\theta)}{\dot{G}[I_0(T;\theta)]} - \dot{G}[I_0(T;\theta)]I_1(T;\theta) + X(T) \Bigg\} h_1 \right)^2 \\
&+ 3\mathbb{P}\left((\omega_{\hat{\alpha}_n} - \omega_{\alpha_0})\Delta_2 \frac{\exp\{-G[I_0(R;\theta)]\}\dot{G}[I_0(R;\theta)]I_1(R;\theta) - \exp\{-G[I_0(L;\theta)]\}\dot{G}[I_0(L;\theta)]I_1(L;\theta)}{\exp\{-G[I_0(L;\theta)]\} - \exp\{-G[I_0(R;\theta)]\}} h_1 \right)^2 \\
&+ 3\mathbb{P}\left((\omega_{\alpha_0} - \omega_{\hat{\alpha}_n})(1 - \Delta_1 - \Delta_2)\dot{G}[I_0(L;\theta)]I_1(L;\theta)h_1 \right)^2
\end{aligned}
$$

Define $I_0(u;\theta) = \int_0^u \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} d\Lambda(t)$, $I_1(u;\theta) = \int_0^u \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} X(t) d\Lambda(t)$, $I_2(u;\theta) = \int_0^u \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} Z(t) d\Lambda(t)$, $I_3(u;\theta) = \int_0^u \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} h_3(t) d\Lambda(t)$. Notice that $c_1 \leq \exp(\beta^\top X(t) + \gamma^\top Z(t)) \leq c_2$ for $\beta, \gamma \in \mathcal{B}$ and $\dot{G}(\cdot)$ is bounded $\Lambda\{T\} = O(1)$.

$$
c_1\Lambda(u) \leq \int_0^u \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} d\Lambda(t) = I_0(u;\theta) \leq \int_0^u c_2 d\Lambda(t) = c_2\Lambda(u)
$$

$$
I_1(u;\theta) = \int_0^u \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} X(t) d\Lambda(t) \leq c_2 k_x \Lambda(u)
$$

Thus, we have

$$
\begin{aligned}
&\left| (\omega_{\hat{\alpha}_n} - \omega_{\alpha_0})\Delta_1 \Bigg\{ \frac{\ddot{G}[I_0(T;\theta)]I_1(T;\theta)}{\dot{G}[I_0(T;\theta)]} - \dot{G}[I_0(T;\theta)]I_1(T;\theta) + X(T) \Bigg\} h_1 \right| \\
&\leq \frac{2}{\sigma} \left| \frac{\ddot{G}[c_2\Lambda(T)]c_2 k_x \Lambda(T)}{\dot{G}[c_1\Lambda(T)]} + \dot{G}[c_2\Lambda(T)]I_1(T;\theta) + X(T) \right| M = O(1)
\end{aligned}
$$

and

$$\left|(\omega_{\hat{\alpha}_n}-\omega_{\alpha_0})\Delta_2\frac{\exp\{-G[I_0(R;\theta)]\}\dot{G}[I_0(R;\theta)]I_1(R;\theta)-\exp\{-G[I_0(L;\theta)]\}\dot{G}[I_0(L;\theta)]I_1(L;\theta)}{\exp\{-G[I_0(L;\theta)]\}-\exp\{-G[I_0(R;\theta)]\}}\right|M$$
$$\leq\frac{2}{\sigma}\left|\frac{\exp\{-G[I_0(L;\theta)]\}\dot{G}[I_0(R;\theta)]I_1(R;\theta)-\exp\{-G[I_0(R;\theta)]\}\dot{G}[I_0(L;\theta)]I_1(L;\theta)}{\exp\{-G[I_0(L;\theta)]\}-\exp\{-G[I_0(R;\theta)]\}}\right|M$$
$$\leq\frac{2}{\sigma}\left|\frac{\exp\{-G[I_0(L;\theta)]\}\dot{G}[I_0(R;\theta)]I_1(R;\theta)}{\exp\{-G[I_0(L;\theta)]\}-\exp\{-G[I_0(R;\theta)]\}}\right|M$$
$$\leq\frac{2M}{\sigma}\left|\frac{\exp\{-G[c_1\Lambda(L)]\}\dot{G}[c_2\Lambda(R)]c_2\Lambda(R)k_x}{\exp\{-G[I_0(L;\theta)]\}-\exp\{-G[I_0(R;\theta)]\}}\right|=O(1)$$

and

$$\left|(\omega_{\hat{\alpha}_n}-\omega_{\alpha_0})(1-\Delta_1-\Delta_2)\dot{G}[I_0(L;\theta)]I_1(L;\theta)h_1\right|$$
$$\leq\frac{2}{\sigma}\left|\dot{G}[c_2\Lambda(L)]c_2k_x\Lambda(L)\right|M=O(1)$$

Since $\omega_{\hat{\alpha}_n}\xrightarrow{\text{P}}\omega_{\alpha_0}$, by dominated convergence theorem and $\|\boldsymbol{h}\|_{\mathcal{H}}\leq M$, we can prove that

$$\sup_{\boldsymbol{h}\in\mathcal{H}}A\xrightarrow{\text{P}}0$$

and by similar argument we can prove that $B\xrightarrow{\text{P}}0$, $C\xrightarrow{\text{P}}0$ and thus,

$$\sup_{\boldsymbol{h}\in\mathcal{H}}\mathbb{P}(\phi^{\omega_{\hat{\alpha}_n}}(\theta)(\boldsymbol{h})-\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}))^2\xrightarrow{\text{P}}0$$

2. By lemma 3.3.5 of van der Vaart and Wellner (1996) and step 1, we have

$$\sup_{\theta\in\Theta}\|\sqrt{n}(\Phi_{n,\theta}-\Phi_\theta)(\hat{\alpha}_n)-\sqrt{n}(\Phi_{n,\theta}-\Phi_\theta)(\alpha_0)\|_{\mathcal{H}}=o_p^*(1+\sqrt{n}\|\hat{\alpha}_n-\alpha_0\|)$$

Also by the fact that $\sqrt{n}(\hat{\alpha}-\alpha_0)=O_p(1)$ from Saegusa and Wellner (2013), we have

$$\sup_{\theta\in\Theta}\|\sqrt{n}(\Phi_{n,\theta}-\Phi_\theta)(\hat{\alpha}_n)-\sqrt{n}(\Phi_{n,\theta}-\Phi_\theta)(\alpha_0)\|_{\mathcal{H}}=o_p^*(1)$$

□

Lemma 4. *The conditional distribution of $\sqrt{n}(\hat{\alpha}-\tilde{\alpha})$ given the data converges weakly to the asymptotic distribution of $\sqrt{n}(\hat{\alpha}-\alpha_0)$*

Proof. We proof the asymptotic equivalence of $\sqrt{n}(\hat{\alpha}-\alpha_0)$ and $\sqrt{n}(\hat{\alpha}-\tilde{\alpha})$ by theorem 5.21 of van der Vaart (2000) and theorem 2.6 of Kosorok (2007). Let $\Psi_n(\alpha)=\mathbb{P}_n\psi_\alpha$, $\tilde{\Psi}_n(\alpha)=\tilde{\mathbb{P}}_n\psi_\alpha$ and

$\Psi(\alpha) = \mathbb{P}\psi_\alpha$. From definition, $\hat{\alpha}_n$ is a sequence of estimator satisfying $\Psi_n(\hat{\alpha}_n) = 0$, $\tilde{\alpha}_n$ is a sequence of estimator satisfying $\tilde{\Psi}_n(\tilde{\alpha}_n) = 0$ and α_0 is the true value and $\Psi(\alpha_0) = 0$. V_{α_0} is the non-singular derivative of $\Psi(\alpha)$. By theorem 5.21 of van der Vaart (2000), we have

$$\begin{aligned}\sqrt{n}(\hat{\alpha}_n - \alpha_0) &= -V_{\alpha_0}^{-1}\sqrt{n}\mathbb{P}_n\psi_{\alpha_0} + o_p(1) \\ &= -V_{\alpha_0}^{-1}\sqrt{n}(\mathbb{P}_n - \mathbb{P})\psi_{\alpha_0} + o_p(1) \\ &\rightsquigarrow -V_{\alpha_0}^{-1}\mathbb{G}_p\psi_{\alpha_0}\end{aligned} \tag{4.3}$$

Since the class $\{\psi_\alpha : \alpha \in \mathcal{B}\}$ is Donsker by Lemma 5, we have

$$\tilde{\mathbb{G}}_n\psi_{\tilde{\alpha}_n} - \tilde{\mathbb{G}}_n\psi_{\alpha_0} + \mathbb{G}_n\psi_{\tilde{\alpha}_n} - \mathbb{G}_n\psi_{\alpha_0} \xrightarrow{P} 0$$

and then,

$$\sqrt{n}\mathbb{P}(\psi_{\alpha_0} - \psi_{\tilde{\alpha}_n}) = \sqrt{n}\tilde{\mathbb{P}}_n\psi_{\alpha_0} + o_p(1)$$

Similarly, we have

$$\sqrt{n}(\tilde{\alpha}_n - \alpha_0) = -V_{\alpha_0}^{-1}\sqrt{n}\tilde{\mathbb{P}}_n\psi_{\alpha_0} + o_p(1) \tag{4.4}$$

Combining equations (4.3) and (4.4), we have

$$\begin{aligned}\sqrt{n}(\tilde{\alpha}_n - \hat{\alpha}_n) &= -V_{\alpha_0}^{-1}\sqrt{n}(\tilde{\mathbb{P}}_n - \mathbb{P})\psi_{\alpha_0} + o_p(1) \\ &= -V_{\alpha_0}^{-1}\tilde{\mathbb{G}}_n\psi_{\alpha_0} + o_p(1) \\ &\rightsquigarrow -V_{\alpha_0}^{-1}\mathbb{G}_p\psi_{\alpha_0}\end{aligned} \tag{4.5}$$

□

Properties of ψ_α: Notice the $\hat{\alpha} = \max l(\alpha;\eta,z) = \max\sum_{i=1}^n\Big[\eta_i\alpha^\top z_i - \log\Big(1+\exp(\alpha^\top z_i)\Big)\Big]$ which equivalent to the zero point of $\sum_{i=1}^n\Big(\eta_i - \frac{1}{1+\exp(-\alpha^\top z_i)}\Big)z_i$. We have $\psi_\alpha(z) = \Big(\eta - \frac{1}{1+\exp(-\alpha^\top z)}\Big)z$ and Ψ. By definition, we have $\Psi_n(\alpha) = \mathbb{P}_n\psi_\alpha$, $\Psi(\alpha) = \mathbb{P}\Psi_n(\alpha)$ and $\Psi_n(\hat{\alpha}) = \Psi(\alpha_0) = 0$.

Lemma 5. *The class $\{\psi_\alpha(Z(t)) : \alpha \in \mathcal{B}, t \in [0,\tau]\}$ is Donsker.*

Proof. From definition, $\psi_\alpha(Z) = \Big(\eta - \frac{1}{1+\exp(-\alpha^\top Z(t))}\Big)z$, where z are uniformly bounded with uniformly bounded total variation over $[0,\tau]$. The class $\{Z(t) : t \in [0,\tau]\}$ are Donkser by theorem 2.7.5 of van der Vaart and Wellner (1996) and Example 19.11 of van der Vaart (2000). The class of $\{\eta\}$ is also P-Donsker since they are bounded and square-integrable (p.270 of van der Vaart (2000)). Since $\alpha \in \mathcal{B}$ which is a compact set, thus $\{\alpha \in \mathcal{B}\}$ is Donsker and so $\{\alpha^\top Z(t) : t \in [0,\tau]\}$. The class $\{\exp\{(-\alpha^\top Z(t))\}\}$ is P-Donkser since exponential function is Lipschitz continuous on compact set. So the class $\{\psi_\alpha(Z(t)) : \alpha \in \mathcal{B}, t \in [0,\tau]\}$ is P-Donsker by the preservation property in van der Vaart and Wellner (1996) Theorem 2.10.6. $\square$

4.3.2 Proof of Theorem 1

Proof. Let $\mathbb{P}_n$ denote the empirical measure of the data $\mathcal{X}_i = \{\Delta_{1i}, \Delta_{2i}, \Delta_{1i}T_i, L_i, \Delta_{2i}R_i, \bar{X}_i, \bar{Z}_i\}$ for $i = 1, \cdots, n$ and $\mathbb{P}$ denote the true probability measure. Let f be a function form $\mathcal{X}_i$ to $\mathbb{R}$. The corresponding empirical process is $\mathbb{G}_n f = n^{\frac{1}{2}}(\mathbb{P}_n f - \mathbb{P} f)$.

For each single subject, the weighted log-likelihood for a single subject is

$$
\begin{aligned}
l^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda) = \Bigg\{&\omega_{\alpha_0}\Delta_1\Bigg[\log \dot{G}\Bigg(\int_0^{T_i} e^{\beta^\top X(s)+\gamma^\top Z(s)} d\Lambda(s)\Bigg) - G\Bigg(\int_0^{T_i} e^{\beta^\top X(s)+\gamma^\top Z(s)} d\Lambda(s)\Bigg) \\
&+ \log \dot{\Lambda}(T_i) + (\beta^\top X(T_i) + \gamma^\top Z(T_i))\Bigg] \\
&+ \omega_{\alpha_0}\Delta_2 \log\Bigg[\exp\Bigg\{-G\Bigg(\int_0^{L_i} \exp\{\beta^\top X(s) + \gamma^\top Z(s)\} d\Lambda(s)\Bigg)\Bigg\} \\
&- \exp\Bigg\{-G\Bigg(\int_0^{R_i} \exp\{\beta^\top X(s) + \gamma^\top Z(s)\} d\Lambda(s)\Bigg)\Bigg\}\Bigg] \\
&- \omega_{\alpha_0}(1-\Delta_1-\Delta_2) G\Bigg(\int_0^{L_i} e^{\beta^\top X(s)+\gamma^\top Z(s)} d\Lambda(s)\Bigg)
\end{aligned}
\tag{4.6}
$$

where ω_1, ω_2 and ω_3 are weights for exact observation, interval censored observation and right censored observation separately. They are defined as

$$
\omega = \frac{\eta}{q_1}\Delta_1 + \frac{\eta}{q_2}\Delta_2 + \frac{\eta}{q_3}(1-\Delta_1-\Delta_2) \tag{4.7}
$$

where q_1, q_2, q_3 are defined as $q_1 = Pr(\eta = 1|Z, \Delta_1 = 1), q_2 = Pr(\eta = 1|Z, \Delta_2 = 1), q_3 = Pr(\eta = 1|Z, 1-\Delta_1-\Delta_2 = 1)$. We also define $\frac{0}{0} = 0$ and $L^\omega(\beta,\gamma,\Lambda) = \exp\{l^\omega(\beta,\gamma,\Lambda)\}$.

Following the proof of Zhou et al. (2021), we first show that $\limsup_n \hat{\Lambda}(\tau) < \infty$. Let $\mathcal{U}_{i0} = \Delta_{1i}T_i$,

$\mathcal{U}_{1i} = \Delta_{2i} L_i$ and $\mathcal{U}_{2i} = \Delta_{2i} R_i$. We define

$$\tilde{\Lambda}_0(t) = \int_0^t \frac{\Lambda_0^{'}(s)}{f_1(s)} d\left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{2} I(\mathcal{U}_{ij} \leq s) \right\}$$

where $f_1(t)$ is the derivative of the function $F_1(t) = E\left\{ \sum_{j=0}^{2} I(\mathcal{U}_{ij} \leq t) \right\}$ for $t \in [\zeta, \tau]$. Then $\tilde{\Lambda}_0(t)$ is a step function with jumps only at $\{t_1, \cdots, t_m\}$. Since $\frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{2} I(\mathcal{U}_{ij} \leq s) \xrightarrow{\text{a.s.}} F_1(s)$ uniformly in $s \in [\zeta, \tau]$ as $n \to \infty$, we have that $\tilde{\Lambda}_0(t) \xrightarrow{\text{a.s.}} \Lambda_0(t)$ uniformly in $t \in [\zeta, \tau]$ as $n \to \infty$. We want to show that the class $\mathcal{L} = \{L(\beta, \gamma, \Lambda) : \beta, \gamma \in \mathcal{B}, \Lambda \in \mathcal{M}\}$ is a Donsker class, where $\mathcal{M}$ denotes the class of non-decreasing functions Λ with bounded total variations in $[0, \tau]$ and satisfying $\Lambda(0) = 0$.

Following Zeng et al. (2016), $X(t), Z(t)$ belong to Donsker classes indexed by t because bounded total variation and

$$\mathcal{F}_0 = \left\{ \int_0^T \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} d\Lambda(t) : \beta, \gamma \in \mathcal{B}, \Lambda \in \mathcal{M} \right\}$$

is a Donsker class since it is a convex hull of functions $\left\{ I(T \geq s) \exp\{\beta^\top X(s) + \gamma^\top Z(s)\} \right\}$. Similarly,

$$\mathcal{F}_1 = \left\{ \int_0^L \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} d\Lambda(t) : \beta \in \mathcal{B}, \Lambda \in \mathcal{M} \right\}$$

and

$$\mathcal{F}_2 = \left\{ \int_0^R \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} d\Lambda(t) : \beta \in \mathcal{B}, \Lambda \in \mathcal{M} \right\}$$

are Donsker classes. Since G is twice continuously differentiable, $\mathcal{L}$ is a Donsker class due to the preservation of the Donsker property under Lipschitz-continuous transformations. Notice that

$$\exp\left(-G\left[\int_0^L \exp\{\beta_0^\top X(t) + \gamma_0^\top Z(t)\} d\tilde{\Lambda}_0(t) \right] \right) - \exp\left(-G\left[\int_0^R \exp\{\beta_0^\top X(t) + \gamma_0^\top Z(t)\} d\tilde{\Lambda}_0(t) \right] \right) I(R < \infty)$$

is bounded away from 0 and $\mathbb{G}_n l^{\omega_{\hat{\alpha}_n}}(\beta_0, \gamma_0, \tilde{\Lambda}_0) \rightsquigarrow \mathbb{G}_P l^{\omega_{\alpha_0}}(\beta_0, \gamma_0, \tilde{\Lambda}_0)$. Therefore, as $n \to 0$,

$$|\mathbb{P}_n l^{\omega_{\hat{\alpha}_n}}(\beta_0, \gamma_0, \tilde{\Lambda}_0) - \mathbb{P} l^{\omega_{\hat{\alpha}_n}}(\beta_0, \gamma_0, \tilde{\Lambda}_0)| \xrightarrow{\text{a.s.}} 0$$

On the other hand, by the definition of $(\hat{\beta}, \hat{\gamma}, \hat{\Lambda})$, we have $\mathbb{P}_n l^{\omega_{\hat{\alpha}_n}}(\hat{\beta}, \hat{\gamma}, \hat{\Lambda}) \geq \mathbb{P}_n l^{\omega_{\hat{\alpha}_n}}(\beta_0, \gamma_0, \tilde{\Lambda}_0)$. By the construction of $\tilde{\Lambda}_0$, which implies $\tilde{\Lambda}_0(t) \xrightarrow{\text{a.s.}} \Lambda_0(t)$, we have $\mathbb{P} l^{\omega_{\hat{\alpha}_n}}(\beta_0, \gamma_0, \tilde{\Lambda}_0) \xrightarrow{\text{a.s.}} \mathbb{P} l^{\omega_{\alpha_0}}(\beta_0, \gamma_0, \Lambda_0)$.

Since $\mathbb{P}l^{\omega_{\alpha_0}}(\beta_0, \gamma_0, \Lambda_0)$ is finite, we obtain that with probability 1,

$$\liminf_n \mathbb{P}_n l^{\omega_{\hat{\alpha}_n}}(\hat{\beta}, \hat{\gamma}, \hat{\Lambda}) \geq O(1)$$

Let c_1 and c_2 be positive constants such $c_1 \leq \exp(\beta^\top X(t) + \gamma^\top Z(t)) \leq c_2$ for $\beta, \gamma \in \mathcal{B}$. Notice that $G^{'}(\cdot)$ is bounded $\hat{\Lambda}\{T\} = O(1)$. We have

$$\begin{aligned}
&\omega_{\hat{\alpha}_n}\Delta_1 \log\Bigg\{G^{'}\Bigg[\int_0^T \exp\{\hat{\beta}^\top X(t) + \hat{\gamma}^\top Z(t)\}d\hat{\Lambda}(t)\Bigg]\exp\{\hat{\beta}^\top X(T) + \hat{\gamma}^\top Z(t)\}\hat{\Lambda}\{T\} \\
&\exp\Bigg(- G\Bigg[\int_0^T \exp\{\hat{\beta}^\top X(t) + \hat{\gamma}^\top Z(t)\}d\hat{\Lambda}(t)\Bigg]\Bigg)\Bigg\} \\
&\leq \omega_{\hat{\alpha}_n}\Delta_1 \log\Bigg\{G'[\int_0^T c_2 d\hat{\Lambda}(t)]c_2 O(1)\exp\Bigg(- G[\int_0^T c_1 d\hat{\Lambda}(t)]\Bigg)\Bigg\} \\
&= \omega_{\hat{\alpha}_n}\Delta_1 \log\Bigg\{G'[c_2\hat{\Lambda}(T)]c_2 O(1)\exp\Bigg(- G[c_1\hat{\Lambda}(T)]\Bigg)\Bigg\} \\
&\leq \omega_{\hat{\alpha}_n} O(1)
\end{aligned}$$

Notice that $l^{\omega_{\hat{\alpha}_n}}(\hat{\beta}, \hat{\gamma}, \hat{\Lambda}) = \sum_{i=1}^n \omega_{\hat{\alpha}_n}\Delta_{1i}\log\Bigg\{G^{'}\Bigg[\int_0^T \exp\{\hat{\beta}^\top X(t) + \hat{\gamma}Z(t)\}d\hat{\Lambda}(t)\Bigg]\exp\{\hat{\beta}^\top X(T) + \hat{\gamma}Z(t)\}\hat{\Lambda}\{T\}\exp\Bigg(-G\Bigg[\int_0^T \exp\{\hat{\beta}^\top X(t) + \hat{\gamma}Z(t)\}d\hat{\Lambda}(t)\Bigg]\Bigg)\Bigg\} + \sum_{i=1}^n \omega_{\hat{\alpha}_n}\Delta_{2i}\log\Bigg\{\exp\Bigg(-G\Bigg[\int_0^{L_i} \exp\{\hat{\beta}^\top X(t) + \hat{\gamma}Z(t)\}d\Lambda(t)\Bigg]\Bigg) - \exp\Bigg(-G\Bigg[\int_0^{R_i} \exp\{\hat{\beta}^\top X(t) + \hat{\gamma}Z(t)\}d\Lambda(t)\Bigg]\Bigg)\Bigg\} + \sum_{i=1}^n \omega_{\hat{\alpha}_n}(1 - \Delta_{1i} - \Delta_{2i})\log\Bigg\{\exp\Bigg(-G\Bigg[\int_0^{L_i} \exp\{\hat{\beta}^\top X(t) + \hat{\gamma}Z(t)\}d\Lambda(t)\Bigg]\Bigg)$. Then we have

$$\begin{aligned}
\mathbb{P}_n l^{\omega_{\hat{\alpha}_n}}(\hat{\beta}, \hat{\gamma}, \hat{\Lambda}) = &\frac{1}{n}\sum_{i=1}^n \omega_{\hat{\alpha}_n}\Delta_{1i}\log\Bigg\{G^{'}\Bigg[\int_0^T \exp\{\hat{\beta}^\top X(t) + \hat{\gamma}Z(t)\}d\hat{\Lambda}(t)\Bigg]\exp\{\hat{\beta}^\top X(T) + \hat{\gamma}Z(t)\}\hat{\Lambda}\{T\} \\
&\exp\Bigg(- G\Bigg[\int_0^T \exp\{\hat{\beta}^\top X(t) + \hat{\gamma}Z(t)\}d\hat{\Lambda}(t)\Bigg]\Bigg)\Bigg\} \\
&+ \frac{1}{n}\sum_{i=1}^n \omega_{\hat{\alpha}_n}\Delta_{2i}\log\Bigg\{\exp\Bigg(- G\Bigg[\int_0^{L_i} \exp\{\hat{\beta}^\top X(t) + \hat{\gamma}Z(t)\}d\Lambda(t)\Bigg]\Bigg) \\
&- \exp\Bigg(- G\Bigg[\int_0^{R_i} \exp\{\hat{\beta}^\top X(t) + \hat{\gamma}Z(t)\}d\Lambda(t)\Bigg]\Bigg)\Bigg\} \\
&+ \frac{1}{n}\sum_{i=1}^n \omega_{\hat{\alpha}_n}(1 - \Delta_{1i} - \Delta_{2i})\log\Bigg\{\exp\Bigg(- G\Bigg[\int_0^{L_i} \exp\{\hat{\beta}^\top X(t) + \hat{\gamma}Z(t)\}d\Lambda(t)\Bigg]\Bigg)
\end{aligned}$$

Thus, we can conclude

$$\liminf_n \mathbb{P}_n l^{\omega_{\hat{\alpha}_n}}(\hat{\beta},\hat{\gamma},\hat{\Lambda})$$
$$= \liminf_n \frac{1}{n}\sum_{i=1}^n \omega_{\hat{\alpha}_n}\Delta_{1i}\log\Bigg\{G^{'}\Bigg[\int_0^T \exp\{\hat{\beta}^\top X(t)+\hat{\gamma}Z(t)\}d\hat{\Lambda}(t)\Bigg]\exp\{\hat{\beta}^\top X(T)+\hat{\gamma}Z(t)\}\hat{\Lambda}\{T\}$$
$$\exp\Bigg(-G\Bigg[\int_0^T \exp\{\hat{\beta}^\top X(t)+\hat{\gamma}Z(t)\}d\hat{\Lambda}(t)\Bigg]\Bigg)\Bigg\}$$
$$+\liminf_n \frac{1}{n}\sum_{i=1}^n \omega_{\hat{\alpha}_n}\Delta_{2i}\log\Bigg\{\exp\Bigg(-G\Bigg[\int_0^{L_i} \exp\{\hat{\beta}^\top X(t)+\hat{\gamma}Z(t)\}d\Lambda(t)\Bigg]\Bigg)-\exp\Bigg(-G\Bigg[\int_0^{R_i} \exp\{\hat{\beta}^\top X(t)+\hat{\gamma}Z(t)$$
$$+\liminf_n \frac{1}{n}\sum_{i=1}^n \omega_{\hat{\alpha}_n}(1-\Delta_{1i}-\Delta_{2i})\log\Bigg\{\exp\Bigg(-G\Bigg[\int_0^{L_i} \exp\{\hat{\beta}^\top X(t)+\hat{\gamma}Z(t)\}d\Lambda(t)\Bigg]\Bigg)$$
$$\leq \limsup_n \frac{1}{n}\sum_{i=1}^n \omega_{\hat{\alpha}_n}\Delta_{1i}O(1)+\limsup_n \frac{1}{n}\sum_{i=1}^n \omega_{\hat{\alpha}_n}\Delta_{2i}\log\Bigg\{\exp\Bigg(-G\Bigg[\int_0^{L_i} \exp\{\hat{\beta}^\top X(t)+\hat{\gamma}Z(t)\}d\Lambda(t)\Bigg]\Bigg)\Bigg\}$$
$$+\limsup_n \frac{1}{n}\sum_{i=1}^n \omega_{\hat{\alpha}_n}(1-\Delta_{1i}-\Delta_{2i})\log\Bigg\{\exp\Bigg(-G\Bigg[\int_0^{L_i} \exp\{\hat{\beta}^\top X(t)+\hat{\gamma}Z(t)\}d\Lambda(t)\Bigg]\Bigg)\Bigg\}$$
$$\leq \limsup_n \mathbb{P}_n(\omega_{\hat{\alpha}_n})O(1)-\limsup_n \mathbb{P}_n\Bigg(\omega_{\hat{\alpha}_n}\Delta_2 G\Bigg[\int_0^L \exp\{\hat{\beta}^\top X(t)+\hat{\gamma}Z(t)\}d\hat{\Lambda}(t)\Bigg]\Bigg)$$
$$-\limsup_n \mathbb{P}_n\Bigg(\omega_{\hat{\alpha}_n}(1-\Delta_1-\Delta_2)G\Bigg[\int_0^L \exp\{\hat{\beta}^\top X(t)+\hat{\gamma}Z(t)\}d\hat{\Lambda}(t)\Bigg]\Bigg)$$
$$\leq O(1)-\limsup_n \mathbb{P}_n\Bigg(\omega_{\hat{\alpha}_n}(1-\Delta_1-\Delta_2)G\Bigg[\int_0^L \exp\{\hat{\beta}^\top X(t)+\hat{\gamma}Z(t)\}d\hat{\Lambda}(t)\Bigg]\Bigg)$$
$$\leq O(1)-\limsup_n \mathbb{P}_n\Bigg(\omega_{\hat{\alpha}_n}(1-\Delta_1-\Delta_2)G\Bigg[c_1\hat{\Lambda}(L)\Bigg]\Bigg)$$
$$\leq O(1)-\limsup_n \mathbb{P}_n\Bigg(\omega_{\hat{\alpha}_n}(1-\Delta_1-\Delta_2)I(L=\tau)G\Bigg[c_1\hat{\Lambda}(L)\Bigg]\Bigg)$$
$$\leq O(1)-\limsup_n \mathbb{P}_n\Bigg(\omega_{\hat{\alpha}_n}(1-\Delta_1-\Delta_2)I(L=\tau)\Bigg)G\Bigg[c_1\hat{\Lambda}(\tau)\Bigg]$$

From above two inequalities, $\liminf_n \mathbb{P}_n l^{\omega_{\hat{\alpha}_n}}(\hat{\beta},\hat{\gamma},\hat{\Lambda}) \geq O(1)$ and $\liminf_n \mathbb{P}_n l^{\omega_{\hat{\alpha}_n}}(\hat{\beta},\hat{\gamma},\hat{\Lambda}) \leq O(1) - \liminf_n P_n\Bigg(\omega_{\hat{\alpha}_n}(1-\Delta_1-\Delta_2)I(L=\tau)\Bigg)G[c_1\hat{\Lambda}(\tau)]$, we have $\limsup_n \mathbb{P}_n\Bigg(\omega_{\hat{\alpha}_n}(1-\Delta_1-\Delta_2)I(L=\tau)\Bigg)G\Bigg[c_1\hat{\Lambda}(\tau)\Bigg] \leq O(1)$, since $\mathbb{P}_n\Bigg(\omega_{\hat{\alpha}_n}(1-\Delta_1-\Delta_2)I(L=\tau)\Bigg) \xrightarrow{a.s.} \mathbb{P}\Bigg(\omega_{\alpha_0}(1-\Delta_1-\Delta_2)I(L=\tau)\Bigg)$ which is positive under condition 4. Thus $\limsup_n \hat{\Lambda}(\tau) < \infty$ with probability 1 from condition 5.

Now we restrict $\hat{\Lambda}$ to a class of functions with uniformly bounded total variation, equipped with the Skorohod topology on $[\zeta,\tau]$. By Helly's selection theorem, for any subsequence $(\hat{\beta},\hat{\gamma},\hat{\Lambda})$, there is a further subsequence such that $\hat{\beta},\hat{\gamma}$ converge to β^*,γ^* and $\hat{\Lambda}$ converges weakly to some Λ^* on $[\zeta,\tau]$.

Since $\mathbb{P}_n \log L^{\omega_{\hat\alpha_n}}(\hat\beta,\hat\gamma,\hat\Lambda) \ge \mathbb{P}_n \log L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0)$, then $\mathbb{P}_n \log \frac{L^{\omega_{\hat\alpha_n}}(\hat\beta,\hat\gamma,\hat\Lambda)}{L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0)} \ge 0$. From concavity of natural log function, we have for $0 < a < 1$,

$$\mathbb{P}_n \log\left(1-a+a\frac{L^{\omega_{\hat\alpha_n}}(\hat\beta,\hat\gamma,\hat\Lambda)}{L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0)}\right) \ge (1-a)\mathbb{P}_n \log 1 + a\mathbb{P}_n \log\frac{L^{\omega_{\hat\alpha_n}}(\hat\beta,\hat\gamma,\hat\Lambda)}{L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0)} \ge 0$$

where $L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0)$ is bounded away from 0. Notice that for any $\beta,\gamma \in \mathcal{B}$ and $\Lambda \in \mathcal{M}$,

1. For each $n \in \mathbb{N}$,

$$\begin{aligned}
&\left|\log\left(1-a+a\frac{L^{\omega_{\hat\alpha_n}}(\beta,\gamma,\Lambda)}{L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0)}\right) - \log\left(1-a+a\frac{L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde\Lambda_0)}\right)\right| \\
&\le O_p(1)\left|\frac{L^{\omega_{\hat\alpha_n}}(\beta,\gamma,\Lambda)}{L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0)} - \frac{L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde\Lambda_0)}\right| \\
&\le O_p(1)\left|L^{\omega_{\hat\alpha_n}}(\beta,\gamma,\Lambda)L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde\Lambda_0) - L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0)\right| \\
&= O_p(1)\Big|L^{\omega_{\hat\alpha_n}}(\beta,\gamma,\Lambda)L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde\Lambda_0) - L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde\Lambda_0) \\
&+ L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde\Lambda_0) - L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0)\Big| \\
&\le O_p(1)\Big(L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde\Lambda_0)\left|L^{\omega_{\hat\alpha_n}}(\beta,\gamma,\Lambda) - L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)\right| \\
&+ L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)\left|L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0) - L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde\Lambda_0)\right|\Big) \\
&\le O_p(1)O(1) \\
&\le O_p(1)
\end{aligned}$$

Also from the fact $\hat\alpha_n \xrightarrow{a.s} \alpha_0$, we have

$$\left|\log\left(1-a+a\frac{L^{\omega_{\hat\alpha_n}}(\beta,\gamma,\Lambda)}{L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0)}\right) - \log\left(1-a+a\frac{L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde\Lambda_0)}\right)\right| \xrightarrow{a.s} 0$$

2. The class $\{\log(1-\alpha+\alpha\frac{L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde\Lambda_0)}) : \beta,\gamma \in \mathcal{B},\ \Lambda \in \mathcal{M}\}$ is Donkser by Zhou et al. (2021)

By dominated convergence theorem and Lemma 19.24 of van der Vaart (2000), we have $\mathbb{G}_n \log\Big(1 - a + a\frac{L^{\omega_{\hat\alpha_n}}(\beta,\gamma,\Lambda)}{L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0)}\Big) \rightsquigarrow \mathbb{G}_P \log\left(1-a+a\frac{L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda)}{L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\tilde\Lambda_0)}\right)$.

Therefore,

$$(\mathbb{P}_n - \mathbb{P})\log\left(1-a+a\frac{L^{\omega_{\hat\alpha_n}}(\hat\beta,\hat\gamma,\hat\Lambda)}{L^{\omega_{\hat\alpha_n}}(\beta_0,\gamma_0,\tilde\Lambda_0)}\right) \xrightarrow{a.s.} 0$$

Thus,

$$\mathbb{P}_n \log\left(1-a+a\frac{L^{\omega_{\hat{\alpha}_n}}(\hat{\beta},\hat{\gamma},\hat{\Lambda})}{L^{\omega_{\hat{\alpha}_n}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)}\right) = (\mathbb{P}_n-\mathbb{P})\log\left(1-a+a\frac{L^{\omega_{\hat{\alpha}_n}}(\hat{\beta},\hat{\gamma},\hat{\Lambda})}{L^{\omega_{\hat{\alpha}_n}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)}\right) + \mathbb{P}\log\left(1-a+a\frac{L^{\omega_{\hat{\alpha}_n}}(\hat{\beta},\hat{\gamma},\hat{\Lambda})}{L^{\omega_{\hat{\alpha}_n}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)}\right) \geq 0$$

Then we have $\mathbb{P}\log\left(1-a+a\frac{L^{\omega_{\hat{\alpha}_n}}(\hat{\beta},\hat{\gamma},\hat{\Lambda})}{L^{\omega_{\hat{\alpha}_n}}(\beta_0,\gamma_0,\tilde{\Lambda}_0)}\right) \geq 0$. By the convergence of $\hat{\alpha}_n \to \alpha_0$, $\tilde{\Lambda}_0 \to \Lambda_0$, $\hat{\beta} \to \beta^*$ and $\hat{\Lambda} \to \Lambda^*$, then $\mathbb{P}\log\left(1-a+a\frac{L^{\omega_{\alpha_0}}(\beta^*,\gamma^*,\Lambda^*)}{L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\Lambda_0)}\right) \geq 0$. So $\left(1-a+a\frac{L^{\omega_{\alpha_0}}(\beta^*,\gamma^*,\Lambda^*)}{L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\Lambda_0)}\right) \geq 1$ which implies $\mathbb{P}\log\left(\frac{L^{\omega_{\alpha_0}}(\beta^*,\gamma^*,\Lambda^*)}{L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\Lambda_0)}\right) \geq 0$ by letting $a \to 0$. From double expectation theorem, we have

$$\begin{aligned}
\mathbb{P}\log\left(\frac{L^{\omega_{\alpha_0}}(\beta^*,\gamma^*,\Lambda^*)}{L^{\omega_{\alpha_0}}(\beta_0,\gamma_0,\Lambda_0)}\right) &= \mathbb{P}\bigg(\omega_{\alpha_0}(\eta,Z)\log\frac{L(\beta^*,\gamma^*,\Lambda^*)}{L(\beta_0,\gamma_0,\Lambda_0)}\bigg)\\
&= \mathbb{P}\bigg(\frac{\eta}{G_e(\alpha_0^\top Z)}\log\frac{L(\beta^*,\gamma^*,\Lambda^*)}{L(\beta_0,\gamma_0,\Lambda_0)}\bigg)\\
&= \mathbb{P}\bigg(\mathbb{P}\bigg(\frac{\eta}{G_e(\alpha_0^\top Z)}\log\frac{L(\beta^*,\gamma^*,\Lambda^*)}{L(\beta_0,\gamma_0,\Lambda_0)}\bigg|Z,\Delta_1,\Delta_2,L,R,T\bigg)\bigg)\\
&= \mathbb{P}\bigg(\frac{\mathbb{P}(\eta|Z,\Delta_1,\Delta_2,L,R,T)}{G_e(\alpha_0^\top Z)}\mathbb{P}\bigg(\log\frac{L(\beta^*,\gamma^*,\Lambda^*)}{L(\beta_0,\gamma_0,\Lambda_0)}\bigg|Z,\Delta_1,\Delta_2,L,R,T\bigg)\bigg) \quad \text{MAR assumpti}\\
&= \mathbb{P}\bigg(\mathbb{P}\bigg(\log\frac{L(\beta^*,\gamma^*,\Lambda^*)}{L(\beta_0,\gamma_0,\Lambda_0)}\bigg|Z,\Delta_1,\Delta_2,L,R,T\bigg)\bigg)\\
&= \mathbb{P}\bigg(\log\frac{L(\beta^*,\gamma^*,\Lambda^*)}{L(\beta_0,\gamma_0,\Lambda_0)}\bigg)
\end{aligned}$$

Hence, $\mathbb{P}\bigg(\log\frac{L(\beta^*,\gamma^*,\Lambda^*)}{L(\beta_0,\gamma_0,\Lambda_0)}\bigg) \geq 0$. On the other hand, by the property of Kullback-Leibler divergence, $\mathbb{P}\bigg(\log\frac{L(\beta^*,\gamma^*,\Lambda^*)}{L(\beta_0,\gamma_0,\Lambda_0)}\bigg) \leq 0$. Therefore, $\mathbb{P}\bigg(\log\frac{L(\beta^*,\gamma^*,\Lambda^*)}{L(\beta_0,\gamma_0,\Lambda_0)}\bigg) = 0$. We have $L(\beta^*,\gamma^*,\Lambda^*) = L(\beta_0,\gamma_0,\Lambda_0)$ with probability 1. Then, in this case we have

$$\int_0^t \exp\{(\beta^*)^\top X(s) + (\gamma^*)^\top Z(s)\}d\Lambda^*(s) = \int_0^t \exp\{(\beta_0)^\top X(s) + (\gamma_0)^\top X(s)\}d\Lambda_0(s)$$

for $t \in [\zeta,\tau]$. Differentiating both sides with respect to $t \in [\zeta,\tau]$ and take logorithm, we have

$$(\beta^*)^\top X(t) + (\gamma^*)^\top Z(t) + \log\Lambda^{*'}(t) = (\beta_0)^\top X(t) + (\gamma_0)^\top Z(t) + \log\Lambda_0^{'}(t)$$

From condition 3, we have $\beta^* = \beta_0, \gamma^* = \gamma_0$ and $\Lambda^{*'}(t) = \Lambda_0^{'}(t)$ for $t \in [\zeta,\tau]$. We let $X(t) = 0$ by redefining $X(t)$ to center at a deterministic function in the support of $X(t)$ and let $t = \zeta$. Then we have

$$\int_0^{\zeta} d\Lambda^*(s) = \Lambda^*(\zeta) = \Lambda_0(\zeta) = \int_0^{\zeta} d\Lambda_0(s)$$

Hence $\Lambda^*(t) = \Lambda_0(t)$ for $t \in [\zeta, \tau]$. It follow that $\hat{\beta} \xrightarrow{\text{a.s.}} \beta_0$ and $\hat{\Lambda} \xrightarrow{\text{a.s.}} \Lambda_0$. The latter convergence strengthened to uniform convergence by the continuity of Λ_0. □

4.3.3 Proof of Theorem 2

Proof. First let $\mathcal{H} = R^d \times BV[0,\tau]$. For $\boldsymbol{h} = (h_1, h_2, h_3) \in \mathcal{H}$, we introduce the norm $\|\boldsymbol{h}\|_{\mathcal{H}} = \|h_1\|_d + \|h_2\|_d + \|h_3\|_V$, where $\|h_1\|_d, \|h_2\|_d$ are the Euclidean norm in R^d, and $\|h_3\|_V$ is the sum of the absolute value of $h_3(0)$ and the total variation of h_3 on $[0,\tau]$. Let H be the subset of $\mathcal{H}$ with $\|\boldsymbol{h}\|_{\mathcal{H}} \leq M < \infty$. Consider submodels $\beta_\epsilon = \beta + \epsilon h_1, \gamma_\epsilon = \gamma + \epsilon h_2$ and $\Lambda_\epsilon(t) = \int_0^t (1 + \epsilon h_3(u)) d\Lambda(u)$, where $\boldsymbol{h} = (h_1, h_2, h_3) \in H$.

Let $\theta = (\beta, \gamma, \Lambda)$ and $\theta_0 = (\beta_0, \gamma_0, \Lambda_0)$. Define $I_0(u;\theta) = \int_0^u \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} d\Lambda(t)$, $I_1(u;\theta) = \int_0^u \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} X(t) d\Lambda(t)$, $I_2(u;\theta) = \int_0^u \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} Z(t) d\Lambda(t)$, $I_3(u;\theta) = \int_0^u \exp\{\beta^\top X(t) + \gamma^\top Z(t)\} h_3(t) d\Lambda(t)$. The derivatives of the observed data log-likelihood for a single subject along the submodels are

$$\begin{aligned}
\phi_1^{\omega_{\alpha_0}}(\theta)(h_1) &= \frac{d \log L^{\omega_{\alpha_0}}(\beta_\epsilon, \gamma, \Lambda)}{d\epsilon}\bigg|_{\epsilon=0} \\
&= \omega_{\alpha_0}\bigg\{\Delta_1 \left[\frac{\ddot{G}[I_0(T;\theta)] I_1(T;\theta)}{\dot{G}[I_0(T;\theta)]} - \dot{G}[I_0(T;\theta)] I_1(T;\theta) + X(T) h_1\right] \\
&+ \Delta_2 \frac{\exp\{-G[I_0(R;\theta)]\}\dot{G}[I_0(R;\theta)] I_1(R;\theta) - \exp\{-G[I_0(L;\theta)]\}\dot{G}[I_0(L;\theta)] I_1(L;\theta)}{\exp\{-G[I_0(L;\theta)]\} - \exp\{-G[I_0(R;\theta)]\}} h_1 \\
&- \Delta_3 \dot{G}[I_0(L;\theta)] I_1(L;\theta) h_1\bigg\}
\end{aligned}$$

$$
\begin{aligned}
\phi_2^{\omega_{\alpha_0}}(\theta)(h_2) &= \frac{d\log L^{\omega_{\alpha_0}}(\beta,\gamma_\epsilon,\Lambda)}{d\epsilon}\bigg|_{\epsilon=0}\\
&= \omega_{\alpha_0}\bigg\{\Delta_1\left[\frac{\ddot G[I_0(T;\theta)]I_2(T;\theta)}{\dot G[I_0(T;\theta)]} - \dot G[I_0(T;\theta)]I_2(T;\theta) + Z(T)h_2\right]\\
&+ \Delta_2\frac{\exp\{-G[I_0(R;\theta)]\}\dot G[I_0(R;\theta)]I_2(R;\theta) - \exp\{-G[I_0(L;\theta)]\}\dot G[I_0(L;\theta)]I_2(L;\theta)}{\exp\{-G[I_0(L;\theta)]\} - \exp\{-G[I_0(R;\theta)]\}}h_2\\
&- \Delta_3\dot G[I_0(L;\theta)]I_2(L;\theta)h_2\bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\phi_3^{\omega_{\alpha_0}}(\theta)(h_3) &= \frac{d\log L^{\omega_{\alpha_0}}(\beta,\gamma,\Lambda_\epsilon)}{d\epsilon}\bigg|_{\epsilon=0}\\
&= \omega_{\alpha_0}\bigg\{\Delta_1\left[\frac{\ddot G[I_0(T;\theta)]I_3(T;\theta)}{\dot G[I_0(T;\theta)]} - \dot G[I_0(T;\theta)]I_3(T;\theta) + h_3(T)\right]\\
&+ \Delta_2\frac{\exp\{-G[I_0(R;\theta)]\}\dot G[I_0(R;\theta)]I_3(R;\theta) - \exp\{-G[I_0(L;\theta)]\}\dot G[I_0(L;\theta)]I_3(L;\theta)}{\exp\{-G[I_0(L;\theta)]\} - \exp\{-G[I_0(R;\theta)]\}}\\
&- \Delta_3\dot G[I_0(L;\theta)]I_3(L;\theta)\bigg\}
\end{aligned}
$$

The score functions along the submodels are $\Phi_{1,n}^{\omega_{\alpha_0}}(\theta)(h_1) = \mathbb{P}_n\phi_1^{\omega_{\alpha_0}}(\theta)(h_1)$, $\Phi_{2,n}^{\omega_{\alpha_0}}(\theta)(h_2) = \mathbb{P}_n\phi_2^{\omega_{\alpha_0}}(\theta)(h_2)$ and $\Phi_{3,n}^{\omega_{\alpha_0}}(\theta)(h_3) = \mathbb{P}_n\phi_1^{\omega_{\alpha_0}}(\theta)(h_3)$. Let $\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) = \phi_1^{\omega_{\alpha_0}}(\theta)(h_1) + \phi_2^{\omega_{\alpha_0}}(\theta)(h_2) + \phi_3^{\omega_{\alpha_0}}(\theta)(h_3)$, $\Phi_n^{\omega_{\alpha_0}}(\theta)(h) = \mathbb{P}_n\phi^{\omega_{\alpha_0}}(\theta)(h) = \Phi_{1,n}^{\omega_{\alpha_0}}(\theta)(h_1) + \Phi_{2,n}^{\omega_{\alpha_0}}(\theta)(h_2) + \Phi_{3,n}^{\omega_{\alpha_0}}(\theta)(h_3)$. We prove the asymptotic normality of $\hat\theta = (\hat\beta, \hat\gamma, \hat\Lambda)$ by the following steps:

1. $\mathbb{G}_n\phi_0^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}) = \sqrt{n}(\Phi_n^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}) - \Phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}))$ converges in distribution to a tight random element Z.

   Let $\mathcal{F} = \{\phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}) : \|h_1\|_d \le 1, \|h_2\|_d \le 1, h_3 \in BV[0,\tau], \|h_3\|_V \le 1\}$. From condition (A4)-(A5), we have $\exp(-G[I_0(L;\theta_0)]) - \exp(-G[I_0(R;\theta_0)])I(R\infty)$ is bounded away from 0. From condition (A1)-(A5), we have $\sup_{\phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})\in\mathcal{F}} \|\mathbb{P}\phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})\| < \infty$. Notice that $\phi_0^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})$ depends on h_1, h_2 linearly. The class $\{h_3(\cdot) : h_3 \in BV[0,\tau], \|h_3\|_V \le 1\}$ is Donsker from the results of theorem 2.1 of Dudley (1992). Since the function $g(\cdot) = \int_0^{\cdot} \exp\{\beta_0^\top X(t) + \gamma_0^\top Z(t)h_3(t)d\Lambda_0(t)$ is monotone, absolute continuous in $\cdot$ by theorem 11, Chapter 6 of Royden (2010) and hence bounded variation. So the class $\{\int_0^{\cdot} \exp\{\beta_0^\top X(t) + \gamma_0^\top Z(t)h_3(t)d\Lambda_0(t)\}, h_3 \in BV[0,\tau], \|h_3\|_V \le 1\}$ is Donsker by example 19.11 of van der Vaart (2000). Therefore, $\phi_0^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})$ belongs to some Donsker class by theorem 2.10.6 and example

2.10.8 of van der Vaart and Wellner (1996) and thus

$$\mathbb{G}_n\phi_0^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}) = \sqrt{n}(\Phi_n^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}) - \Phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})) \rightsquigarrow Z \in l^{\infty}(H)$$

2. Show that $\Phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h})$ is Frechet differentiable of θ at $\theta = \theta_0$.

   Since $\Phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) = \mathbb{P}\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) = \mathbb{P}\phi_1^{\omega_{\alpha_0}}(\theta)(h_1) + \mathbb{P}\phi_2^{\omega_{\alpha_0}}(\theta)(h_2) + \mathbb{P}\phi_3^{\omega_{\alpha_0}}(\theta)(h_3)$. The Frechet derivative $\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}}(\theta-\theta_0)(\boldsymbol{h})$ can be computed as

$$\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}}(\theta-\theta_0)(\boldsymbol{h}) = \left.\frac{\mathrm{d}\Phi^{\omega_{\alpha_0}}(\theta_0 + \epsilon(\theta-\theta_0))}{\mathrm{d}\epsilon}\right|_{\epsilon=0}$$

   It is clear that $\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}}(\boldsymbol{h})$ is a linear operator. To show $\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}}(\boldsymbol{h})$ is continuous invertible, it is suffice to show that $\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}}(\boldsymbol{h})$ is a one-to-one map and thus invertible. If $\boldsymbol{h} = 0$, then $\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}}(\boldsymbol{h}) = 0$ for any (β, Λ) in the neighborhood of $(\beta_0, \gamma_0, \Lambda_0)$. Choosing (β, γ, Λ) of the form $\beta_\epsilon = \beta_0 + \epsilon h_1$, $\gamma_\epsilon = \gamma_0 + \epsilon h_2$, $\Lambda(t) = \int_0^t (1 + \epsilon h_3(s)) d\Lambda_0(s)$, by the likelihood properties and double expectation,

$$\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}}(\boldsymbol{h}) = \mathbb{P}\dot{\phi}^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}) = -\mathbb{P}\{\phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})\}^2 = 0 \tag{4.8}$$

   If $\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}}(\boldsymbol{h}) = 0$, then $\phi^{\omega_{\alpha_0}}(\boldsymbol{h}) = 0$ with probability 1. Let $\Delta_2 = 1$, we have

$$\exp\{-G[I_0(R;\theta_0)]\}\dot{G}[I_0(R;\theta_0)]I_1(R;\theta_0)h_1 - \exp\{-G[I_0(L;\theta_0)]\}\dot{G}[I_0(L;\theta_0)]I_1(L;\theta_0)h_1 = 0$$

$$\exp\{-G[I_0(R;\theta_0)]\}\dot{G}[I_0(R;\theta_0)]I_2(R;\theta_0)h_2 - \exp\{-G[I_0(L;\theta_0)]\}\dot{G}[I_0(L;\theta_0)]I_2(L;\theta_0)h_2 = 0$$

$$\exp\{-G[I_0(R;\theta_0)]\}\dot{G}[I_0(R;\theta_0)]I_3(R;\theta_0) - \exp\{-G[I_0(L;\theta_0)]\}\dot{G}[I_0(L;\theta_0)]I_3(L;\theta_0) = 0$$

   and we obtain $h_1 = 0, h_2 = 0$ and $h_3(\cdot) = 0$ for $t \in [\zeta, \tau]$.

3. Under conditions $1-5$, we can show that $\{\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) - \phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}) : \|\theta-\theta_0\| < \delta, \boldsymbol{h} \in H\}$ is P-Donsker for some $\delta > 0$, and

$$\sup_{h\in H} P(\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) - \phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}))^2 \to 0, \quad \text{as} \quad \theta \to \theta_0$$

   Since $\hat{\theta} \xrightarrow{\text{a.s.}} \theta_0$, we have

$$\|\mathbb{G}_n(\phi^{\omega_{\alpha_0}}(\hat{\theta})(\boldsymbol{h}) - \phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}))\|_H = o_{p^*}(1 + \sqrt{n}\|\hat{\theta} - \theta_0\|)$$

where o_{p^*} stands for converging in outer probability.

4. The above equation can be written as

$$\sqrt{n}(\Phi_n^{\omega_{\alpha_0}} - \Phi^{\omega_{\alpha_0}})(\hat{\theta}) - \sqrt{n}(\Phi_n^{\omega_{\alpha_0}} - \Phi^{\omega_{\alpha_0}})(\theta_0) = o_{p^*}(1 + \sqrt{n}\|\hat{\theta} - \theta_0\|)$$

Since $\Phi_n^{\omega_{\alpha_0}}(\hat{\theta}) = 0$ and $\Phi^{\omega_{\alpha_0}}(\theta_0) = 0$, from theorem 3.3.1 of van der Vaart and Wellner (1996), we have

$$\sqrt{n}\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}}(\hat{\theta}_n - \theta_0) = -\sqrt{n}(\Phi_n^{\omega_{\alpha_0}} - \Phi^{\omega_{\alpha_0}})(\theta_0) + o_{p^*}(1)$$

and by continuous invertible of $\dot{\Phi}^{\omega_{\alpha_0}}$, we have $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} -(\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}})^{-1}Z$.

5. Next we prove the estimator $\hat{\theta}^{\omega_{\hat{\alpha}_n}} = (\hat{\beta}^{\omega_{\hat{\alpha}_n}}, \hat{\gamma}^{\omega_{\hat{\alpha}_n}}, \hat{\Lambda}^{\omega_{\hat{\alpha}_n}})$ converges weakly to some tight random elements in $l^\infty(H)$ under the estimated weight $\omega_{\hat{\alpha}_n}$. Notice that $\hat{\alpha}_n \xrightarrow{p} \alpha_0$ and we have $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = \mathbb{G}_n\psi + o_p^*(1)$ by proposition A1 of Saegusa and Wellner (2013). By theorem 1 of Breslow and Wellner (2008), we have

$$\begin{aligned}
\sqrt{n}(\hat{\theta}^{\omega_{\hat{\alpha}_n}} - \theta_0) &= -(\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}})^{-1}[\sqrt{n}(\Phi_n^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}) - \Phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})) + \sqrt{n}(\Phi^{\omega_{\hat{\alpha}_n}}(\theta_0)(\boldsymbol{h}) - \Phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}))] + o_p^*(1) \\
&= -(\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}})^{-1}[\sqrt{n}(\Phi_n^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h}) - \Phi^{\omega_{\alpha_0}}(\theta_0)(\boldsymbol{h})) + \dot{\Phi}_\alpha\mathbb{G}_n\psi] + o_p^*(1) \\
&\rightsquigarrow -(\dot{\Phi}_{\theta_0}^{\omega_{\alpha_0}})^{-1}(\mathcal{Z} + \dot{\Phi}_\alpha\mathbb{G}_p\psi_{\alpha_0})
\end{aligned}$$

Thus the weak convergence of $\hat{\theta}^{\omega_{\hat{\alpha}_n}} = (\hat{\beta}^{\omega_{\hat{\alpha}_n}}, \hat{\gamma}^{\omega_{\hat{\alpha}_n}}, \hat{\Lambda}^{\omega_{\hat{\alpha}_n}})$ established. □

4.3.4 Proof of Theorem 3

Proof: Let $u_1, u_2, \cdots, u_n$ be a sequence of IID exponentially distributed random variables with $\mu = \mathbb{P}(u_1) = 1$ and $\sigma^2 = \text{Var}(u_1) = 1$. We assume that $u_1, u_2, \cdots, u_n$ are independent of the observed data $O_i = (\Delta_{1i}, \Delta_{2i}, \eta_i, \Delta_{1i}T_i, L_i, \Delta_{2i}R_i, \eta_i X_i, Z_i)$. Let $\tilde{u}_i = \frac{u_i}{\bar{u}}$ where $\bar{u} = \frac{1}{n}\sum_{i=1}^n u_i$. Let $\tilde{\mathbb{P}}_n f = \frac{1}{n}\sum_{i=1}^n \tilde{u}_i f(O_i)$ be the weighted bootstrapped empirical process for any measurable function f. Let $\tilde{\Phi}_n^{\omega_{\alpha_0}}$ be $\Phi_n^{\omega_{\alpha_0}}$ with $\mathbb{P}_n$ replaced by $\tilde{\mathbb{P}}_n$ and $\tilde{\theta} = (\tilde{\vartheta}, \tilde{\Lambda})$ be the weighted bootstrap estimator which solves $\tilde{\Phi}_n^{\omega_{\alpha_0}}(\theta) = 0$. Let $\tilde{\Phi}^{\omega_{\alpha_0}}(\theta) = \mathbb{P}(\tilde{u} \cdot \Phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}))$. From above, the class of functions $\{\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) : \theta \in B_\delta(\theta_0), h \in BV[0,\tau]\}$ is P-Donsker for some $\delta > 0$. Thus the class $\{\tilde{u} \cdot \phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) : \theta \in B_\delta(\theta_0), h \in BV[0,\tau]\}$ from Kosorok (2007) theorem 10.1.

Since $e_1, e_2, \cdots, e_n$ is a sequence of IID exp(1) variable. Let $\xi_i = u_i - 1$, then we have $E(\xi_i) = 0$ and $\text{Var}(\xi_i) = 1$. Since $\{\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) : \theta \in B_\delta(\theta_0), h \in BV[0,\tau]\}$ is P-Donsker, we have $\mathbb{G}_n\phi^{\omega_{\alpha_0}}$ converges to a tight random element, $\mathbb{G}_P\phi^{\omega_{\alpha_0}}$. Define weighted bootstrap empirical process as

$$\tilde{\mathbb{G}}_n\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) = \sqrt{n}\left\{\tilde{\mathbb{P}}\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) - \mathbb{P}_n\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h})\right\}$$

On the other hand, by theorem 10.1 of Kosorok (2007) $(i \to ii)$, we have

$$\tilde{\mathbb{G}}_n\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) - \mathbb{G}_n\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) \rightsquigarrow Z_0 \in l^\infty(H)$$

therefore, $\tilde{\mathbb{G}}_n\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h})$ converges to some tight random elements in $l^\infty(H)$ and hence the class $\{\tilde{e}\cdot\Phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}) : \theta \in B_\delta(\theta_0), h \in BV[0,\tau]\}$ is P-Donsker. We also have $\mathbb{P}\Big(\tilde{e}\cdot\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h})\Big) = \mathbb{P}\Big(\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h})\Big)$ which implies $\tilde{\Phi}^{\omega_{\alpha_0}}(\theta) = \Phi^{\omega_{\alpha_0}}(\theta)$.

By Taylor expansion, we have the following

$$\begin{aligned}
0 &= \tilde{\mathbb{P}}_n\phi^{\omega_{\alpha_0}}(\tilde{\theta})(\boldsymbol{h}) - \tilde{\mathbb{P}}_n\phi^{\omega_{\alpha_0}}(\hat{\theta})(\boldsymbol{h}) + \tilde{\mathbb{P}}_n\phi^{\omega_{\alpha_0}}(\hat{\theta})(\boldsymbol{h}) - \mathbb{P}_n\phi^{\omega_{\alpha_0}}(\hat{\theta})(\boldsymbol{h}) \\
&= \left(\frac{\partial\tilde{\mathbb{P}}_n\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h})}{\partial\theta}\bigg|_{\theta=\hat{\theta}}\right)(\tilde{\theta}-\theta) + (\tilde{\mathbb{P}}_n - \mathbb{P}_n)\phi^{\omega_{\alpha_0}}(\hat{\theta})(\boldsymbol{h}) + o_p(\|\tilde{\theta}-\theta_0\| + \|\hat{\theta}-\theta_0\|)
\end{aligned} \tag{4.9}$$

The consistency of $\tilde{\theta}$ can be proved by the similar arguments of consistency proof. From theorem 2.6 of Kosorok (2007), the empirical process $\tilde{\mathbb{G}}_n$ given the data is asymptotically equivalent to $\mathbb{G}_n$ since $\mu = \sigma^2 = 1$. Hence the equation (4.9) above can be written as

$$\begin{aligned}
\sqrt{n}\cdot 0 =& \sqrt{n}\left(\frac{\partial\tilde{\mathbb{P}}_n\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h})}{\partial\theta}\bigg|_{\theta=\hat{\theta}}\right)(\tilde{\theta}-\theta) + \sqrt{n}(\tilde{\mathbb{P}}_n - \mathbb{P}_n)\phi^{\omega_{\alpha_0}}(\hat{\theta})(\boldsymbol{h}) + \sqrt{n}\cdot o_p(\|\tilde{\theta}-\theta_0\| + \|\hat{\theta}-\theta_0\|) \\
=& \sqrt{n}\left(\frac{\partial\tilde{\mathbb{P}}_n\phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h})}{\partial\theta}\bigg|_{\theta=\hat{\theta}}\right)(\tilde{\theta}-\theta) + \sqrt{n}(\tilde{\mathbb{P}}_n - \mathbb{P})\phi^{\omega_{\alpha_0}}(\hat{\theta})(\boldsymbol{h}) + o_p(1) \\
=& \sqrt{n}\dot{\Phi}^{\omega_{\alpha_0}}_{\theta_0}(\tilde{\theta}-\hat{\theta}) + \mathbb{G}_n\phi^{\omega_{\alpha_0}}(\hat{\theta})(\boldsymbol{h}) + o_p(1) \\
\rightsquigarrow& \sqrt{n}\dot{\Phi}^{\omega_{\alpha_0}}_{\theta_0}(\tilde{\theta}-\hat{\theta}) + Z
\end{aligned}$$

Therefore, we have

$$\sqrt{n}(\tilde{\theta}-\hat{\theta}) \rightsquigarrow -(\dot{\Phi}^{\omega_{\alpha_0}}_{\theta_0})^{-1}\cdot Z \tag{4.10}$$

Now we prove the similar results still hold under estimated weight $\omega_{\hat{\alpha}_n}$. Similarly, let $u_1, u_2, \cdots, u_n$ be a sequence of IID exponentially distributed random variables with $\mu = \mathbb{P}(u_1) = 1$ and $\sigma^2 = \text{Var}(u_1) = 1$. Since the class $\{\phi^{\omega_{\alpha_0}}(\boldsymbol{h}) : \|h_1\|_d \leq 1, \|h_2\|_d \leq 1, h_3 \in BV_1[0,\tau]\}$ is Donkser and we proved

$$\sup_{\boldsymbol{h}\in\mathcal{H}} \mathbb{P}(\phi^{\omega_{\hat{\alpha}_n}}(\theta)(\boldsymbol{h}) - \phi^{\omega_{\alpha_0}}(\theta)(\boldsymbol{h}))^2 \xrightarrow{\text{P}} 0$$

in lemma 3, thus by lemma 19.24 of van der Vaart (2000), we have

$$\mathbb{G}_n\phi^{\omega_{\hat{\alpha}_n}} \rightsquigarrow \mathbb{G}_p\phi^{\omega_{\alpha_0}} = Z \in l^{\infty}(H)$$

By definition we have $\tilde{\Phi}^{\omega_{\tilde{\alpha}_n}}(\tilde{\theta}) = 0$, $\Phi_n^{\omega_{\hat{\alpha}_n}}(\hat{\theta}) = 0$, similarly we can have the following

$$\begin{aligned}
0 &= \tilde{\mathbb{P}}_n\phi^{\omega_{\tilde{\alpha}_n}}(\tilde{\theta})(\boldsymbol{h}) - \tilde{\mathbb{P}}_n\phi^{\omega_{\tilde{\alpha}_n}}(\hat{\theta})(\boldsymbol{h}) + \hat{\mathbb{P}}_n\phi^{\omega_{\tilde{\alpha}_n}}(\hat{\theta})(\boldsymbol{h}) - \mathbb{P}_n\phi^{\omega_{\hat{\alpha}_n}}(\hat{\theta})(\boldsymbol{h}) \\
&= \tilde{\mathbb{P}}_n\phi^{\omega_{\tilde{\alpha}_n}}(\tilde{\theta})(\boldsymbol{h}) - \tilde{\mathbb{P}}_n\phi^{\omega_{\tilde{\alpha}_n}}(\hat{\theta})(\boldsymbol{h}) + \tilde{\mathbb{P}}_n\phi^{\omega_{\tilde{\alpha}_n}}(\hat{\theta})(\boldsymbol{h}) - \tilde{\mathbb{P}}_n\phi^{\omega_{\hat{\alpha}_n}}(\hat{\theta})(\boldsymbol{h}) + \tilde{\mathbb{P}}_n\phi^{\omega_{\hat{\alpha}_n}}(\hat{\theta})(\boldsymbol{h}) - \mathbb{P}_n\phi^{\omega_{\hat{\alpha}_n}}(\hat{\theta})(\boldsymbol{h}) \\
&= \left(\frac{\partial\tilde{\mathbb{P}}_n\phi^{\omega_{\tilde{\alpha}_n}}(\theta)(\boldsymbol{h})}{\partial\theta}\bigg|_{\theta=\hat{\theta}}\right)(\tilde{\theta} - \hat{\theta}) + \left(\frac{\partial\tilde{\mathbb{P}}_n\phi^{\omega_{\alpha}}(\hat{\theta})(\boldsymbol{h})}{\partial\alpha}\bigg|_{\alpha=\hat{\alpha}}\right)(\tilde{\alpha} - \hat{\alpha}) + (\tilde{\mathbb{P}}_n - \mathbb{P}_n)\phi^{\omega_{\hat{\alpha}_n}}(\hat{\theta})(\boldsymbol{h}) \\
&+ o_p(\|\tilde{\theta} - \theta_0\| + \|\hat{\theta} - \theta_0\| + \|\tilde{\alpha} - \alpha_0\| + \|\hat{\alpha} - \alpha_0\|)
\end{aligned} \tag{4.11}$$

Notice that the conditional distribution of $(\tilde{\mathbb{P}}_n - \mathbb{P}_n)\phi^{\omega_{\hat{\alpha}_n}}(\hat{\theta})(\boldsymbol{h})$ given the data is asymptotic equivalent to the distribution of $(\mathbb{P}_n - \mathbb{P})\phi^{\omega_{\hat{\alpha}_n}}(\hat{\theta})(\boldsymbol{h})$ and $\sqrt{n}(\hat{\alpha} - \alpha_0) \rightsquigarrow \mathbb{G}_p\psi$ by Lemma in Appendix, then the equation 4.11 can be re-written as

$$0 = \sqrt{n}\dot{\Phi}_{\theta}^{\omega_{\tilde{\alpha}_n}}(\tilde{\theta} - \hat{\theta}) + \dot{\Phi}_{\alpha}\sqrt{n}(\tilde{\alpha} - \hat{\alpha}) + \tilde{\mathbb{G}}_n\phi^{\omega_{\hat{\alpha}_n}}(\hat{\theta})(\boldsymbol{h}) + o_p(1)$$

By Lemma 4 in the appendix, we have $\sqrt{n}(\tilde{\alpha}_n - \alpha_0) = -V_{\alpha_0}^{-1}\sqrt{n}\tilde{\mathbb{P}}_n\psi_{\alpha_0} + o_p(1) = \mathbb{G}_p\psi_{\alpha_0} + o_p(1)$ and thus

$$\sqrt{n}(\tilde{\theta} - \hat{\theta}) \rightsquigarrow -(\dot{\Phi}_{\theta}^{\omega_{\alpha_0}})^{-1}(\dot{\Phi}_{\alpha}\mathbb{G}_p\psi_{\alpha_0} + \mathcal{Z}) \tag{4.12}$$

4.4 Tables and Figures for Chapter 2

Table 4.1: Simulation results for estimation of the regression parameters under the model configuration (1) (Missing completely at random with two baselines)

		$\beta = 0.5$				$\gamma = -0.5$			
n	pt	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
$r = 0$		10% missing for cases, 70% missing for non-cases							
800	0	0.022	0.123	0.115	0.922	-0.027	0.131	0.115	0.910
	0.2	0.011	0.115	0.108	0.932	-0.013	0.112	0.108	0.934
	0.5	0.008	0.110	0.104	0.944	-0.009	0.107	0.105	0.944
	1	0.008	0.099	0.099	0.958	-0.005	0.104	0.100	0.932
1200	0	0.016	0.100	0.093	0.924	-0.017	0.103	0.094	0.920
	0.2	0.004	0.096	0.090	0.940	-0.005	0.096	0.090	0.932
	0.5	0.001	0.092	0.086	0.932	-0.002	0.091	0.090	0.942
	1	0.003	0.077	0.076	0.948	0.005	0.078	0.077	0.947
$r = 0.5$		10% missing for cases, 70% missing for non-cases							
800	0	0.023	0.156	0.150	0.923	-0.013	0.161	0.150	0.928
	0.2	-0.004	0.138	0.141	0.952	-0.004	0.148	0.143	0.946
	0.5	0.007	0.133	0.137	0.965	-0.000	0.141	0.137	0.950
	1	0.005	0.135	0.129	0.930	0.002	0.138	0.129	0.930
1200	0	0.008	0.127	0.122	0.930	-0.017	0.125	0.123	0.942
	0.2	0.010	0.123	0.115	0.937	-0.017	0.122	0.116	0.932
	0.5	0.009	0.119	0.111	0.938	-0.014	0.117	0.112	0.930
	1	-0.002	0.113	0.107	0.934	-0.004	0.110	0.107	0.952
$r = 1$		10% missing for cases, 70% missing for non-cases							
800	0	0.025	0.183	0.174	0.930	-0.011	0.184	0.177	0.925
	0.2	0.011	0.170	0.164	0.950	-0.007	0.173	0.165	0.936
	0.5	0.005	0.163	0.158	0.948	0.007	0.166	0.158	0.944
	1	0.003	0.149	0.149	0.964	-0.003	0.156	0.150	0.930
1200	0	0.009	0.149	0.142	0.935	-0.023	0.147	0.144	0.944
	0.2	0.013	0.136	0.134	0.941	0.008	0.130	0.134	0.955
	0.5	0.010	0.132	0.129	0.930	0.008	0.125	0.129	0.952
	1	-0.002	0.128	0.122	0.938	-0.005	0.128	0.123	0.938

Table 4.2: Simulation results for estimation of the regression parameters under the model configuration (2) (Missing at random with two baselines)

n	pt	$\beta = 0.5$ Bias	SSD	ESE	CP	$\gamma = -0.5$ Bias	SSD	ESE	CP
$r = 0$				10% missing for cases, 60% missing for noncases					
800	0	0.015	0.089	0.084	0.925	-0.007	0.112	0.088	0.934
	0.2	0.011	0.102	0.098	0.923	0.008	0.104	0.104	0.935
	0.5	0.004	0.097	0.094	0.930	0.002	0.109	0.102	0.923
	1	0.008	0.097	0.091	0.935	-0.006	0.101	0.091	0.943
1200	0	0.015	0.089	0.084	0.925	-0.007	0.112	0.088	0.934
	0.2	0.004	0.083	0.084	0.938	0.003	0.085	0.084	0.942
	0.5	0.003	0.080	0.078	0.946	0.002	0.084	0.083	0.946
	1	0.003	0.077	0.074	0.940	-0.005	0.084	0.081	0.940
$r = 0.5$				10% missing for cases, 60% missing for noncases					
800	0	0.011	0.128	0.117	0.912	-0.006	0.125	0.123	0.950
	0.2	-0.000	0.128	0.126	0.958	0.007	0.130	0.130	0.955
	0.5	-0.000	0.123	0.123	0.953	0.006	0.126	0.123	0.952
	1	0.000	0.113	0.117	0.956	0.001	0.123	0.123	0.950
1200	0	0.017	0.110	0.107	0.938	-0.008	0.113	0.109	0.938
	0.2	0.011	0.106	0.103	0.937	0.002	0.105	0.105	0.941
	0.5	0.011	0.103	0.100	0.944	0.004	0.101	0.103	0.946
	1	0.005	0.095	0.095	0.954	-0.006	0.100	0.099	0.942
$r = 1$				10% missing for cases, 60% missing for noncases					
800	0	0.015	0.173	0.166	0.945	0.007	0.177	0.185	0.933
	0.2	0.011	0.169	0.164	0.950	0.007	0.173	0.165	0.936
	0.5	0.011	0.132	0.129	0.930	0.009	0.129	0.125	0.952
	1	-0.000	0.130	0.134	0.960	-0.002	0.141	0.142	0.950
1200	0	0.022	0.126	0.126	0.957	-0.008	0.133	0.127	0.930
	0.2	0.006	0.123	0.119	0.947	0.008	0.127	0.123	0.961
	0.5	0.006	0.119	0.116	0.948	0.006	0.124	0.120	0.952
	1	0.004	0.108	0.109	0.952	-0.006	0.119	0.116	0.952

Table 4.3: Simulation results for estimation of the regression parameters under the model configuration (3) (Missing at random with time-dependent covariate with two baselines)

		$\beta = 0.5$				$\gamma = -0.5$			
n	pt	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
$r = 0$		10% missing for cases, 65% missing for noncases							
800	0	-0.013	0.111	0.105	0.927	0.025	0.208	0.190	0.933
	0.2	-0.018	0.103	0.100	0.930	0.036	0.201	0.185	0.920
	0.5	-0.017	0.101	0.097	0.940	0.033	0.190	0.183	0.923
	1	-0.006	0.100	0.096	0.940	0.023	0.189	0.180	0.940
1200	0	0.011	0.092	0.085	0.928	0.027	0.160	0.155	0.930
	0.2	-0.007	0.084	0.082	0.940	0.019	0.165	0.151	0.943
	0.5	-0.002	0.084	0.081	0.940	0.011	0.162	0.152	0.950
	1	-0.008	0.083	0.080	0.933	0.010	0.153	0.151	0.940
$r = 0.5$		10% missing for cases, 65% missing for noncases							
800	0	-0.012	0.137	0.130	0.950	0.031	0.235	0.235	0.944
	0.2	-0.002	0.136	0.130	0.941	0.005	0.236	0.233	0.948
	0.5	-0.002	0.133	0.128	0.954	0.008	0.238	0.231	0.945
	1	-0.007	0.132	0.127	0.935	0.015	0.231	0.228	0.940
1200	0	-0.000	0.118	0.115	0.943	0.020	0.189	0.193	0.956
	0.2	0.002	0.114	0.110	0.924	0.004	0.196	0.190	0.940
	0.5	-0.001	0.114	0.105	0.930	0.004	0.193	0.188	0.940
	1	0.005	0.106	0.104	0.944	0.024	0.185	0.186	0.952
$r = 1$		10% missing for cases, 65% missing for noncases							
800	0	0.005	0.161	0.150	0.937	-0.003	0.278	0.272	0.951
	0.2	0.001	0.158	0.148	0.934	0.023	0.275	0.267	0.937
	0.5	0.008	0.155	0.147	0.943	0.014	0.267	0.265	0.942
	1	0.009	0.153	0.145	0.956	-0.009	0.265	0.262	0.950
1200	0	-0.008	0.087	0.080	0.930	0.021	0.153	0.150	0.940
	0.2	0.003	0.126	0.122	0.946	0.007	0.223	0.217	0.946
	0.5	0.006	0.125	0.121	0.950	0.005	0.225	0.215	0.940
	1	0.007	0.125	0.120	0.948	0.008	0.201	0.214	0.958

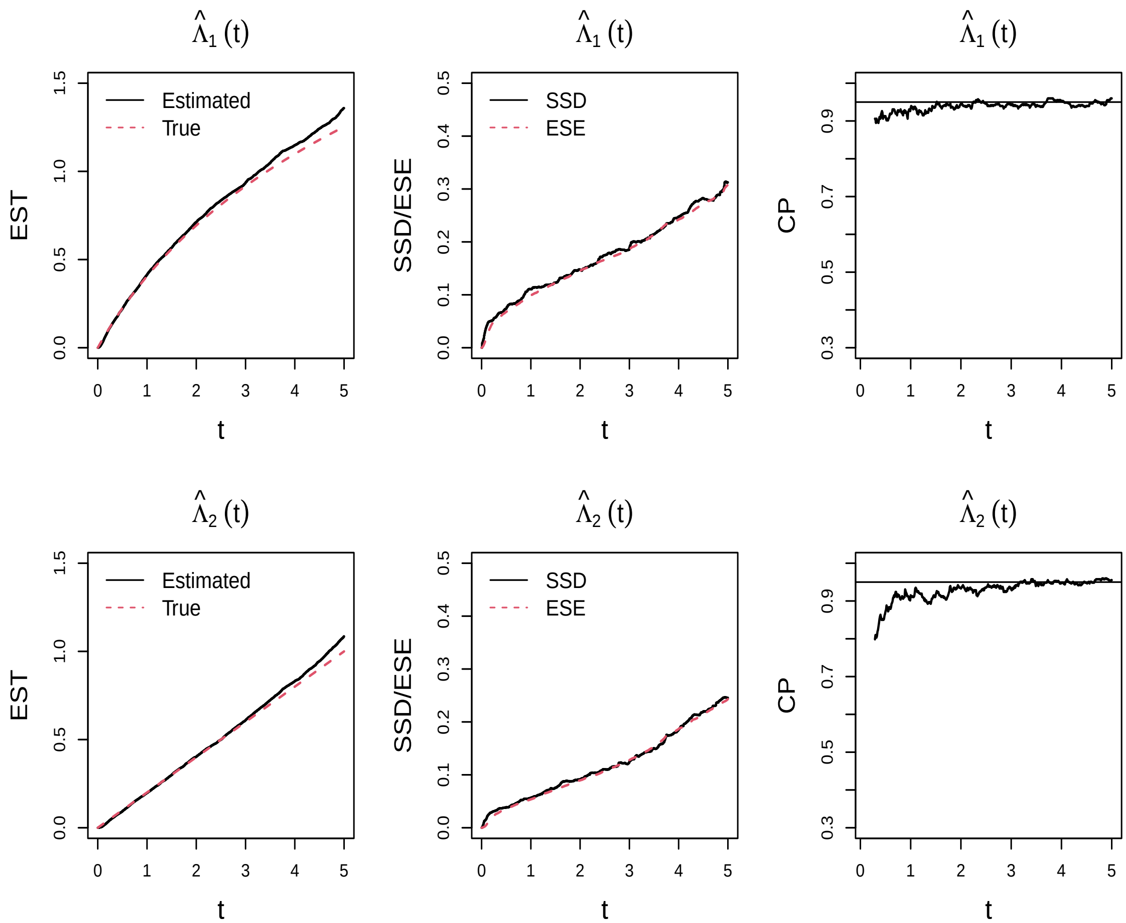

Figure 4.1: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 1 with $r = 0$ and $pt = 0$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

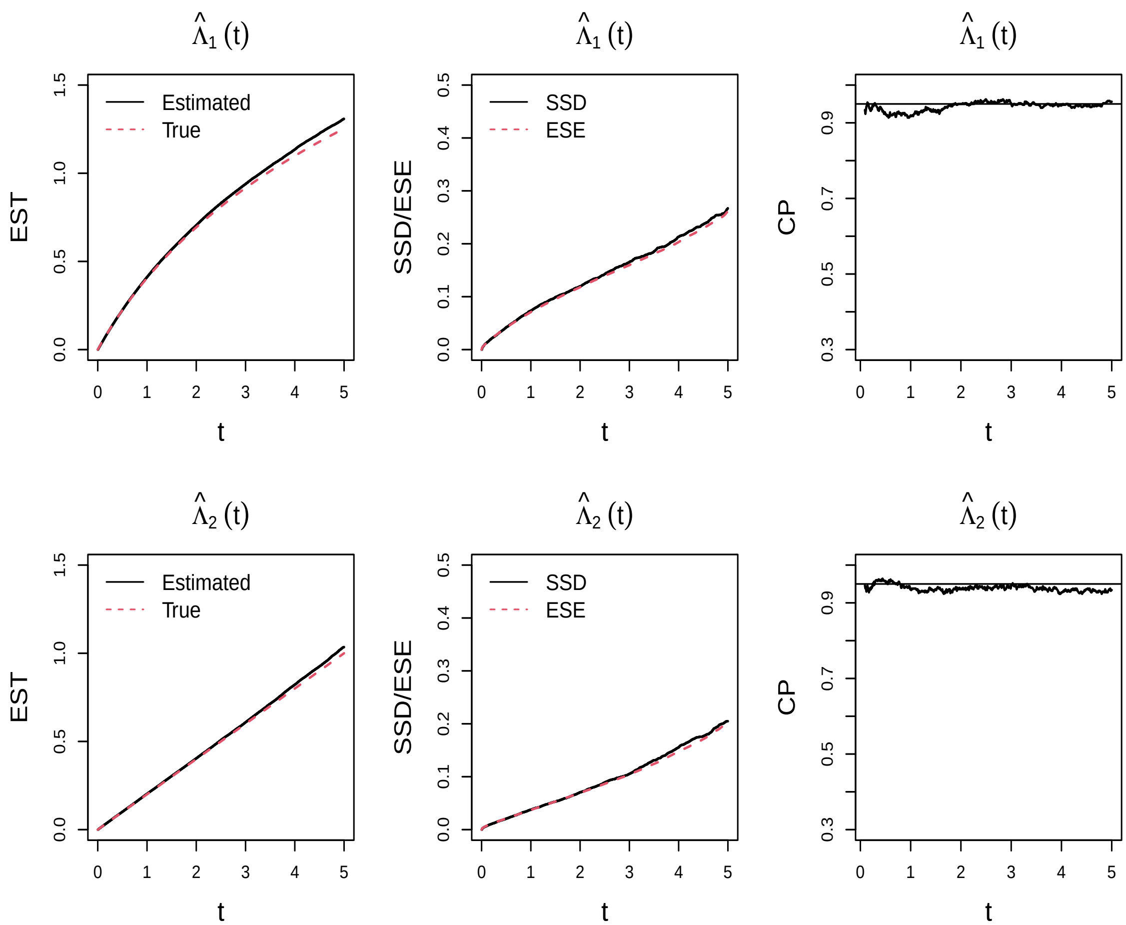

Figure 4.2: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 1 with $r = 0$ and $pt = 0.5$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

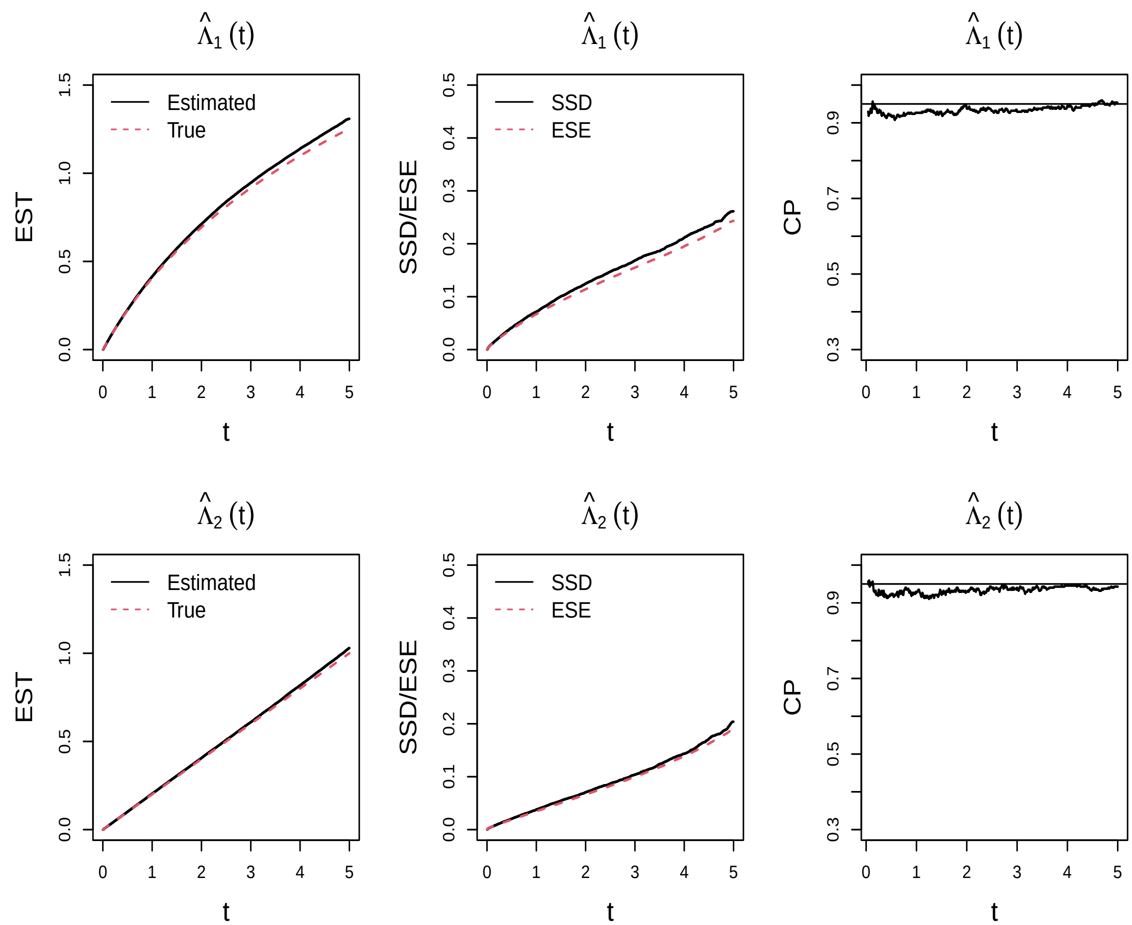

Figure 4.3: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 1 with $r = 0$ and $pt = 1$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

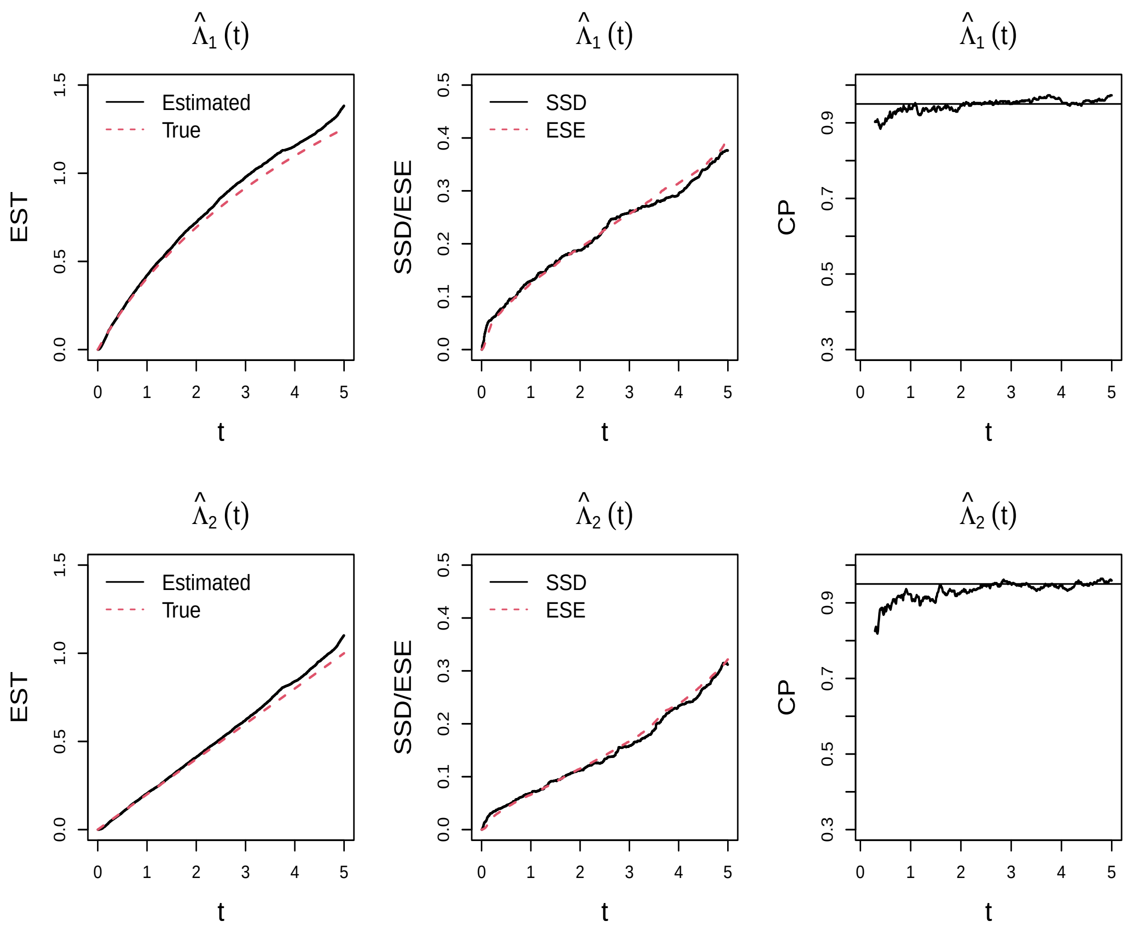

Figure 4.4: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 1 with $r = 0.5$ and $pt = 0$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

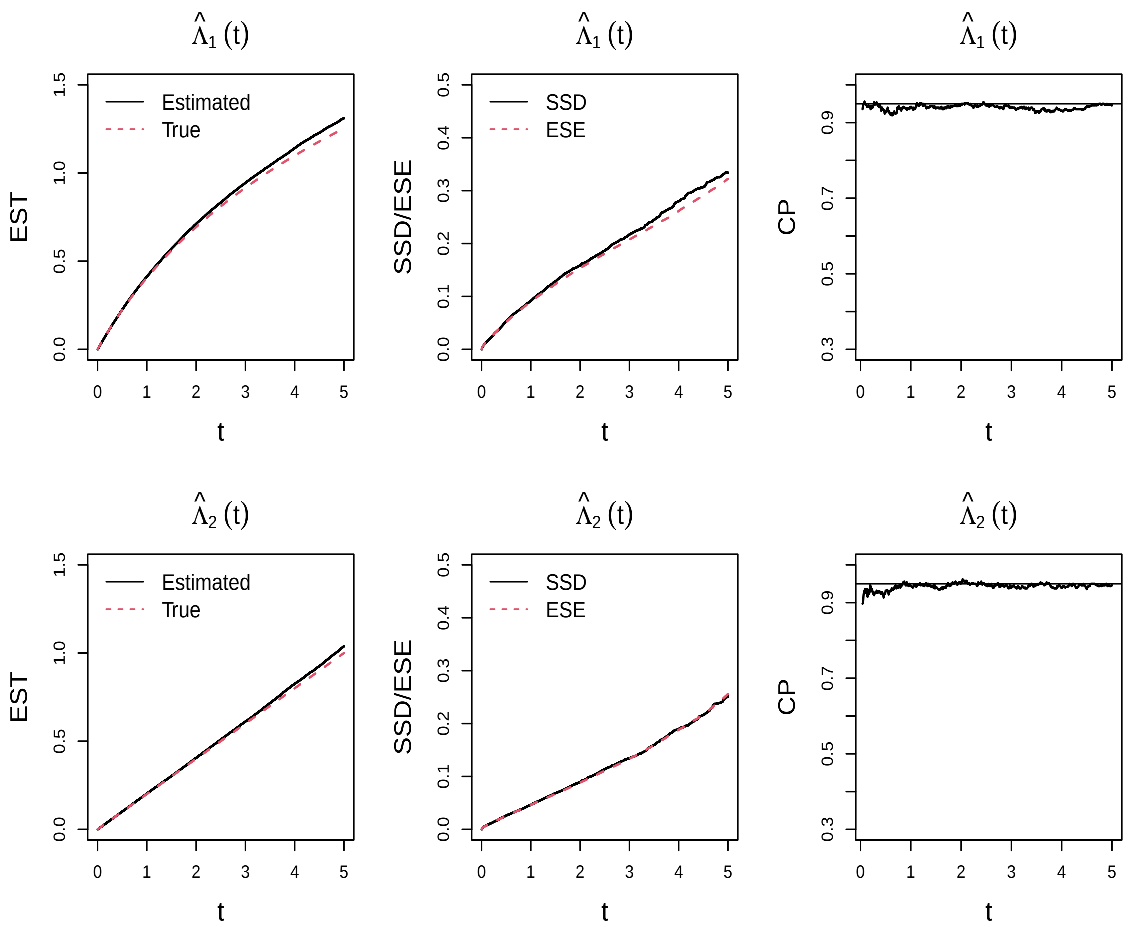

Figure 4.5: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 1 with $r = 0.5$ and $pt = 0.5$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

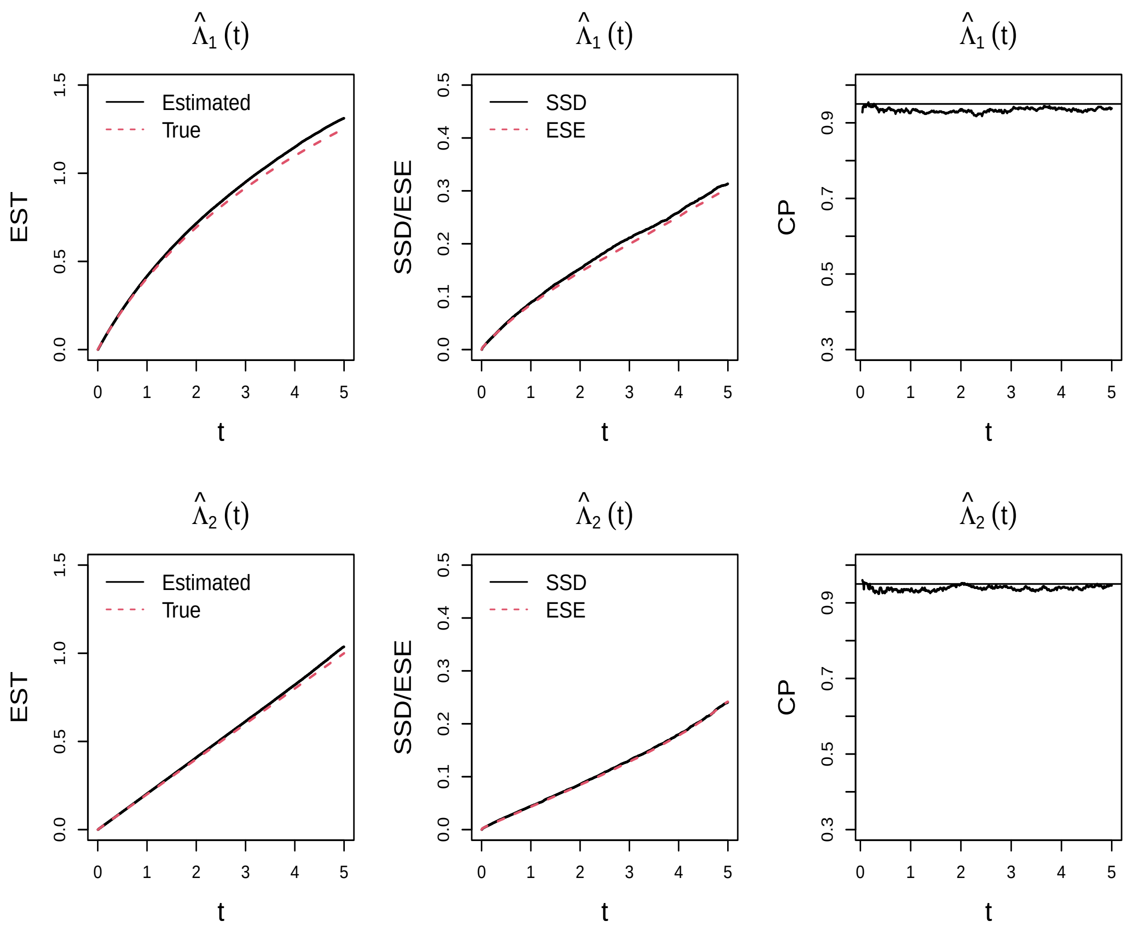

Figure 4.6: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 1 with $r = 0.5$ and $pt = 1$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

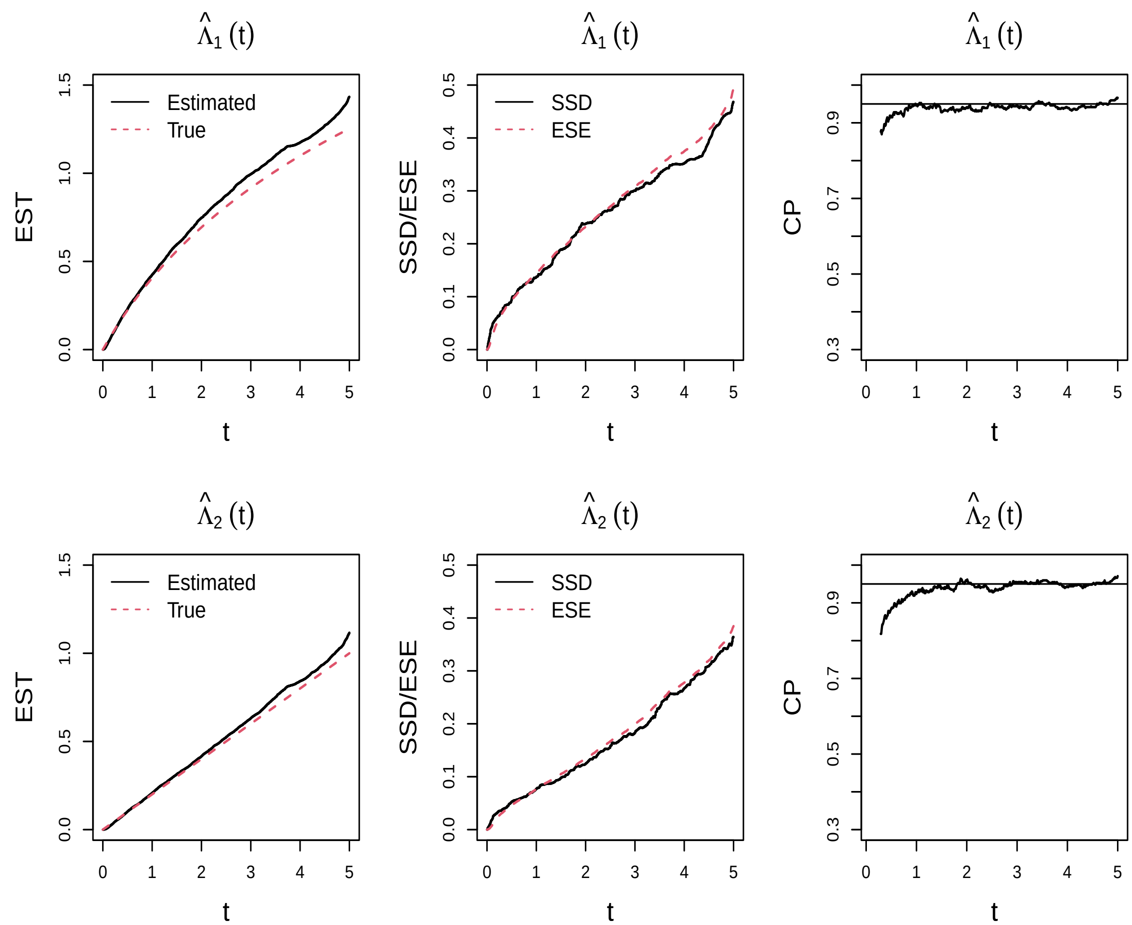

Figure 4.7: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 1 with $r = 1$ and $pt = 0$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

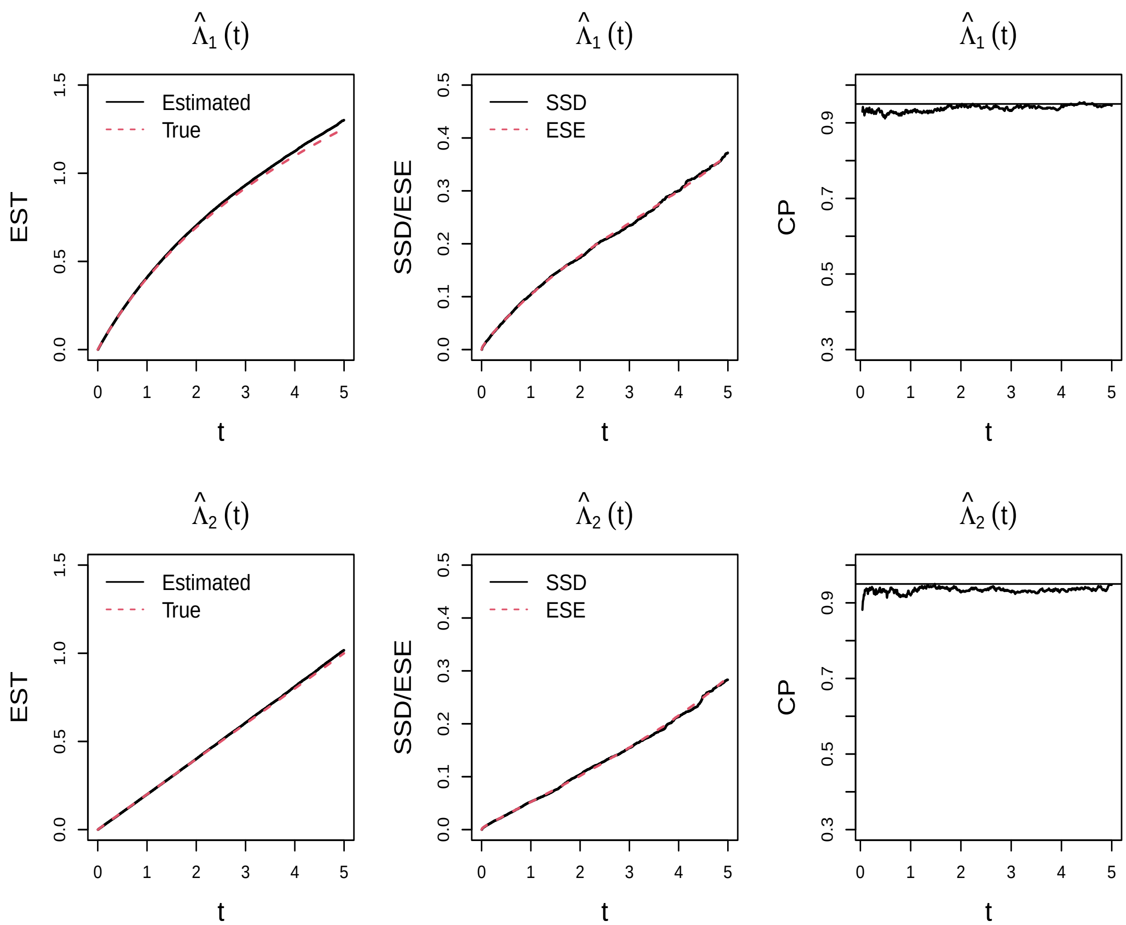

Figure 4.8: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 1 with $r = 1$ and $pt = 0.5$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

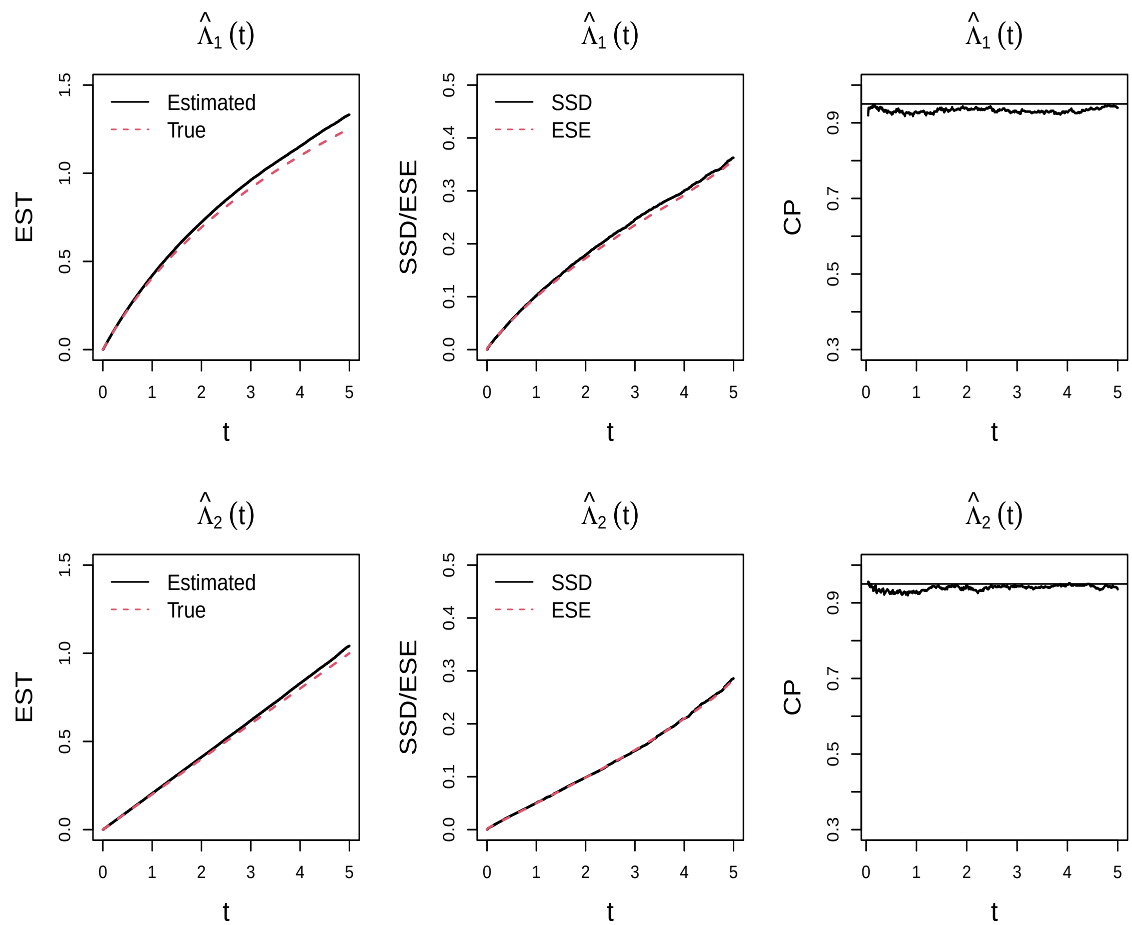

Figure 4.9: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 1 with $r = 1$ and $pt = 1$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

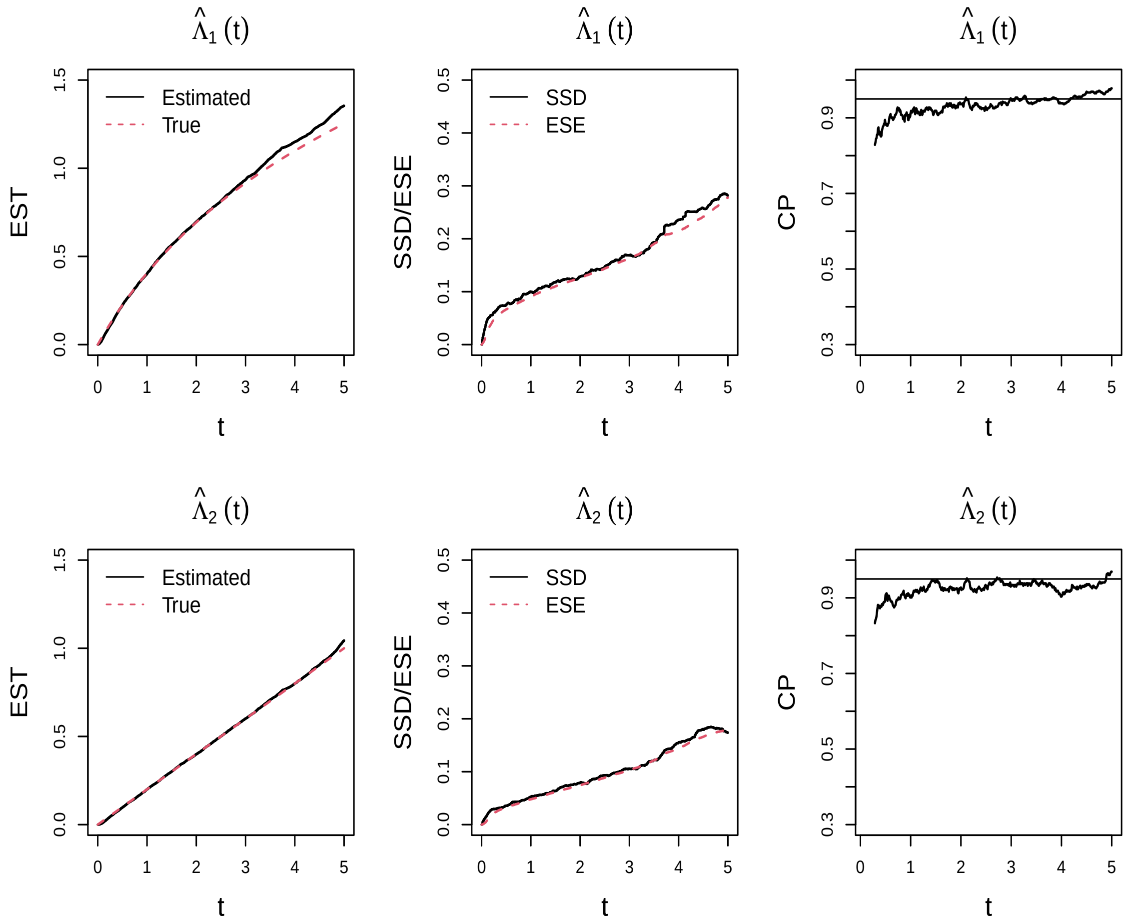

Figure 4.10: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 2 with $r = 0$ and $pt = 0$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

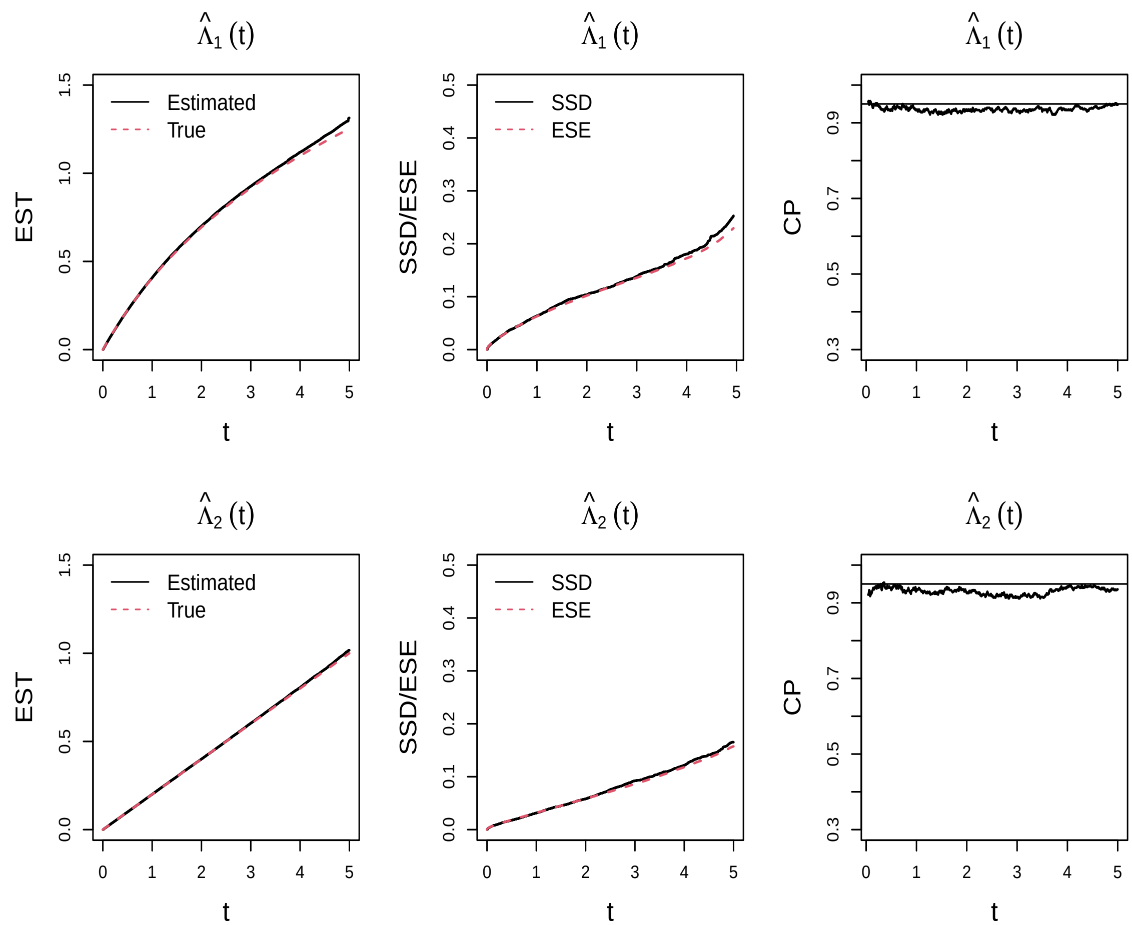

Figure 4.11: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 2 with $r = 0$ and $pt = 0.5$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

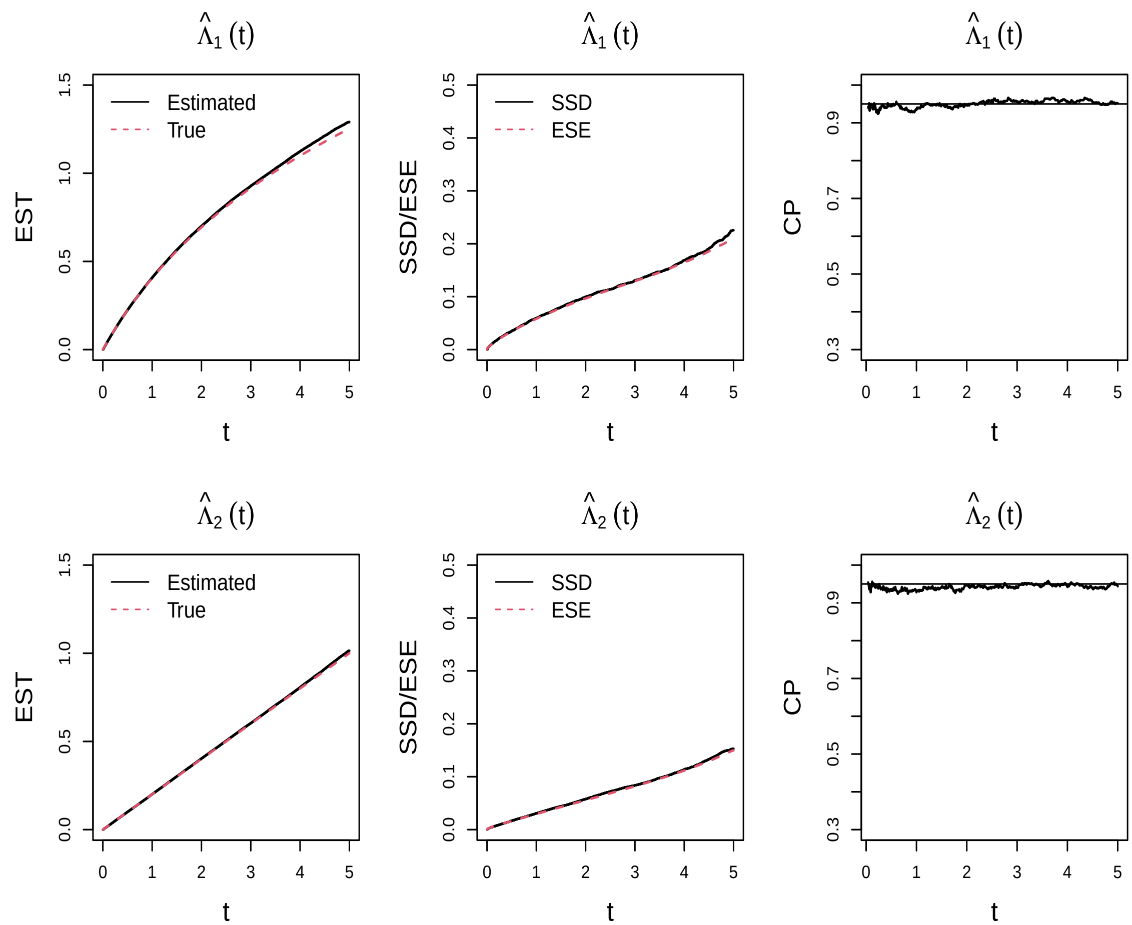

Figure 4.12: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 2 with $r = 0$ and $pt = 1$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

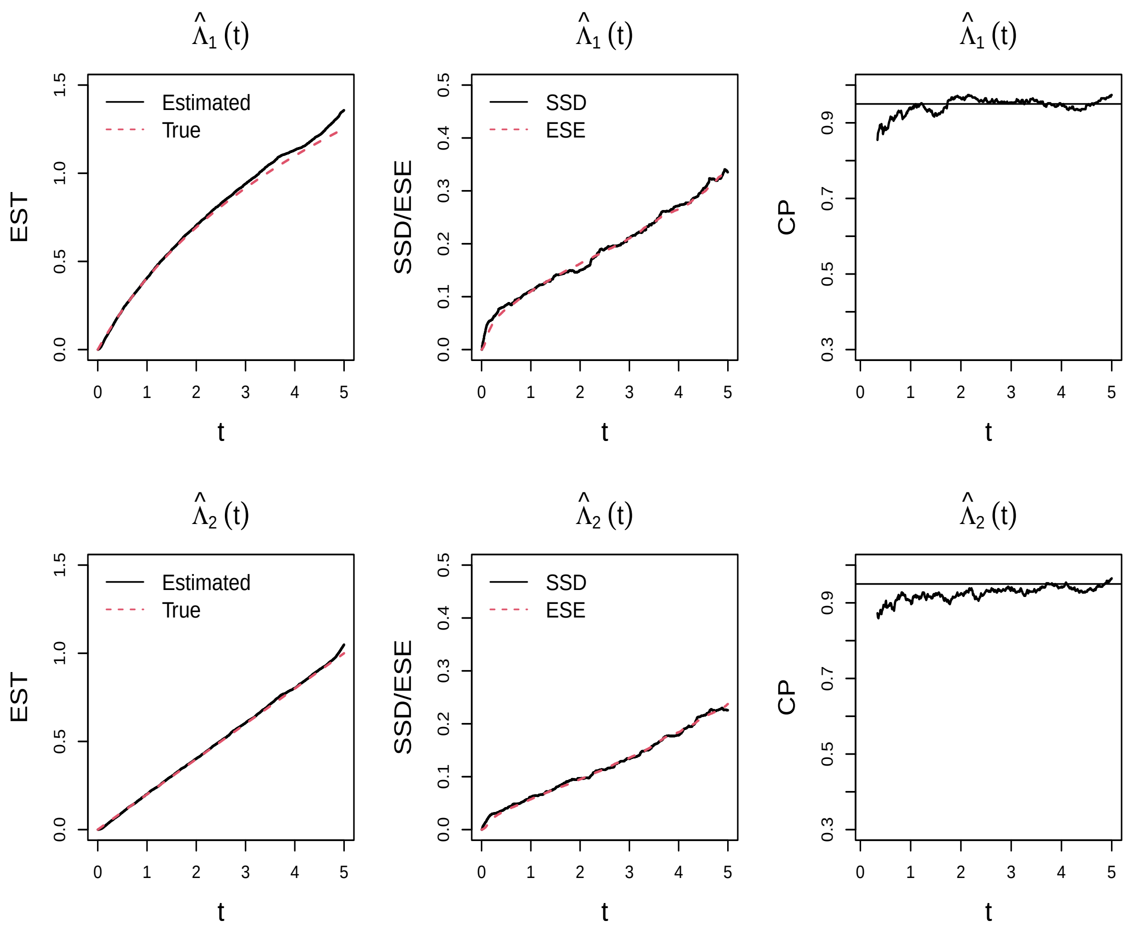

Figure 4.13: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 2 with $r = 0.5$ and $pt = 0$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

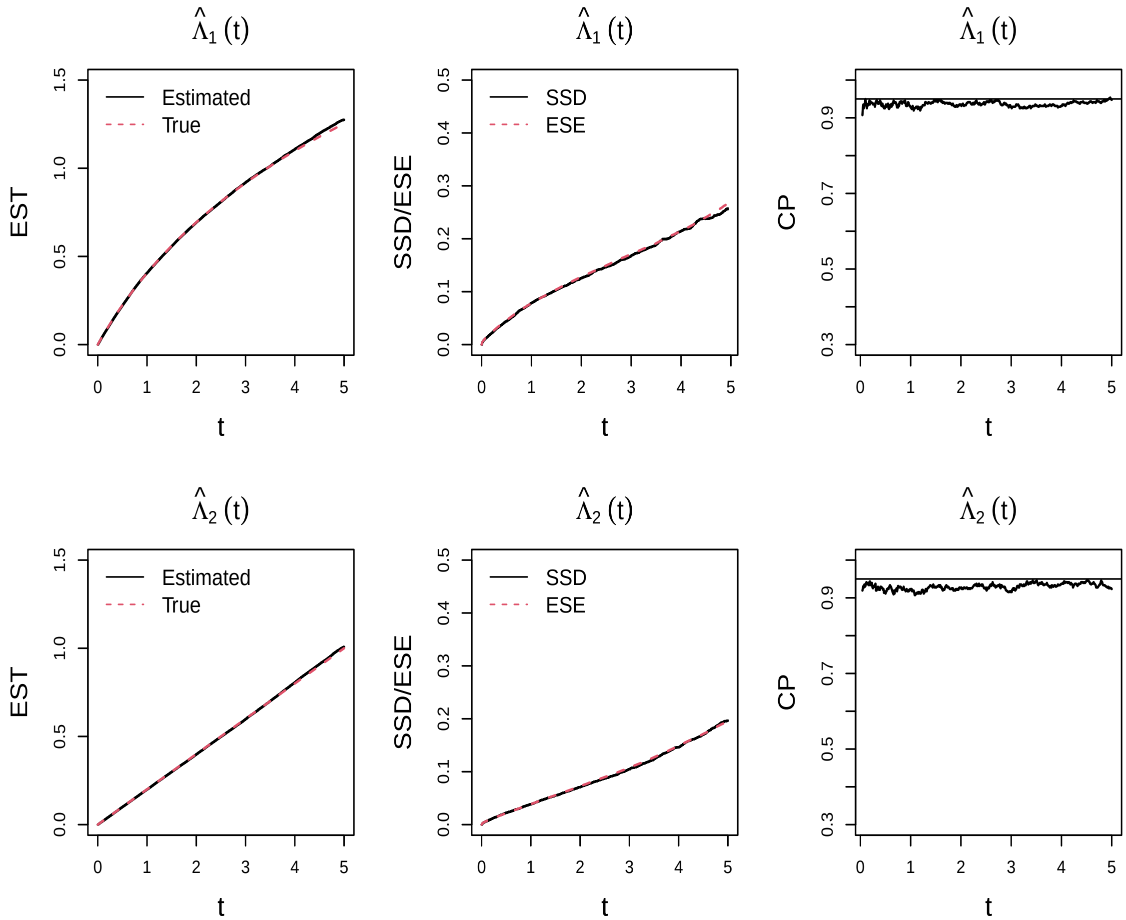

Figure 4.14: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 2 with $r = 0.5$ and $pt = 0.5$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

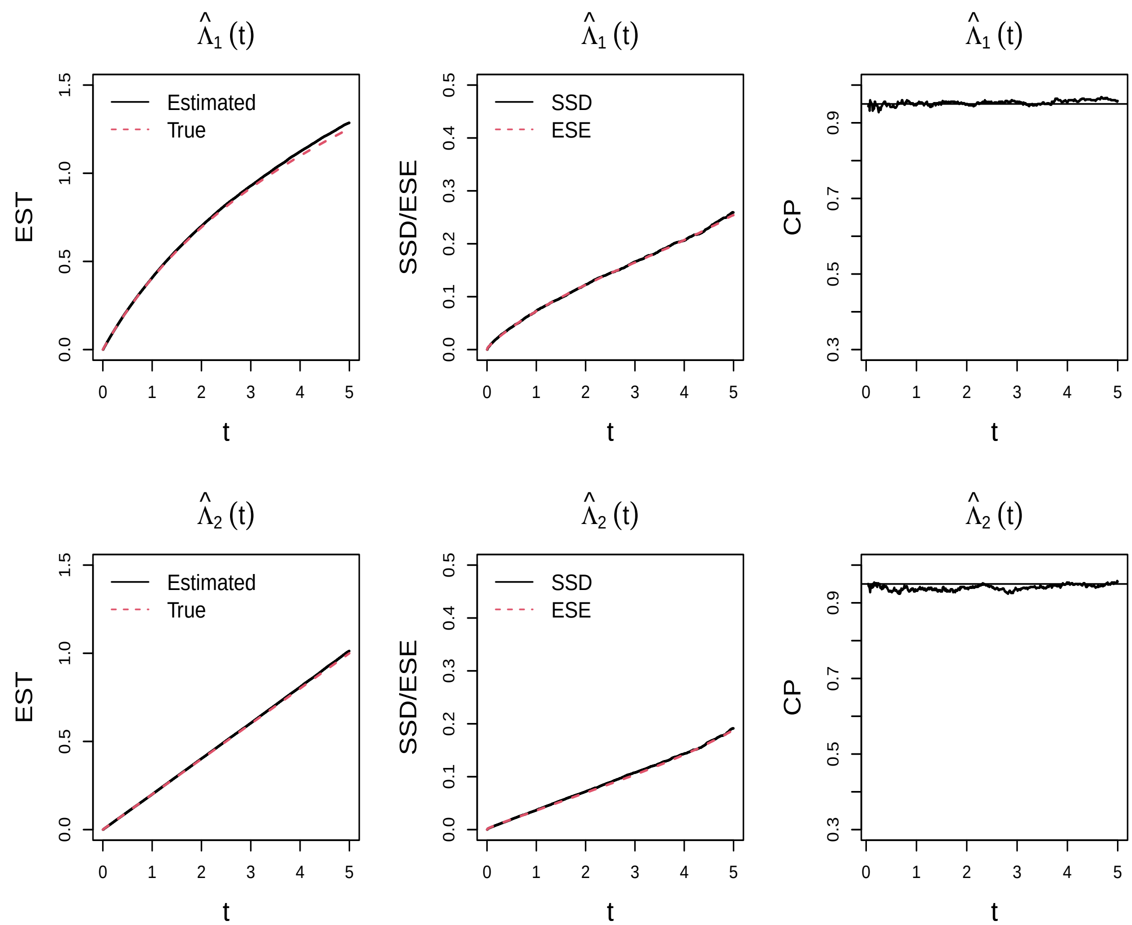

Figure 4.15: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 2 with $r = 0.5$ and $pt = 1$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

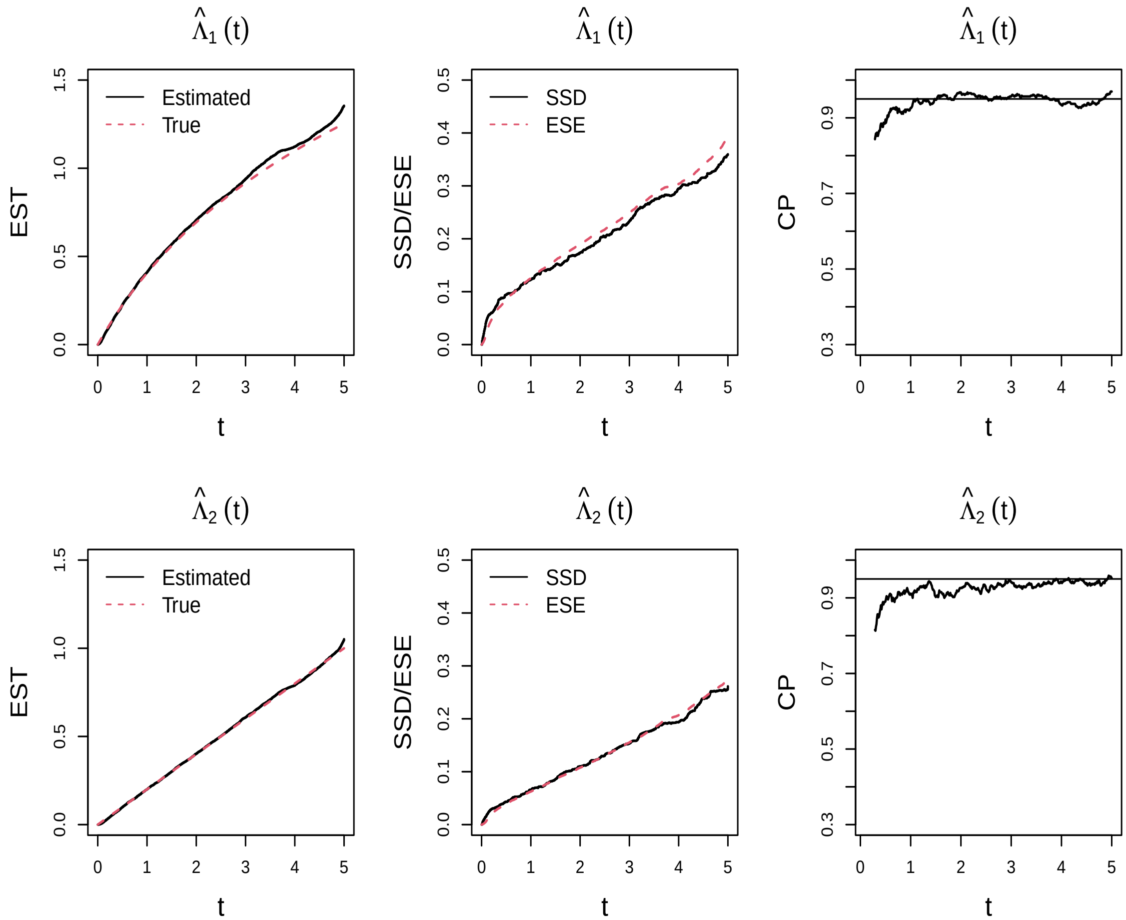

Figure 4.16: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 2 with $r = 1$ and $pt = 0$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

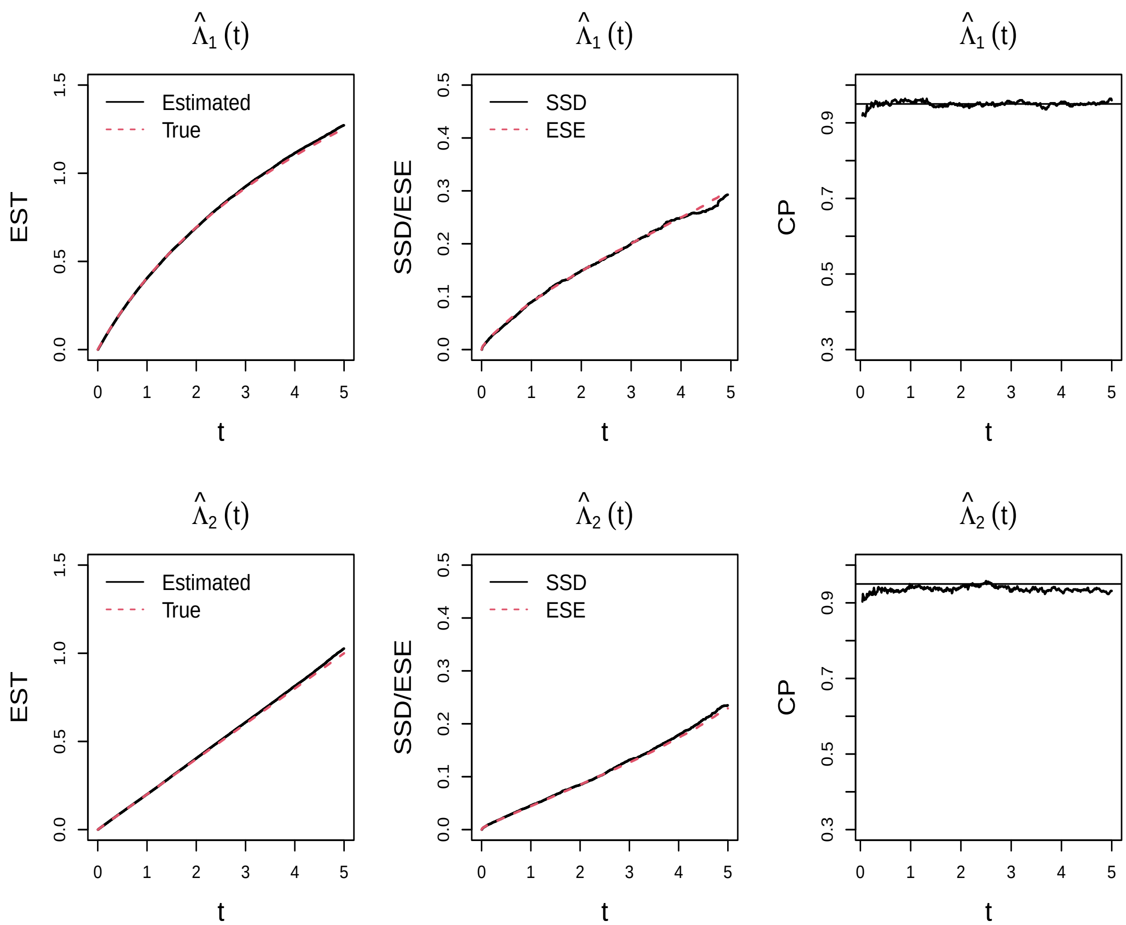

Figure 4.17: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 2 with $r = 1$ and $pt = 0.5$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

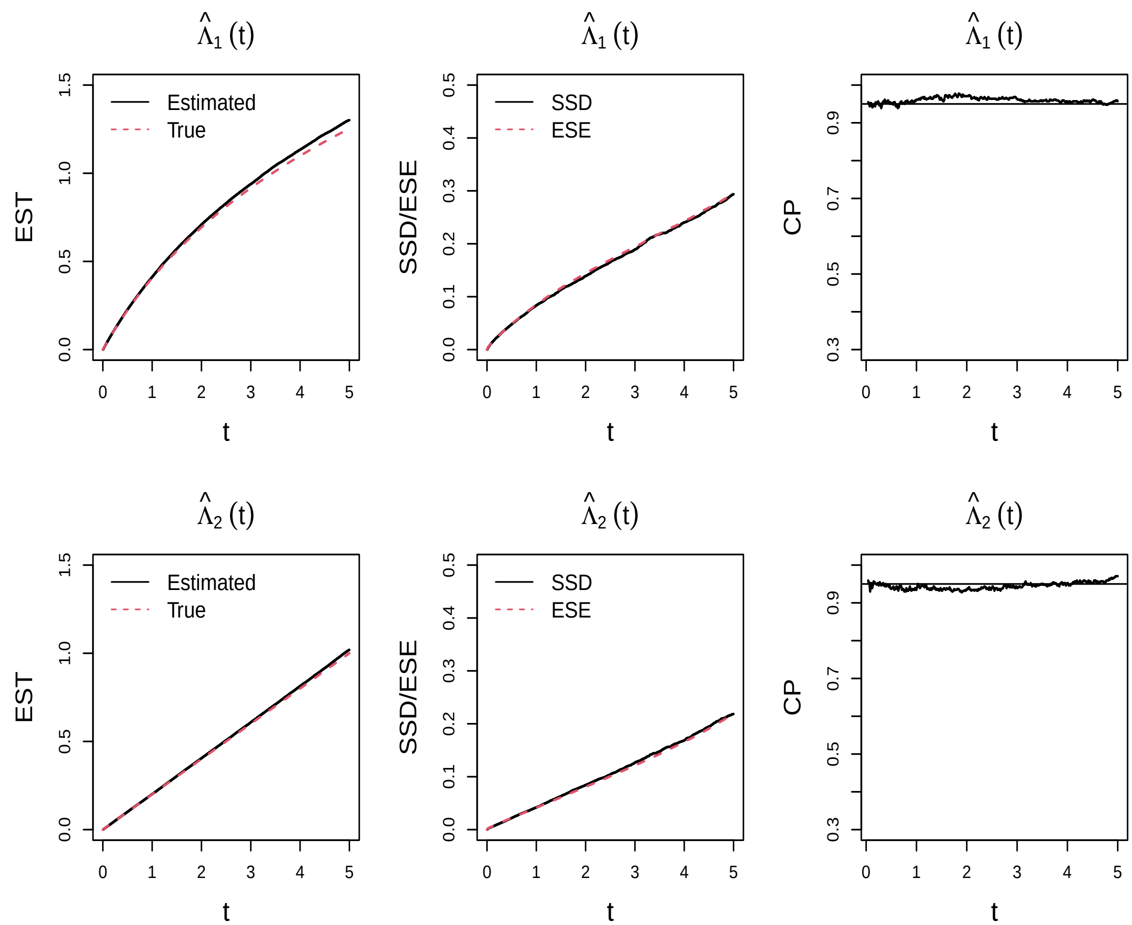

Figure 4.18: Estimation results for (a) $\Lambda_1(t) = \log(1 + \frac{t}{2})$ and (b) $\Lambda_2(t) = 0.2t$ in Scenario 2 with $r = 1$ and $pt = 1$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

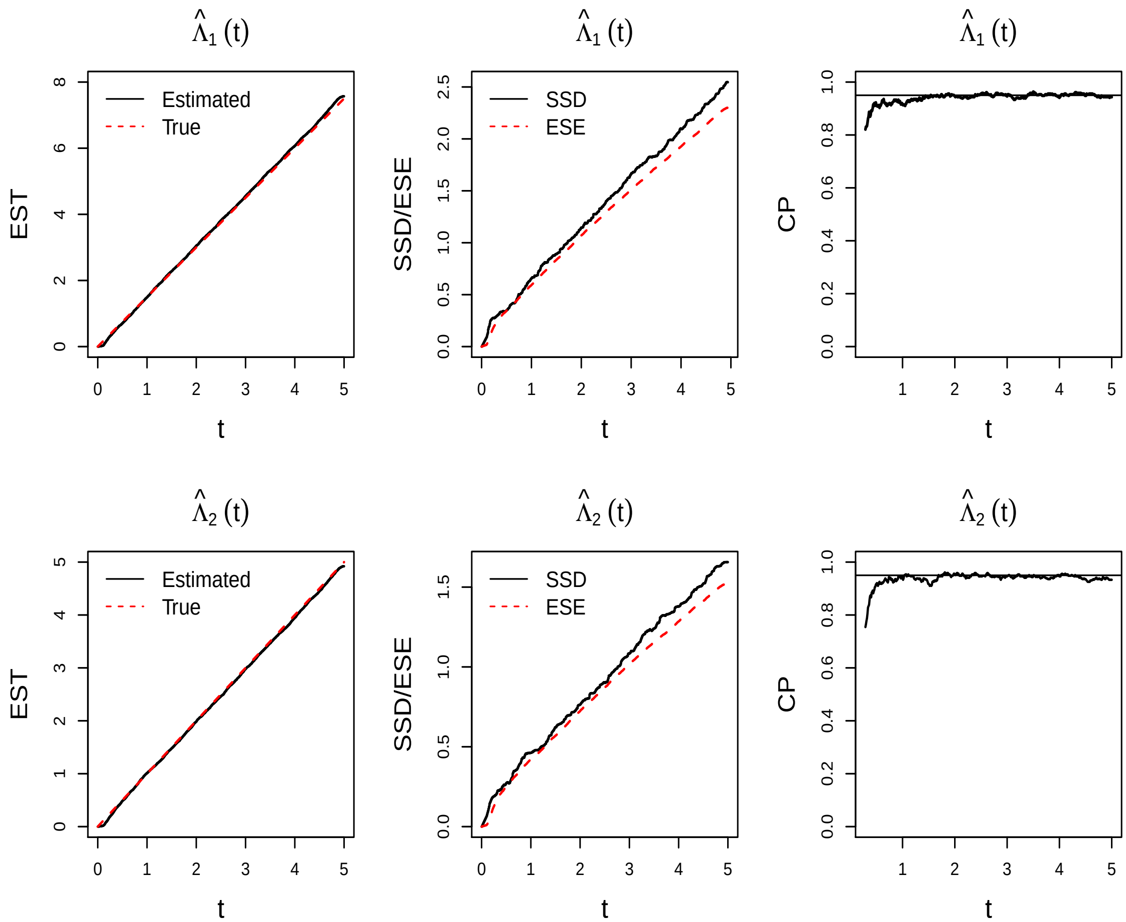

Figure 4.19: Estimation results for (a) $\Lambda_1(t) = 1.5t$ and (b) $\Lambda_2(t) = t$ in Scenario 3 with $r = 0$ and $pt = 0$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

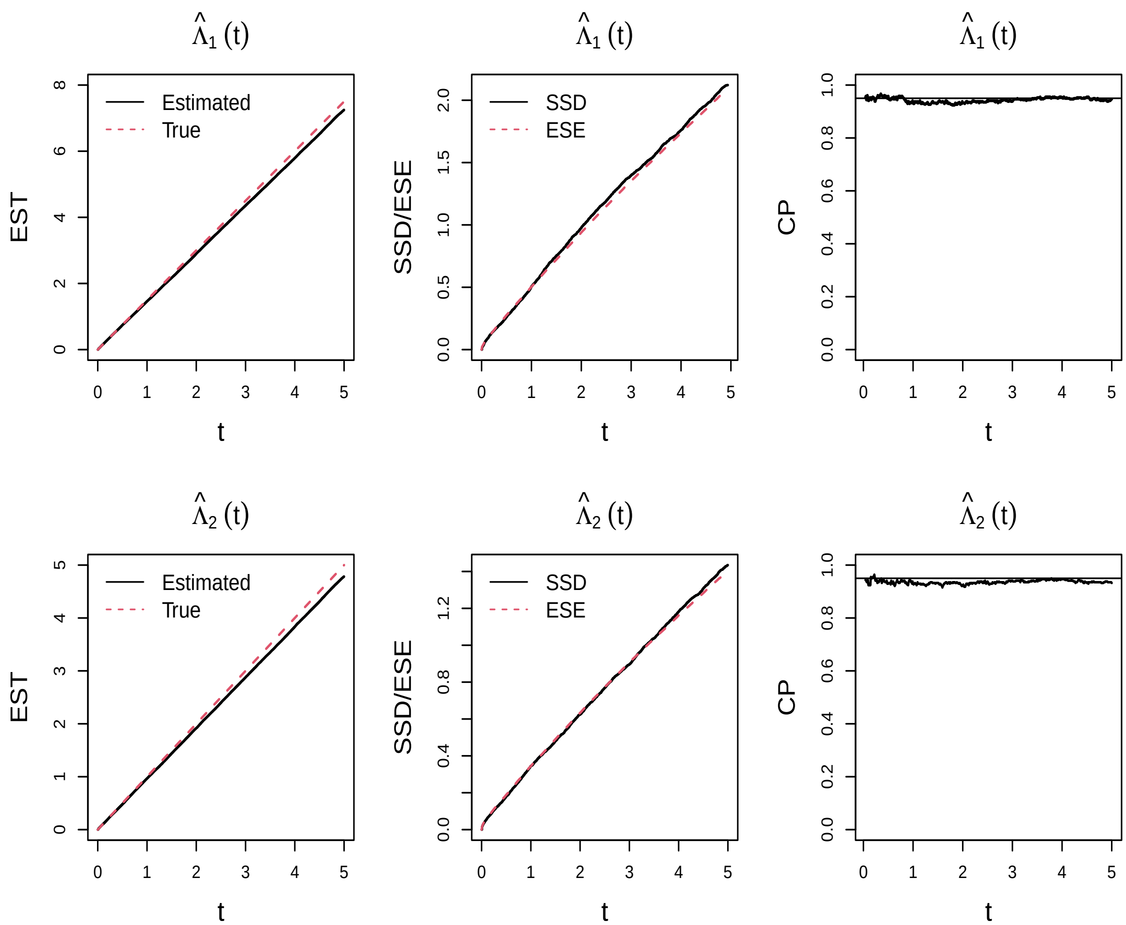

Figure 4.20: Estimation results for (a) $\Lambda_1(t) = 1.5t$ and (b) $\Lambda_2(t) = t$ in Scenario 3 with $r = 0$ and $pt = 0.5$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

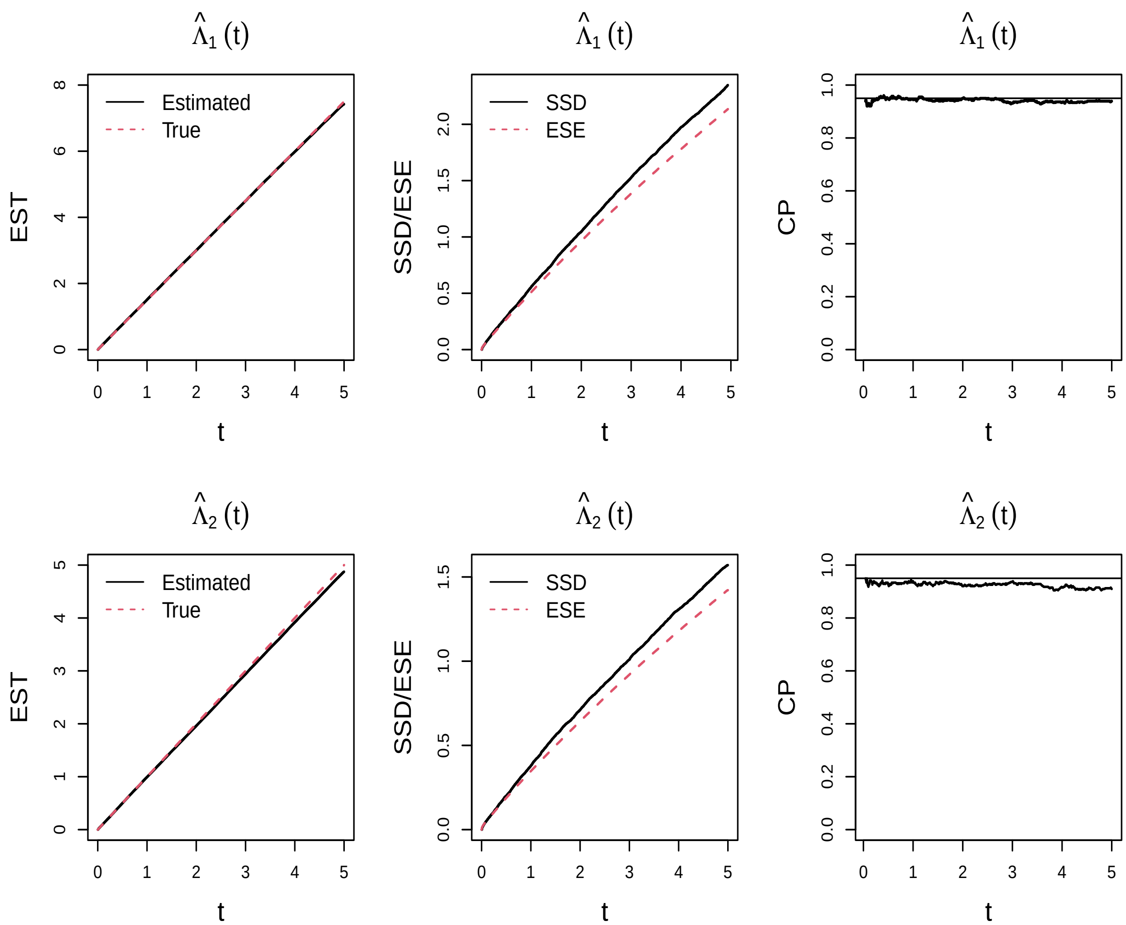

Figure 4.21: Estimation results for (a) $\Lambda_1(t) = 1.5t$ and (b) $\Lambda_2(t) = t$ in Scenario 3 with $r = 0$ and $pt = 1$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

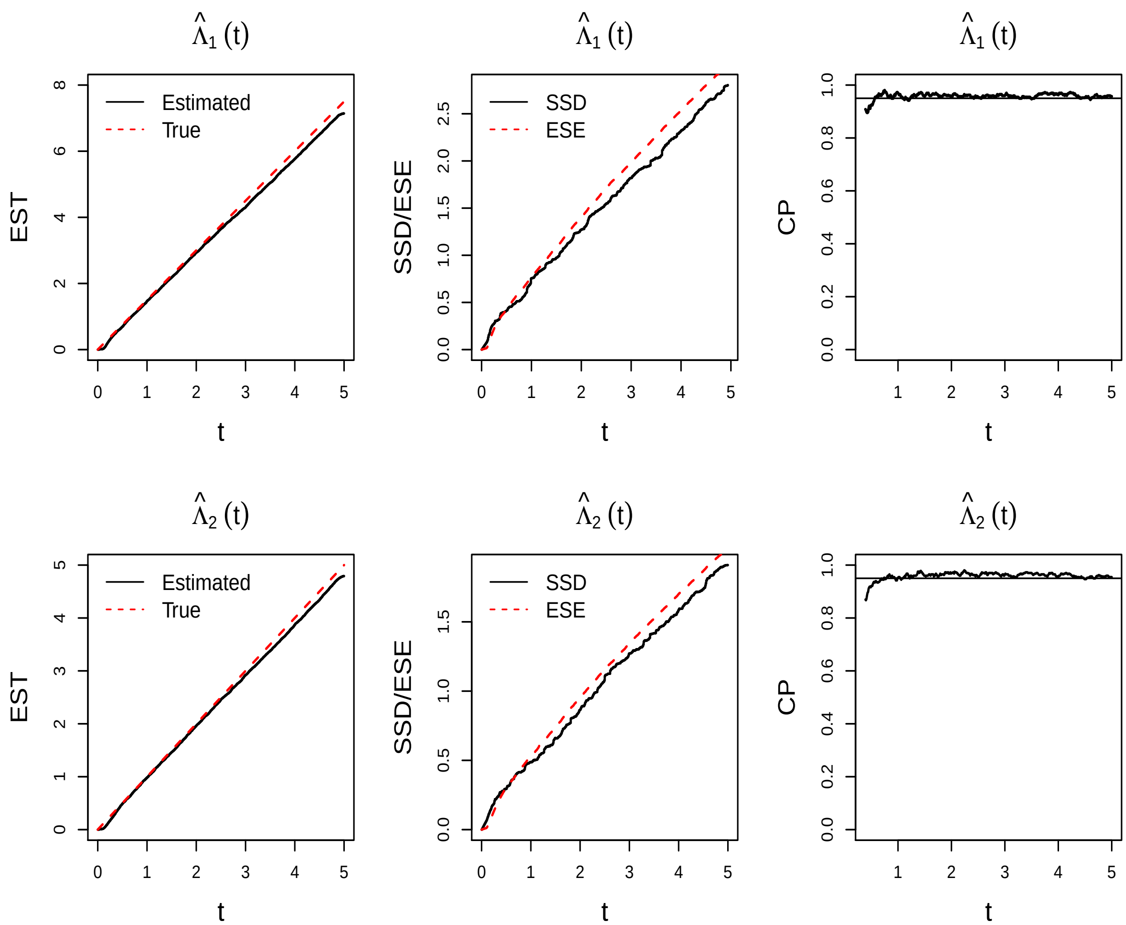

Figure 4.22: Estimation results for (a) $\Lambda_1(t) = 1.5t$ and (b) $\Lambda_2(t) = t$ in Scenario 3 with $r = 0.5$ and $pt = 0$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

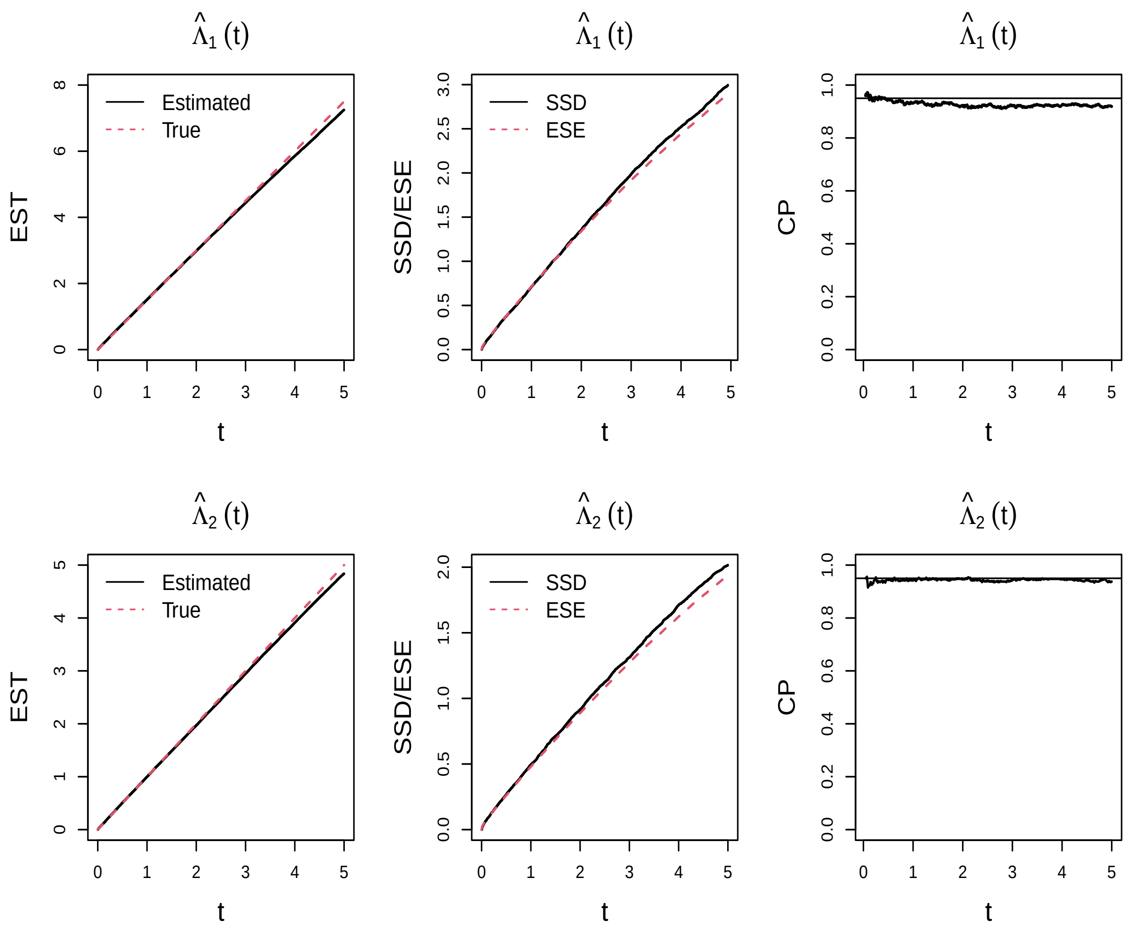

Figure 4.23: Estimation results for (a) $\Lambda_1(t) = 1.5t$ and (b) $\Lambda_2(t) = t$ in Scenario 3 with $r = 0.5$ and $pt = 0.5$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

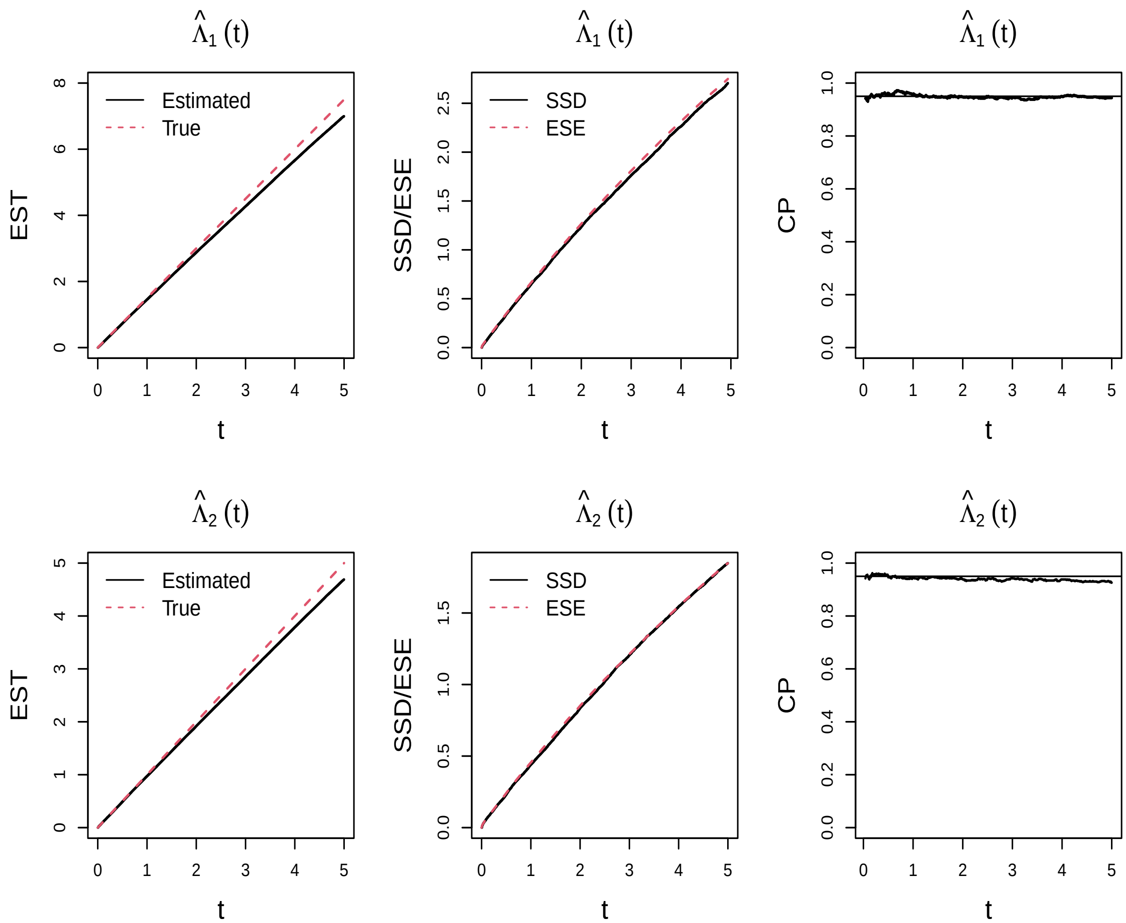

Figure 4.24: Estimation results for (a) $\Lambda_1(t) = 1.5t$ and (b) $\Lambda_2(t) = t$ in Scenario 3 with $r = 0.5$ and $pt = 1$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

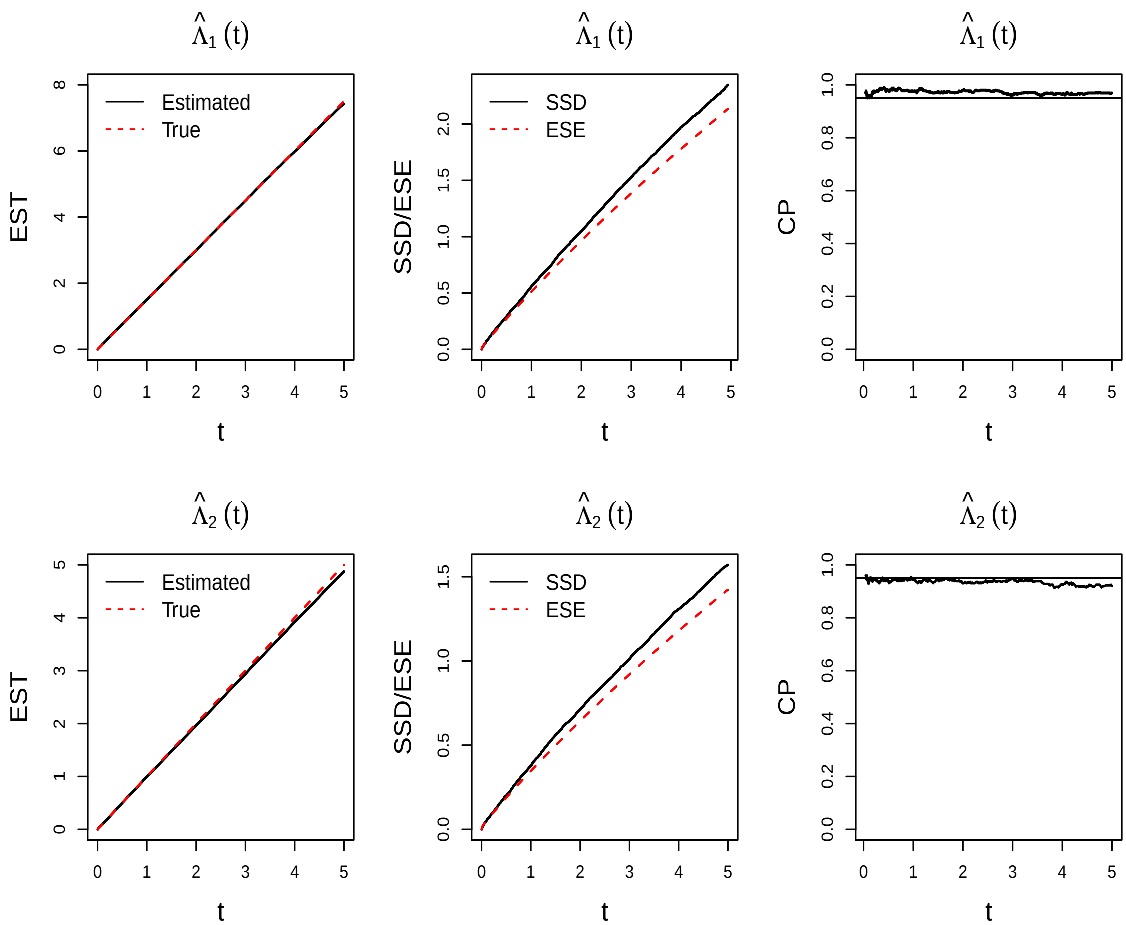

Figure 4.25: Estimation results for (a) $\Lambda_1(t) = 1.5t$ and (b) $\Lambda_2(t) = t$ in Scenario 3 with $r = 1$ and $pt = 0$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

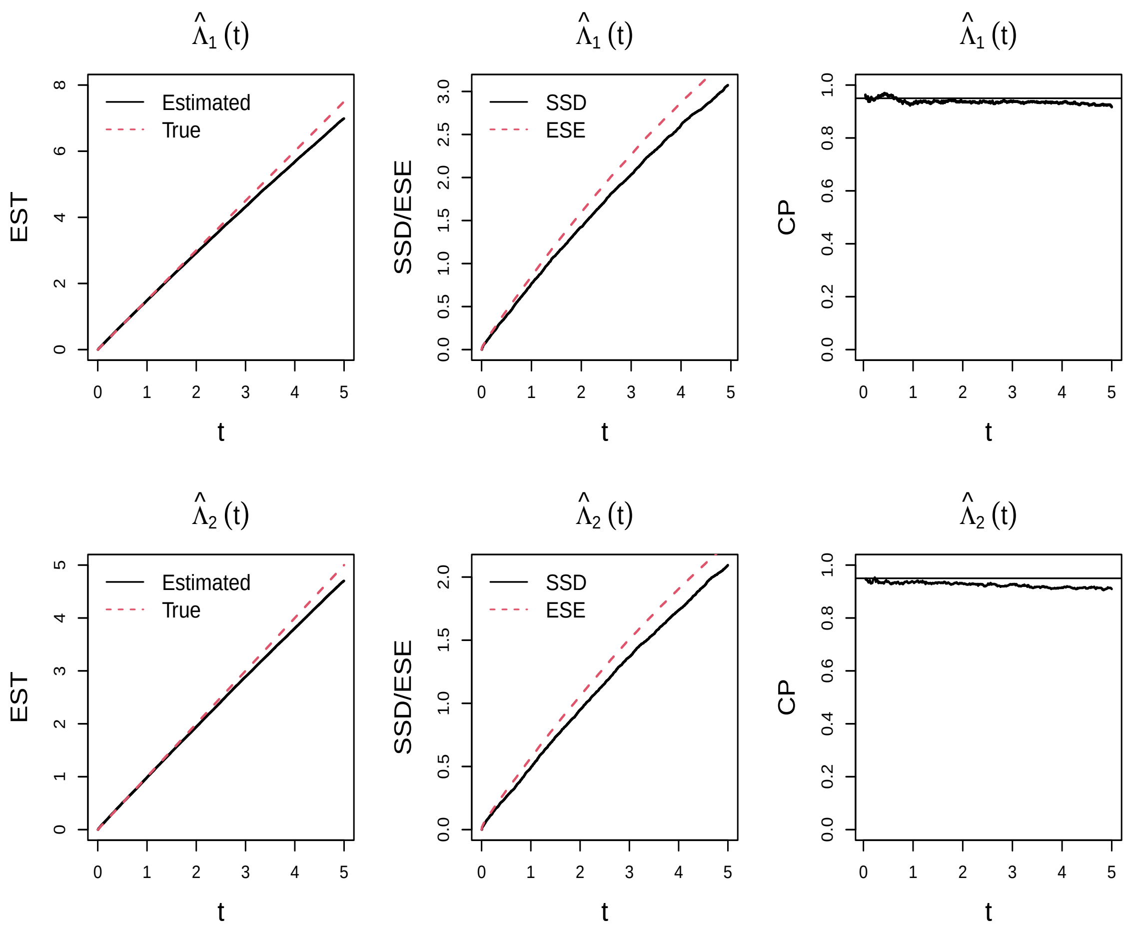

Figure 4.26: Estimation results for (a) $\Lambda_1(t) = 1.5t$ and (b) $\Lambda_2(t) = t$ in Scenario 3 with $r = 1$ and $pt = 0.5$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

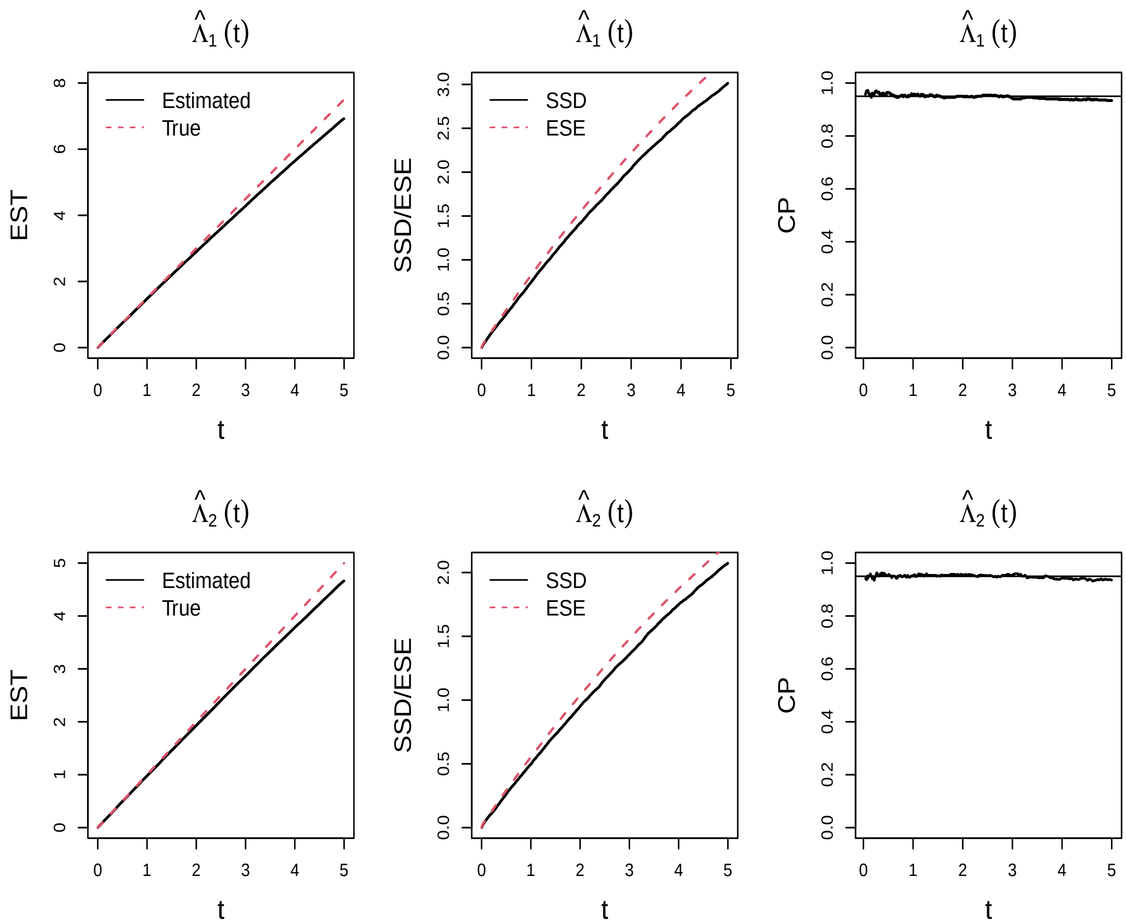

Figure 4.27: Estimation results for (a) $\Lambda_1(t) = 1.5t$ and (b) $\Lambda_2(t) = t$ in Scenario 3 with $r = 1$ and $pt = 1$. Results comes from $n = 1200$. Est means the average of baseline hazards over 500 replicates, SE, SEE, and CP comes from descriptions before

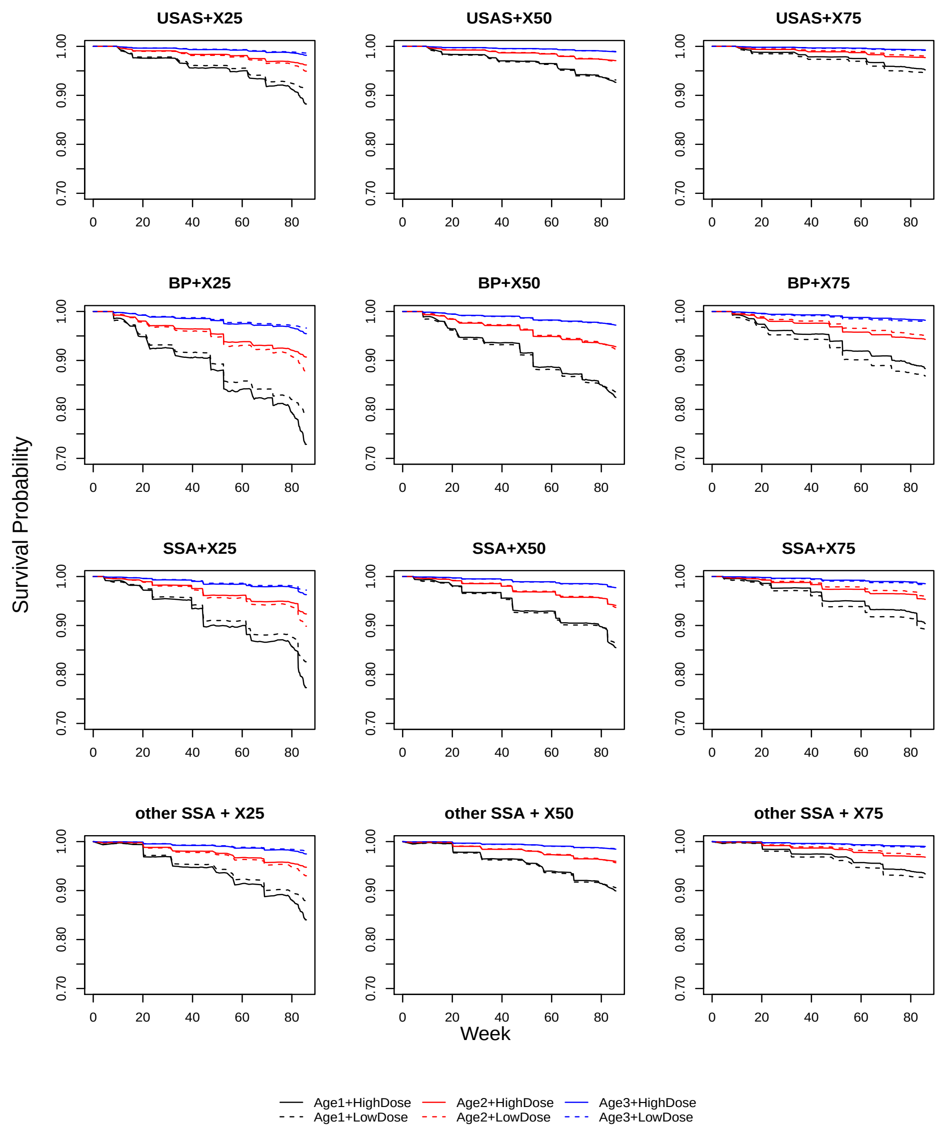

Figure 4.28: Estimated survival curves under the proportional hazards model: $\Lambda_j(t|X(\cdot), Z) = \int_0^t \exp\{\beta_1 X(s) + \beta_2 X(s)\text{HighDose} + \gamma^\top Z\} d\Lambda_j(s)$ for regions USAS, BP, SSA, and other SSA, where $j = 1, \cdots, 4$. Here, $X(\cdot)$ represents log(VRC) over time, and Z denotes time-independent covariates, including age groups and high-dose treatment groups. Here, 'Age1', 'Age2' and 'Age3' stand for the age goups, < 20, $[20, 30]$, > 30, respectively and $X25$, $X50$, $X75$ represent logVRC at its 25th, 50th, 75th percentiles, respectively. The title of each figure, for example, 'USAS+X25' corresponds to the region USAS with log(VRC) being the 25th percentile of log(VRC).

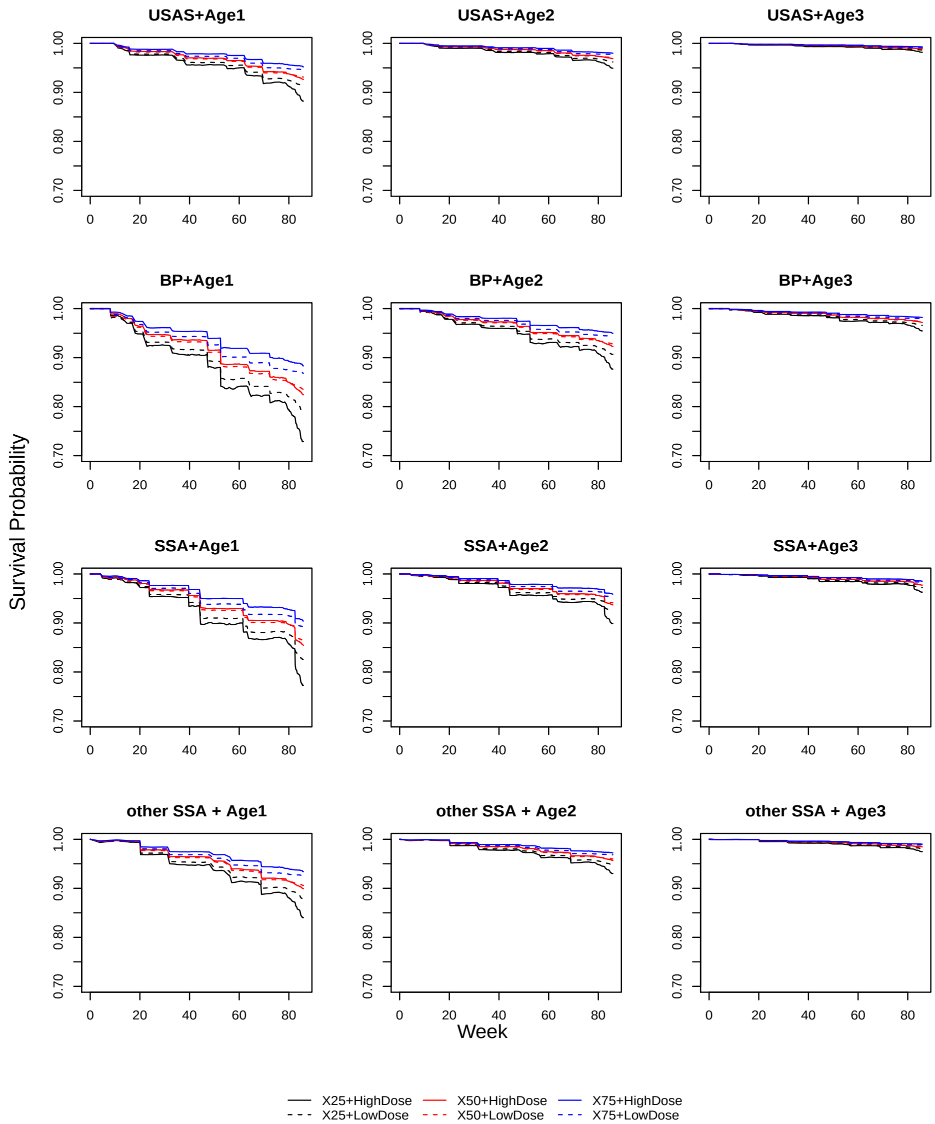

Figure 4.29: Estimated survival curves by considering different combinations of covariates under the proportional hazards model $\Lambda_j(t|X(\cdot), Z) = \int_0^t \exp\{\beta_1 X(s) + \beta_2 X(s)\text{HighDose} + \gamma^\top Z\} d\Lambda_j(s)$, $j = 1, \cdots, 4$, for regions USAS, BP, SSA, other SSA. $X(\cdot)$ denotes the log(VRC) over time and Z denotes time-independent covariates, including age groups and high-dose treatment groups. Here, 'Age1', 'Age2' and 'Age3' stand for the age goups, < 20, $[20, 30]$, > 30, respectively and $X25$, $X50$, $X75$ represent logVRC at its 25th, 50th, 75th percentiles, respectively. The title of each figure, for example, 'USAS+Age1' corresponds to the region USAS with Age less than 20.

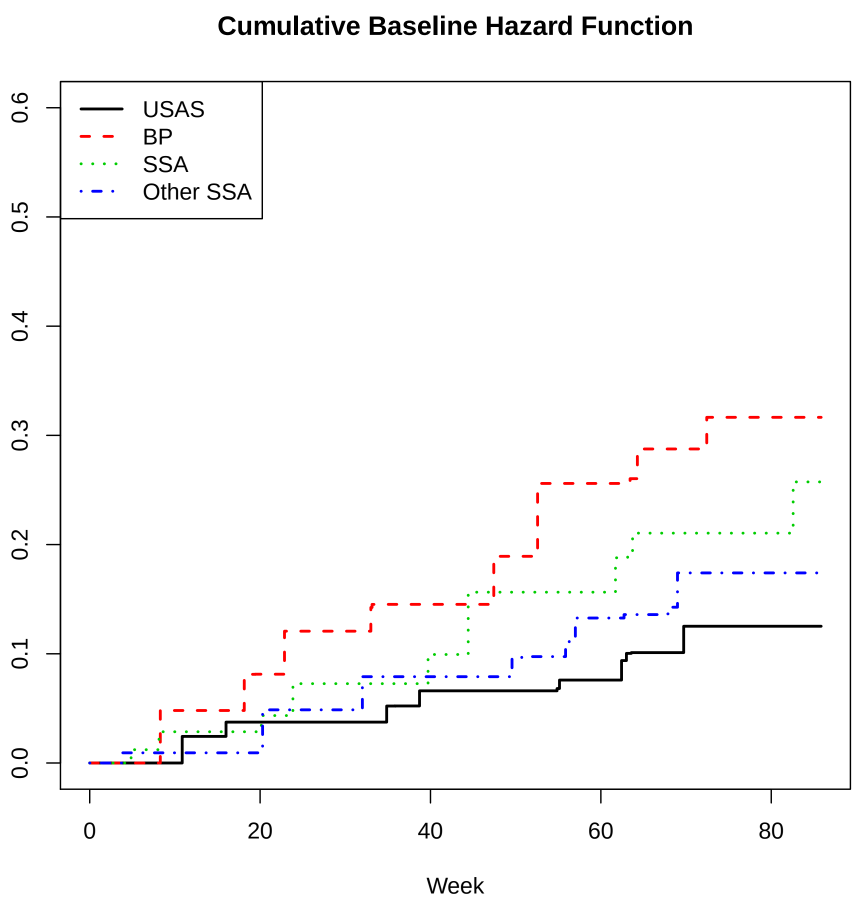

Figure 4.30: Estimated baseline cumulative hazard function, $\hat{\Lambda}_j(t)$, $j = 1, \cdots, 4$, for regions USAS, BP, SSA, other SSA under the proportional hazards model $\Lambda_j(t|X(\cdot), Z) = \int_0^t \exp\{\beta_1 X(s) + \beta_2 X(s)\text{HighDose} + \gamma^\top Z\} d\Lambda_j(s)$, $j = 1, \cdots, 4$ denotes log(VRC) over time and Z denotes time-independent covariates, including age groups and high-dose treatment groups.

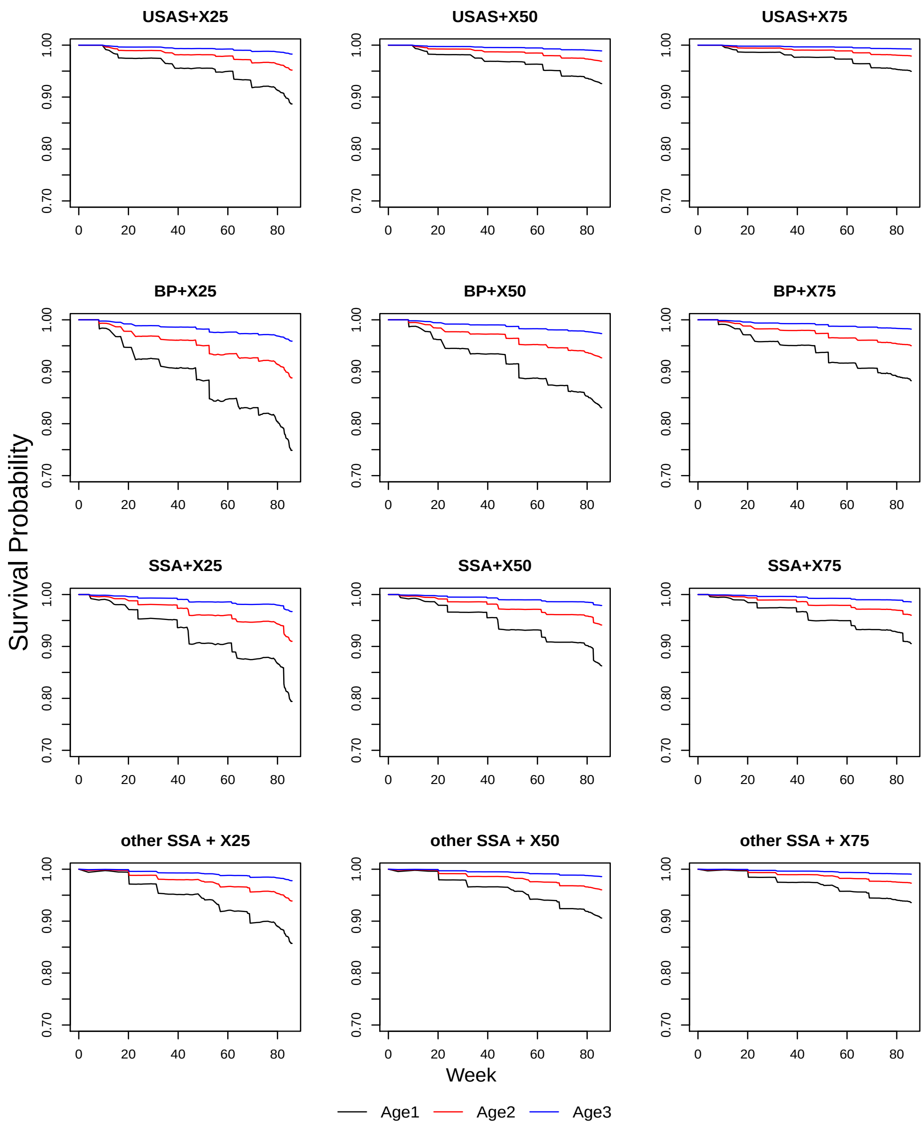

Figure 4.31: Estimated survival curves by considering different combinations of covariates under the proportional hazards model $\Lambda_j(t|X(\cdot), Z) = \int_0^t \exp\{\beta X(s) + \gamma^\top Z\} d\Lambda_j(s)$, $j = 1, \cdots, 4$, for regions USAS, BP, SSA, other SSA. $X(\cdot)$ denotes logVRC over time and Z denotes time independent covariates, the age groups. Here, 'Age1', 'Age2' and 'Age3' stand for the age goups, < 20, $[20, 30]$, > 30, respectively. The title of each figure, for example, 'USAS+X25' corresponds to the region USAS with X being the 25th percentile of log(VRC).

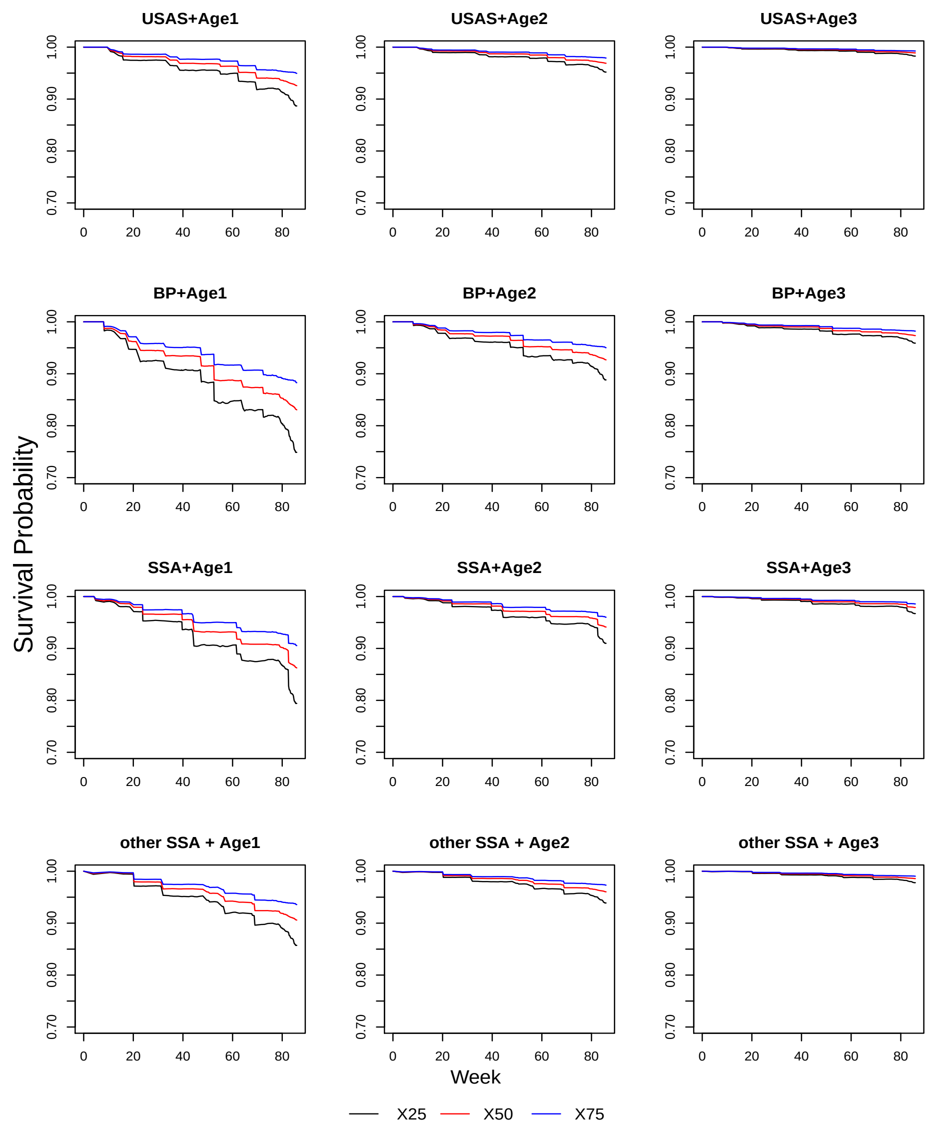

Figure 4.32: Estimated survival curves by considering different combinations of covariates under the proportional hazards model $\Lambda_j(t|X(\cdot), Z) = \int_0^t \exp\{\beta X(s) + \gamma^\top Z\} d\Lambda_j(s)$, $j = 1, \cdots, 4$, for regions USAS, BP, SSA, other SSA. $X(\cdot)$ denotes logVRC over time and Z denotes time independent covariates, the age groups. Here, 'Age1', 'Age2' and 'Age3' stand for the age groups, < 20, $[20, 30]$, > 30, respectively. The title of each figure, for example, 'USAS+Age1' corresponds to the region USAS with Age < 20.

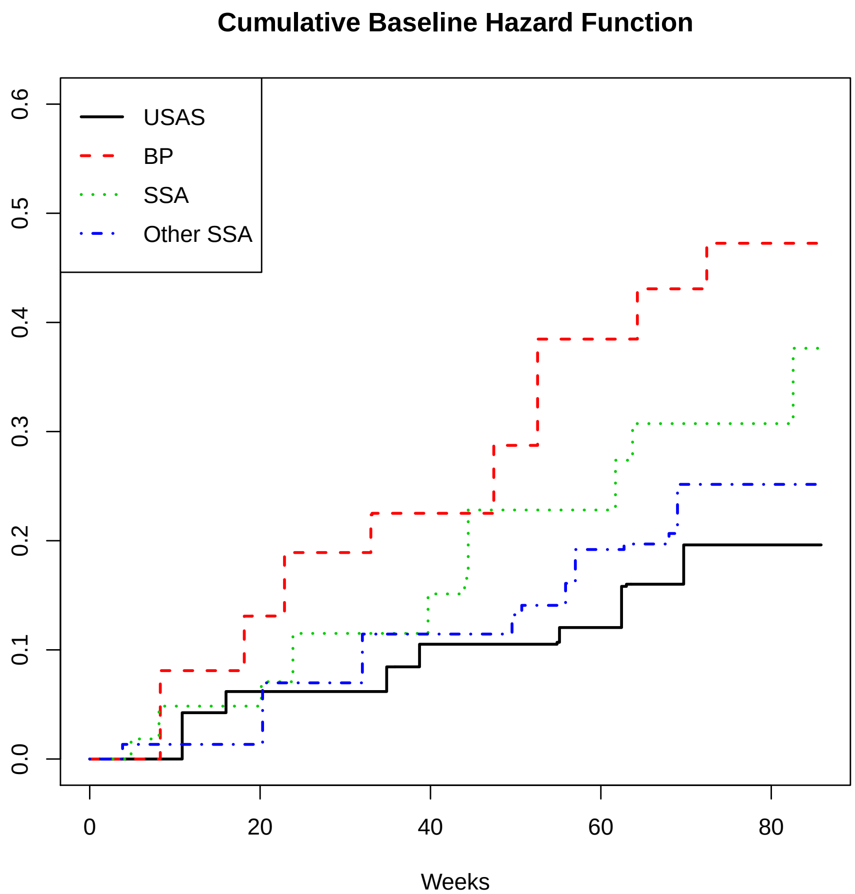

Figure 4.33: Estimated baseline cumulative hazard function, $\hat{\Lambda}_j(t)$, $j = 1, \cdots, 4$, for regions USAS, BP, SSA, other SSA under the proportional hazards model $\Lambda_j(t|X(\cdot), Z) = \int_0^t \exp\{\beta X(s) + \gamma^\top Z\} d\Lambda_j(s)$, where $X(\cdot)$ denotes logVRC over time and Z denotes time independent covariates, the age groups.

Table 4.4: Summary of HIV data

	Characteristic		Placebo	Low Dose Total (VRC)	High-Dose Total (VRC)
	Size		637	642 (51)	645 (41)
	Gender	Male	0	0 (0)	0 (0)
		Female	637	642 (51)	645 (41)
HVTN-703	Age	< 20	38	54 (2)	57 (5)
		$20 - 30$	454	449 (38)	443 (29)
		> 30	145	139 (11)	145 (7)
	Region	SSA	340	338 (27)	341 (25)
		Other SSA	286	292 (24)	294 (16)
	Size		898	895 (53)	894 (51)
	Gender	Male	887	888 (53)	886 (51)
		Female	11	7 (0)	8 (0)
HVTN-704	Age	< 20	57	57 (9)	65 (11)
		$20 - 30$	531	508 (30)	518 (28)
		> 30	310	330 (14)	311 (12)
	Region	USAS	470	474 (14)	469 (13)
		BP	428	421 (39)	425 (38)
	Size		1535	1537 (104)	1539 (92)
	Gender	Male	887	888 (53)	886 (51)
		Female	653	649 (51)	653 (41)
		< 20	95	111 (11)	122 (16)
Combined	Age	$20 - 30$	985	957 (68)	961 (57)
		> 30	455	469 (25)	456 (19)
		USAS	470	474 (14)	469 (13)
	Region	BP	428	421 (39)	425 (38)
		SSA	340	338 (27)	341 (25)
		Other SSA	286	292 (24)	294 (16)

Table 4.5: Results from fitting the logistic model (2.14)

		Estimate	Std. Error	z value	Pr(>|z|)
	(Intercept)	-3.3464	0.2275	-14.71	0.0000
Combined	USAS	-0.9087	0.3602	-2.52	0.0117
	BP	0.1353	0.2921	0.46	0.6433
	SA	0.1243	0.3055	0.41	0.6841
HVTN-703	(Intercept)	-3.3464	0.2275	-14.71	0.0000
	SA	0.1243	0.3055	0.41	0.6841
HVTN-704	(Intercept)	-3.2111	0.1832	-17.53	0.0000
	USAS	-1.0440	0.3340	-3.13	0.0018

CHAPTER 5: SUPPLEMENTAL RESULTS FOR CHAPTER 3

5.1 Tables and Figures for Chapter 3

Table 5.1: Estimation results of regression parameters under the model configuration (1)

			$\beta = 0.5$				$\gamma = -\log(2)$			
n	pt	σ	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
800	0.25	0.1	0.017	0.350	0.335	0.927	-0.021	0.169	0.157	0.930
	0.75	0.1	-0.014	0.336	0.330	0.932	-0.022	0.161	0.154	0.936
	0.25	0.2	0.032	0.426	0.384	0.920	-0.022	0.169	0.157	0.933
	0.75	0.2	-0.014	0.387	0.354	0.924	-0.017	0.160	0.153	0.934
1200	0.25	0.1	-0.005	0.256	0.232	0.940	-0.002	0.126	0.128	0.954
	0.75	0.1	-0.015	0.265	0.250	0.933	-0.012	0.136	0.128	0.942
	0.25	0.2	-0.046	0.279	0.255	0.935	-0.024	0.133	0.128	0.950
	0.75	0.2	-0.025	0.269	0.262	0.933	-0.012	0.135	0.126	0.949

Table 5.2: Estimation results of the regression parameters under the model configuration (2)

			$\beta = 0.5$				$\gamma = -\log(2)$			
n	pt	σ	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
800	0.25	0.1	0.017	0.394	0.320	0.910	-0.021	0.169	0.157	0.931
	0.75	0.1	-0.015	0.372	0.330	0.902	-0.019	0.160	0.153	0.937
	0.25	0.2	0.051	0.446	0.384	0.903	-0.022	0.169	0.157	0.933
	0.75	0.2	-0.014	0.397	0.354	0.910	-0.017	0.160	0.153	0.934
1200	0.25	0.1	0.039	0.239	0.210	0.934	-0.010	0.101	0.100	0.941
	0.75	0.1	-0.027	0.237	0.220	0.954	-0.002	0.104	0.100	0.932
	0.25	0.2	0.052	0.255	0.250	0.946	-0.005	0.101	0.100	0.934
	0.75	0.2	0.029	0.281	0.260	0.947	-0.001	0.104	0.100	0.935

Table 5.3: Estimation results of the regression parameters under the model configuration (3)

			$\beta = 0.5$				$\gamma = -\log(2)$			
n	pt	σ	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
800	0.25	0.1	-0.049	0.288	0.250	0.920	-0.010	0.135	0.124	0.920
	0.75	0.1	-0.017	0.279	0.267	0.931	-0.005	0.132	0.122	0.931
	0.25	0.2	-0.053	0.314	0.260	0.920	-0.009	0.135	0.124	0.920
	0.75	0.2	-0.023	0.324	0.285	0.920	-0.005	0.131	0.123	0.930
1200	0.25	0.1	-0.021	0.232	0.220	0.932	0.000	0.103	0.101	0.943
	0.75	0.1	-0.008	0.229	0.221	0.940	-0.002	0.101	0.100	0.942
	0.25	0.2	-0.012	0.252	0.235	0.941	0.001	0.103	0.101	0.948
	0.75	0.2	-0.030	0.250	0.233	0.935	-0.005	0.102	0.100	0.943

Table 5.4: Estimation results of the regression parameters under the model configuration (4)

			$\beta = 0.5$				$\gamma = -\log(2)$			
n	pt	σ	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
800	0.25	0.1	-0.040	0.270	0.236	0.920	-0.011	0.134	0.122	0.920
	0.75	0.1	-0.014	0.270	0.254	0.930	-0.006	0.126	0.120	0.940
	0.25	0.2	-0.049	0.273	0.240	0.920	-0.010	0.134	0.123	0.920
	0.75	0.2	-0.016	0.283	0.257	0.925	-0.005	0.126	0.120	0.940
1200	0.25	0.1	-0.030	0.221	0.213	0.930	-0.011	0.113	0.114	0.940
	0.75	0.1	-0.018	0.212	0.210	0.940	-0.001	0.098	0.102	0.948
	0.25	0.2	-0.023	0.233	0.225	0.932	-0.011	0.113	0.114	0.940
	0.75	0.2	-0.021	0.213	0.205	0.943	-0.005	0.102	0.106	0.940

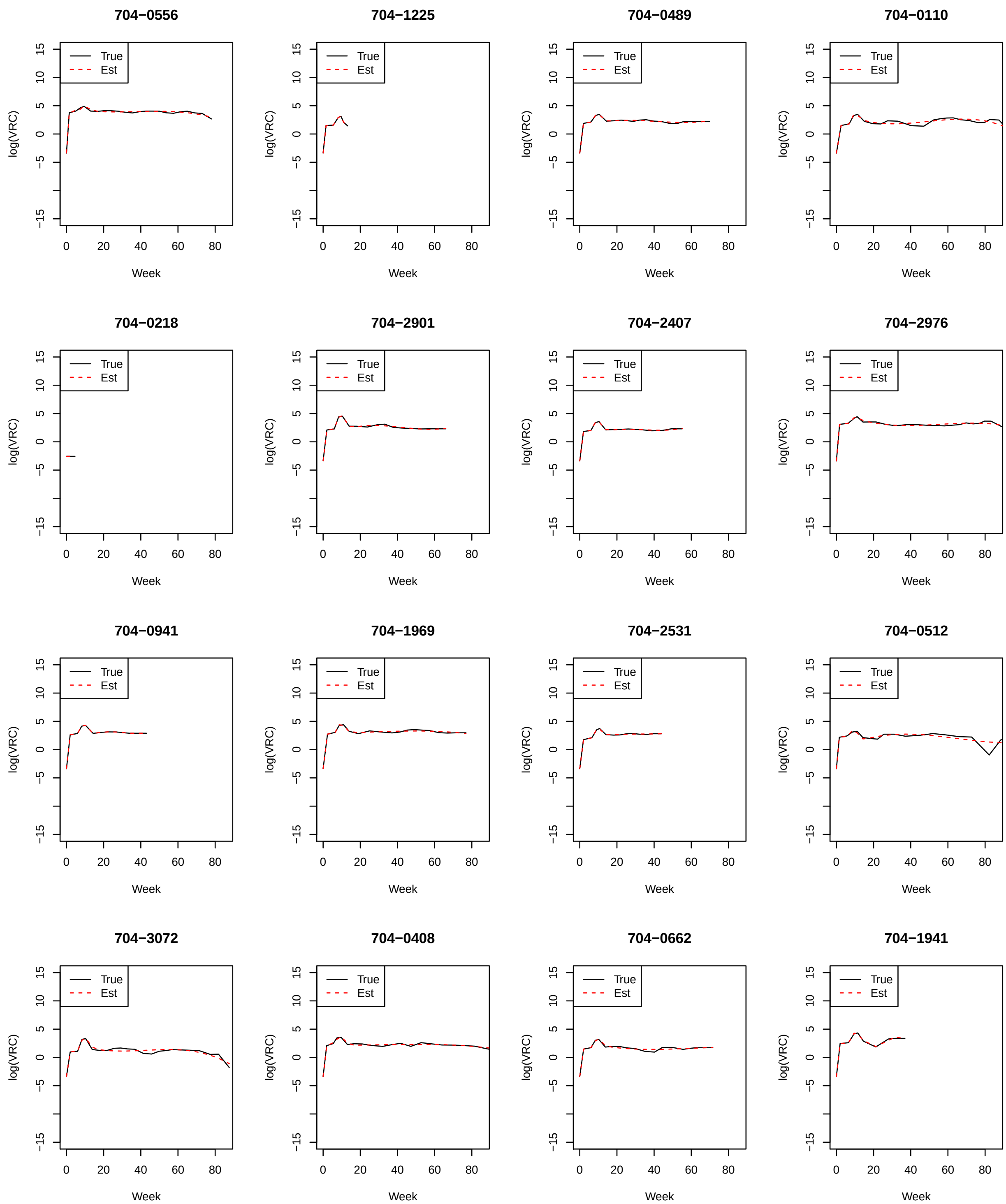

Figure 5.1: True log(VRC) curve versus least squared estimated log(VRC) curve for the participants in United State and Switzerland

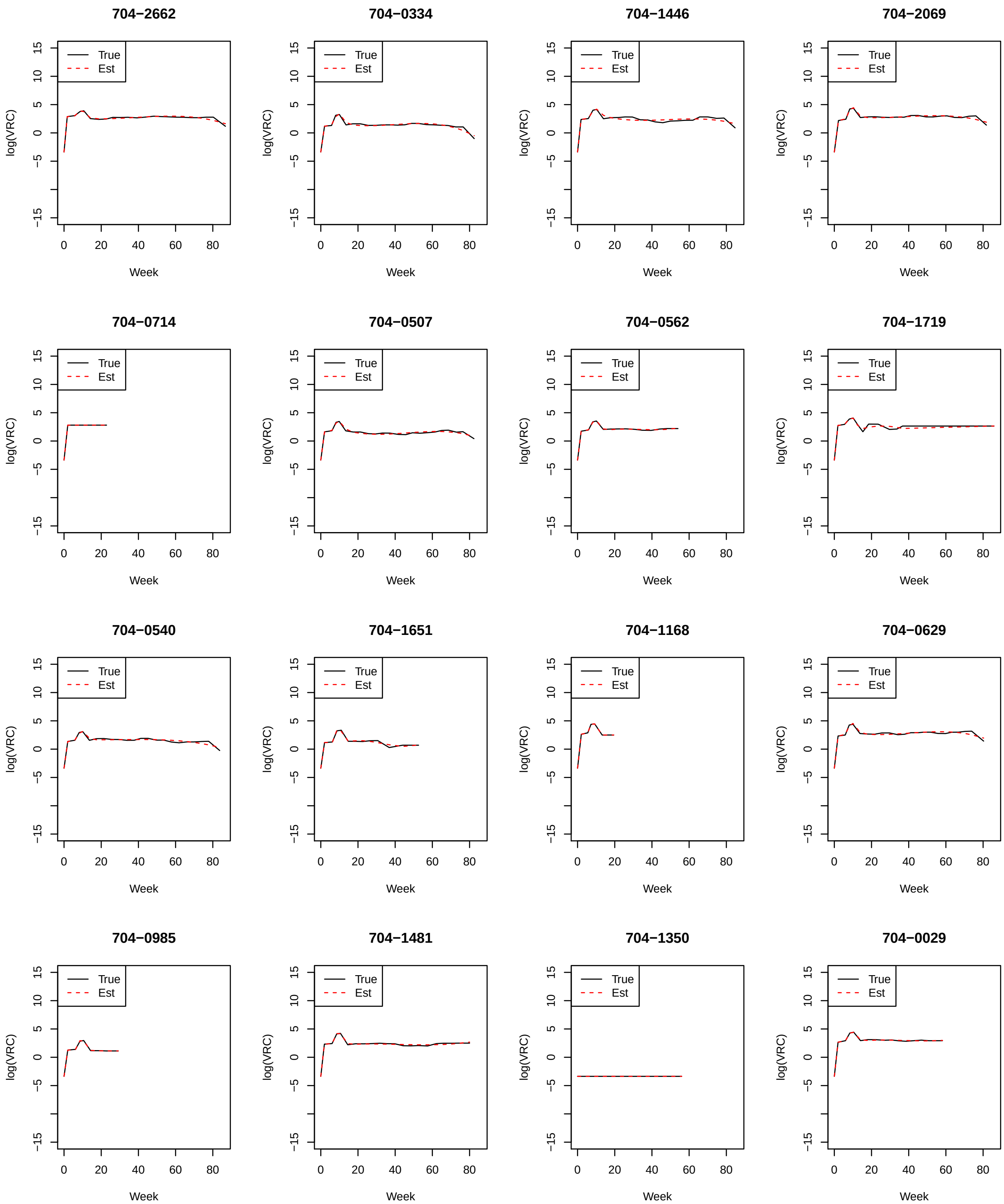

Figure 5.2: True log(VRC) curve versus least squared estimated log(VRC) curve for the participants in Brazil and Peru

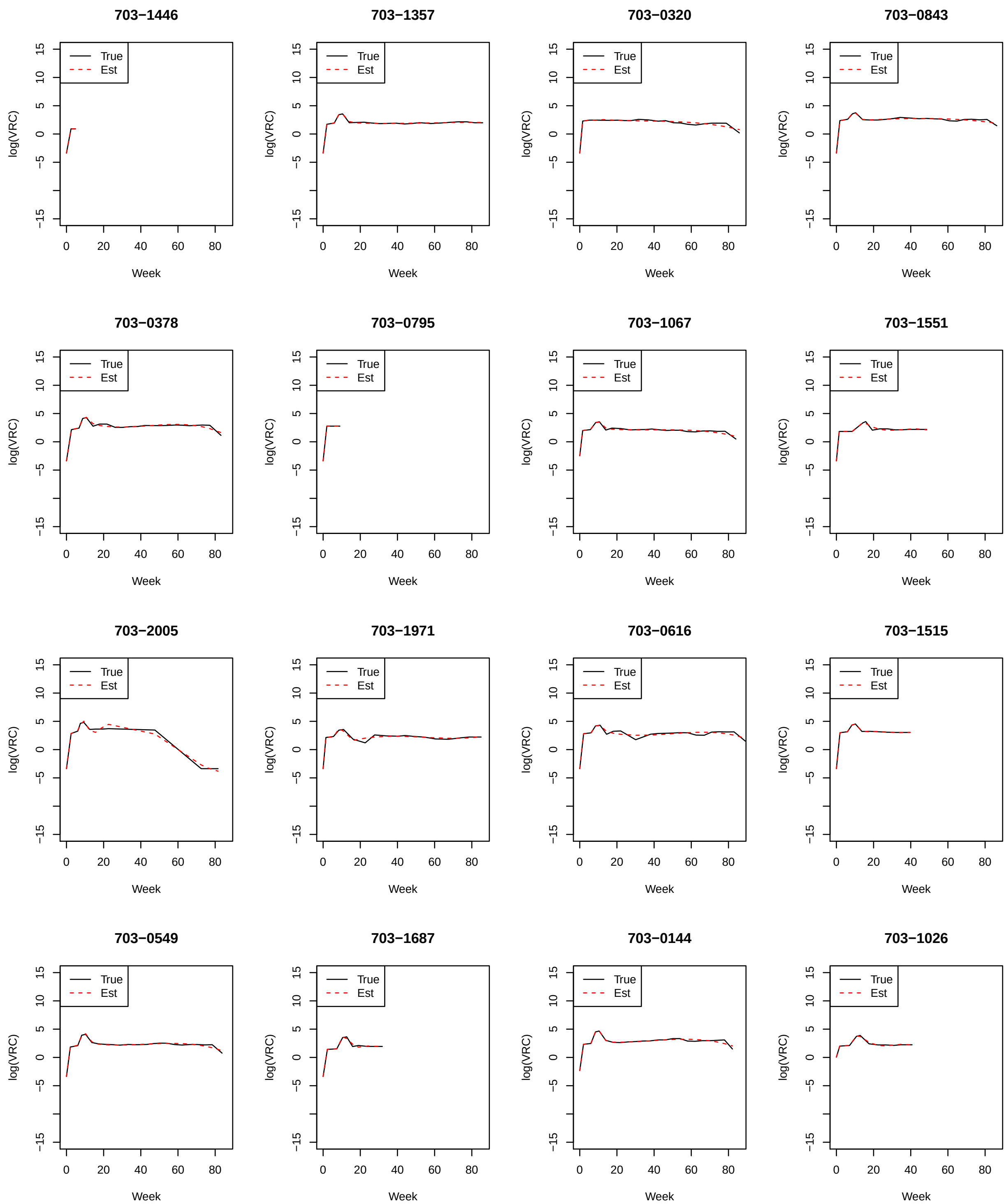

Figure 5.3: True log(VRC) curve versus least squared estimated log(VRC) curve for the participants in Sub-Saharan Countries

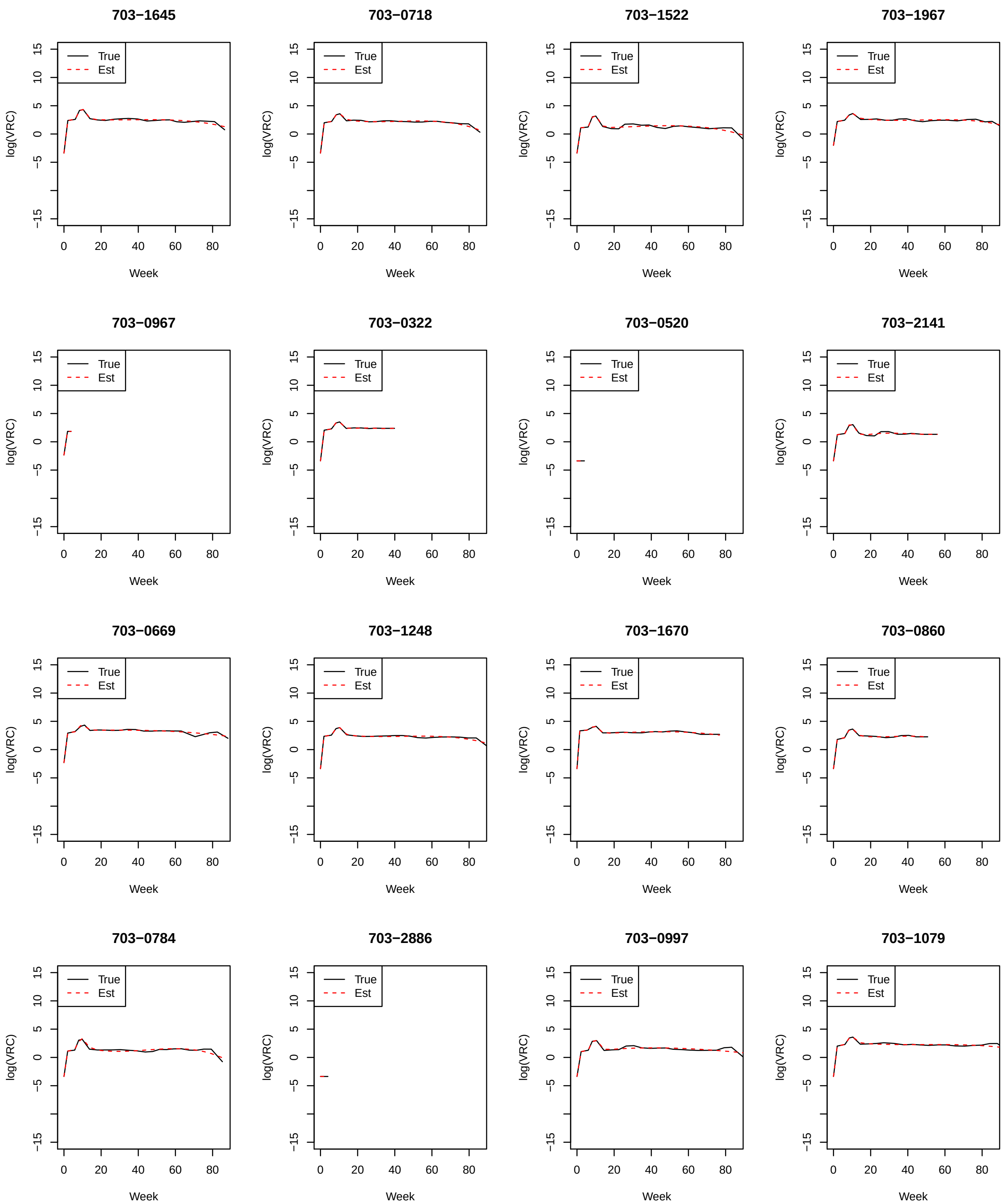

Figure 5.4: True log(VRC) curve versus least squared estimated log(VRC) curve for the participants in other Sub-Saharan Countries

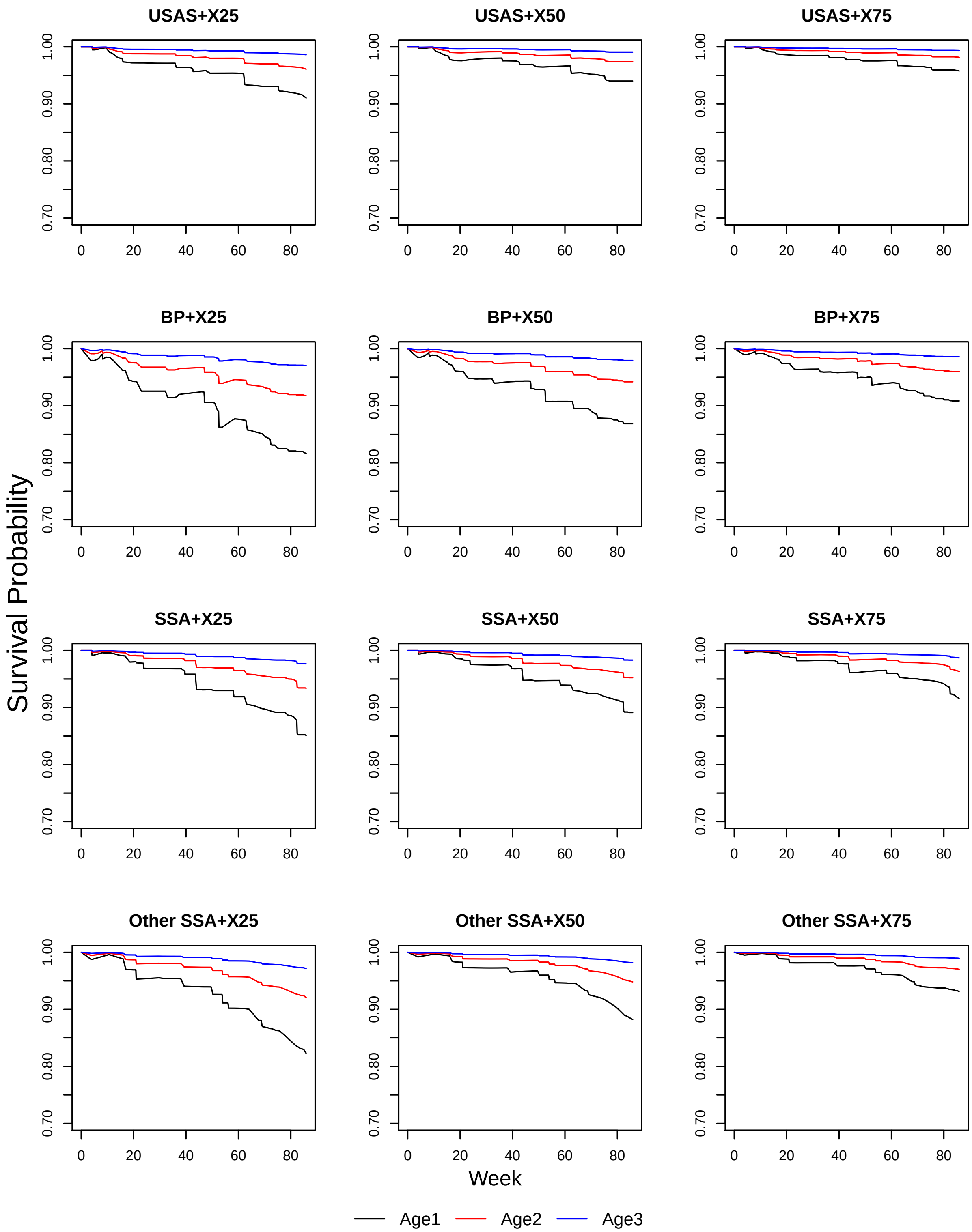

Figure 5.5: Estimated survival curves by considering different combinations of covariates under models (3.21) and (3.23) using the proposed method. Here, 'Age1', 'Age2' and 'Age3' stand for the age goups, < 20, $[20, 30]$, > 30, respectively and X25, X50, X75 represent logVRC at its 25th, 50th, 75th percentiles, respectively. The title of each figure, for example, 'USAS+X25' corresponds to the region USAS with X being the 25th percentile of estimated log(VRC).

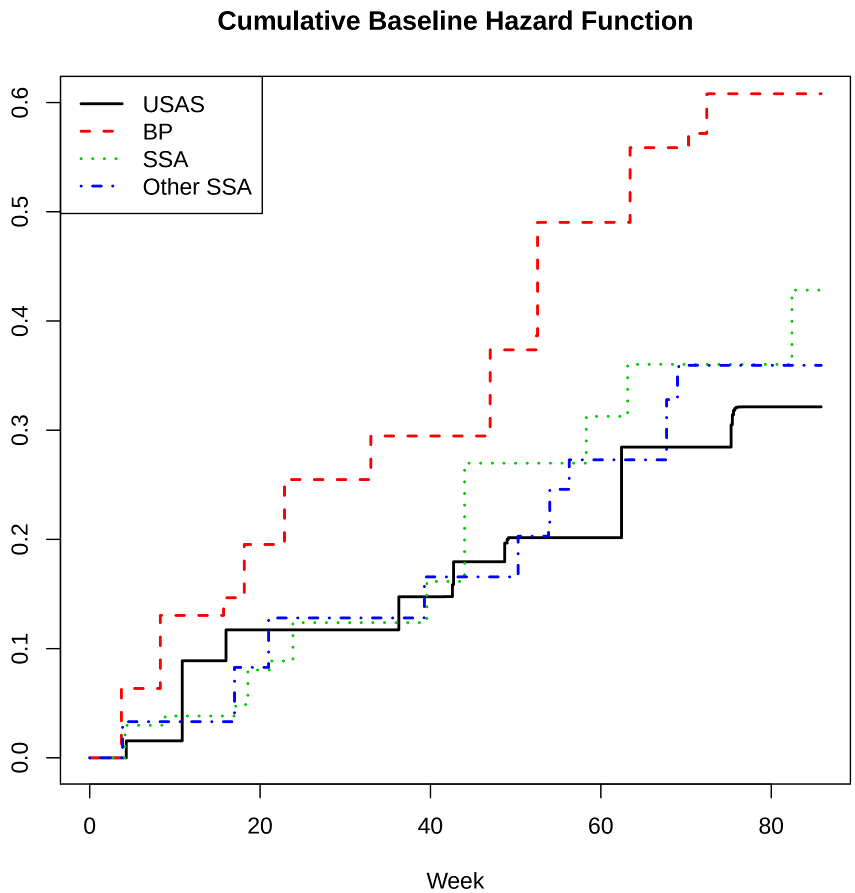

Figure 5.6: Estimated baseline cumulative hazard function, $\hat{\Lambda}_j(t)$, $j = 1, \cdots, 4$, for regions USAS, BP, SSA, other SSA under models (3.21) and (3.23) using the proposed method.

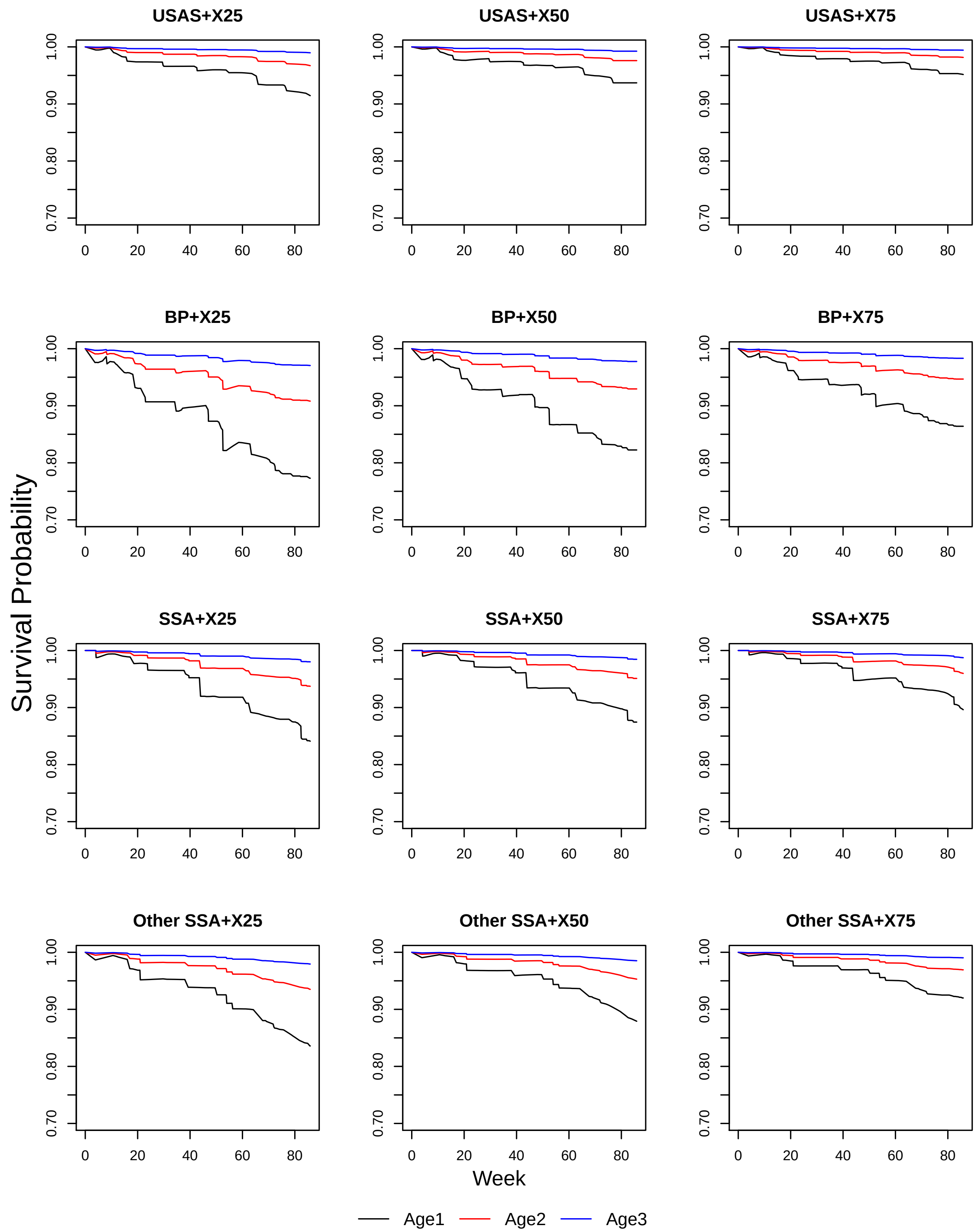

Figure 5.7: Estimated survival curves under the proportional hazards model (3.21), $\Lambda_j(t|X(\cdot), Z) = \int_0^t \exp\{\beta X(s) + \gamma^\top Z\} d\Lambda_j(s)$ without accounting for measurement error, using the naive plug-in method where $X(\cdot) = \log \text{VRC}$ is replaced by $\hat{X}(\cdot)$ and Z is the age groups, where $\hat{X}(\cdot)$ is the estimated $X(\cdot)$ based on model (3.23). Here, 'Age1', 'Age2' and 'Age3' stand for the age goups, < 20, $[20, 30]$, > 30, respectively and $X25$, $X50$, $X75$ represent estimated logVRC at its 25th, 50th, 75th percentiles, respectively. The title of each figure, for example, 'USAS+X25' corresponds to the region USAS with X being the 25th percentile of estimated log(VRC).

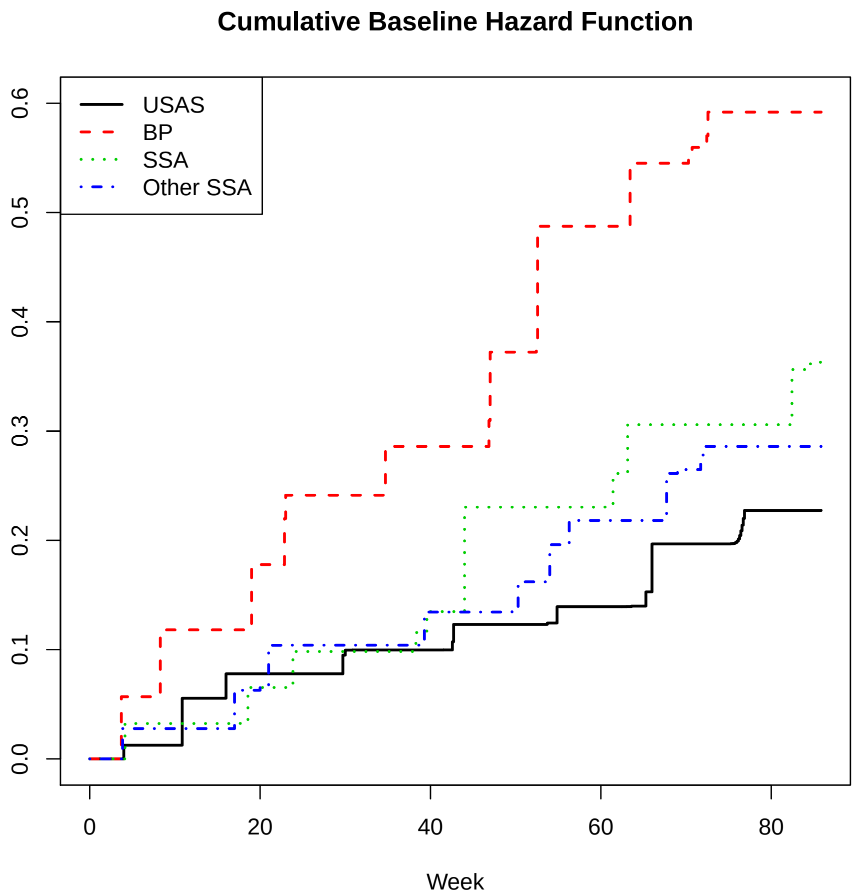

Figure 5.8: Estimated baseline cumulative hazard function, $\hat{\Lambda}_j(t)$, $j = 1, \cdots, 4$, for regions USAS, BP, SSA, other SSA under the proportional hazards model (3.21), $\Lambda_j(t|X(\cdot), Z) = \int_0^t \exp\{\beta X(s) + \gamma^\top Z\} d\Lambda_j(s)$ without accounting for measurement error, using the naive plug-in method where $X(\cdot) = \log \text{VRC}$ is replaced by $\hat{X}(\cdot)$ and Z is the age groups, where $\hat{X}(\cdot)$ is the estimated $X(\cdot)$ based on model (3.23).

References

Breslow, N. E. and Wellner, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scandinavian Journal of Statistics*, 34(1):86–102.

Breslow, N. E. and Wellner, J. A. (2008). A Z-theorem with estimated nuisance parameters and correction note for 'weighted likelihood for semiparametric models and two-phase stratified samples, with application to cox regression'. *Scandinavian Journal of Statistics*, 35(1):186–192.

Cai, T. and Zheng, Y. (2013). Resampling procedures for making inference under nested case-control studies. *Journal of the American Statistical Association*, 108(504):1532–1544. PMID: 24436503.

Chen, K. (2001). Generalized case-cohort sampling. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63(4):791–809.

Corey, L., Gilbert, P. B., Juraska, M., Montefiori, D. C., Morris, L., Karuna, S. T., Edupuganti, S., Mgodi, N. M., deCamp, A. C., Rudnicki, E., Huang, Y., Gonzales, P., Cabello, R., Orrell, C., Lama, J. R., Laher, F., Lazarus, E. M., Sanchez, J., Frank, I., Hinojosa, J., Sobieszczyk, M. E., Marshall, K. E., Mukwekwerere, P. G., Makhema, J., Baden, L. R., Mullins, J. I., Williamson, C., Hural, J., McElrath, M. J., Bentley, C., Takuva, S., Gomez Lorenzo, M. M., Burns, D. N., Espy, N., Randhawa, A. K., Kochar, N., Piwowar-Manning, E., Donnell, D. J., Sista, N., Andrew, P., Kublin, J. G., Gray, G., Ledgerwood, J. E., Mascola, J. R., and Cohen, M. S. (2021). Two randomized trials of neutralizing antibodies to prevent HIV-1 acquisition. *New England Journal of Medicine*, 384(11):1003–1014. PMID: 33730454.

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

Dudley, R. M. (1992). Frechet Differentiability, p-Variation and Uniform Donsker Classes. *The Annals of Probability*, 20(4):1968–1982.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, 42(4):845–854.

Gao, F., Zeng, D., Couper, D., and Lin, D. Y. (2019). Semiparametric regression analysis of multiple right- and interval-censored events. *Journal of the American Statistical Association*, 114(527):1232–1240. PMID: 31588157.

Gao, F., Zeng, D., and Lin, D.-Y. (2017). Semiparametric estimation of the accelerated failure time model with partly interval-censored data. *Biometrics*, 73(4):1161–1168.

Hammer, S., Katzenstein, D., Hughes, M., Gundacker, H., Schooley, R., Haubrich, R., Henry, W., Lederman, M., Phair, J., Niu, M., Hirsch, M., and Merigan, T. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *The New England Journal of Medicine*, 335(15):1081—1090.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

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

Huang, J. and Wellner, J. A. (1997). Interval censored survival data: A review of recent progress.

Kim, J. S. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(2):489–502.

Kosorok, M. (2007). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer New York.

Mandal, S., Wang, S., and Sinha, S. (2019). Analysis of linear transformation models with covariate measurement error and interval censoring. *Statistics in Medicine*, 38(23):4642–4655.

Payne, R., Yang, M., Zheng, Y., Jensen, M. K., and Cai, T. (2016). Robust risk prediction with biomarkers under two-phase stratified cohort design. *Biometrics*, 72(4):1037–1045.

Qi, L., Wang, C. Y., and Prentice, R. L. (2005). Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association*, 100(472):1250–1263.

Royden, H. L. (2018 - 2010). *Real analysis / H.L. Royden, Stanford University, P.M. Fitzpatrick, University of Maryland, College Park.* Pearson modern classic. Pearson, New York, NY, fourth edition [2018 reissue]. edition.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Saegusa, T. and Wellner, J. A. (2013). Weighted likelihood estimation under two-phase sampling. *The Annals of Statistics*, 41(1):269–295.

Schick, A. and Yu, Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scandinavian Journal of Statistics*, 27(1):45–55.

Schumaker, L. L. (1981). *Spline Functions: Basic Theory.* John Wiley & Sons, New York.

Song, X. and Ma, S. (2008). Multiple augmentation for interval-censored data with measurement error. *Statistics in Medicine*, 27(16):3178–3190.

Sun, Y., Zhou, Q., and Gilbert, P. B. (2023). Analysis of the Cox model with longitudinal covariates with measurement errors and partly interval censored failure times, with application to an AIDS clinical trial. *Statistics in Biosciences*, 15(2):430–454.

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

van der Vaart, A. (2000). *Asymptotic Statistics.* Asymptotic Statistics. Cambridge University Press.

van der Vaart, A. and Wellner, J. A. (1996). Weak convergence and empirical processes: With applications to statistics.

Wen, C.-C. and Chen, Y.-H. (2014). Functional inference for interval-censored data in proportional odds model with covariate measurement error. *Statistica Sinica*, 24(3):1301–1317.

Zeng, D., Gao, F., and Lin, D. Y. (2017). Maximum likelihood estimation for semiparametric regression models with multivariate interval-censored data. *Biometrika*, 104(3):505–525.

Zeng, D. and Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 93(3):627–640.

Zeng, D., Mao, L., and Lin, D. (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika*, 103(2):253–271.

Zhang, Y., Hua, L., and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the cox model with interval-censored data. *Scandinavian Journal of Statistics*, 37(2):338–354.

Zhou, Q., Sun, Y., and Gilbert, P. B. (2021). Semiparametric regression analysis of partly interval-censored failure time data with application to an aids clinical trial. *Statistics in Medicine*, 40(20):4376–4394.