ANALYSIS OF SEMIPARAMETRIC REGRESSION MODELS FOR THE CUMULATIVE INCIDENCE FUNCTIONS UNDER THE TWO-PHASE SAMPLING DESIGNS

by

Unkyung Lee

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

2016

Approved by:

____

Dr. Yanqing Sun

____

Dr. Jiancheng Jiang

____

Dr. Yang Li

____

Dr. Wei Zhao

ABSTRACT

UNKYUNG LEE. Analysis of semiparametric regression models for the cumulative incidence functions under the two-phase sampling designs. (Under the direction of DR. YANQING SUN)

Competing risks often arise where a subject may be exposed to two or more mutually exclusive causes of failure. The cumulative incidence function is a proper summary quantity for analyzing competing risks data. In epidemiologic cohort studies, case-cohort study designs have been widely used to evaluate the effects of covariates on failure times when the occurrence of the failure event is rare. Under the case-cohort design, the covariate histories are investigated only for the subjects who experience the event of interest (cases) during the follow-up period and for a relatively small random sample (the subcohort) from the original cohort (Prentice, 1986). In this dissertation, we study estimating procedures for the cumulative incidence function based on competing risks data under case-cohort/two-phase sampling designs.

First, we introduce a semiparametric model for the cumulative incidence function under two-phase sampling data. The estimation procedure is based on the direct binomial regression model (Scheike et al., 2008), which enables us to evaluate the effects of the covariates directly when there exists competing risks. We develop an estimating equation for the missing model by using the inverse probability weighting of the complete cases. However, this method loses the efficiency because it still uses only the complete data of subjects.

Second, we propose an estimating equation by using augmented inverse probability of complete cases for the semiparametric model using an identity link function. The

AIPW method is doubly robust and it can improve efficiency.

The asymptotic properties of the proposed IPW and AIPW estimators have been established. The simulation studies show that these two methods have satisfactory finite-sample performances. We use the auxiliary variables that may improve efficiency through their correlation with the phase-two covariates in the simulation studies for the AIPW method. The proposed IPW and AIPW estimating methods are applied to analyze data from the RV144 vaccine efficacy trial, respectively.

ACKNOWLEDGMENTS

I would like to give my deepest gratitude to my dissertation adviser, Dr. Yanqing Sun. I would not have finished my PhD program without her guidance, insights, patience and encouragement. The lessons and experience I have learned from her direction are invaluable in my life.

I also would like to thank the committee members, Drs. Jianchen Jiang, Yang Li, and Wei Zhao for their constructive comments and valuable suggestions. I would like to specially thank Dr. Yang Li who was willing to have a discussion with me on the material at the beginning of the dissertation process. I also would like to thank Dr. Thomas H. Scheike, University of Copenhagen, Denmark, for sharing his valuable experience and guidance for the programming and to thank Dr. Peter Gilbert at Fred Hutchinson Cancer Research Center for guidance on the data application. This research was partially supported by NIAID NIH award number R37AI054165, and National Science Foundation grant DMS-1208978 and DMS-1513072.

My sincere thanks goes out to the wonderful professors who taught me and the staff in the department of Applied Mathematics and Statistics. I would like to express my sincere gratitude to Dr. Mohammad Kazaemi and Dr. Hae-Soo Oh for encouragement. My thanks also goes out to Mr. Mark Hamrick who helped me with the technical issues of computing.

Finally, I also would like to express my appreciation to my friends, family and husband. I would like to thank my dear friends who shared lots of memories with me. I would like to express my deepest gratitude to my parents who offered endless support. My special thanks are reserved for my husband, Bennett Giles Clayton for staying with me whenever I needed him the most.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

This chapter aims to review previous work and introduce our missing data for the cumulative incidence function. In section 1.1, we review the basic background for competing risks data, how to summarize the competing risks probabilities and estimate those probabilities. In section 1.2, we review literature on regression models to estimate covariate effects for the cumulative incidence function, focusing on direct regression models such as Fine and Gray model (Fine and Gray, 1999) and direct binomial model (Scheike et al., 2008). We also review case-cohort study design (Prentice, 1986), which is a form of two-phase sampling (Kulich and Lin, 2004). In section 1.3, we introduce a semiparametric model proposed by Scheike and others (Scheike et al., 2008) by allowing some covariates to be missing.

1.1 Competing Risks Data

The competing risks models are concerned with the situation where each individual may be exposed to two or more mutually exclusive causes of failure. These causes may compete with each other, but the eventual failure occurs due to exactly one of these. Each of these causes is called a competing risk.

Competing risks data have arisen in many research areas such as biomedical, public health, actuarial science, social science and engineering. In cancer study, death due to cancer may be of interest and deaths due to other causes such as surgical mortality and old age are competing risks (Putter et al., 2007). The study of failures of engines fitted to heavy duty vehicles (Hinds, 1996) is associated with five different competing risks. A competing risks model is also used in modeling the unemployment time (Flinn and Heckman, 1983), where failure time is the waiting time till the end of unemployment and reasons for leaving unemployment were considered as competing risks.

In the presence of competing risks data, a problem arises when the occurrence of the event of interest is precluded by the occurrence of another event.

1.1.1 Modeling Competing Risks Data

Let T_k be the kth latent failure time with $k = 1, 2, \ldots, K$ and let Z be a possibly time dependent covariate vector. Let $T = \min_{1 \le k \le K}\{T_k\}$ be the first observed failure time with the cause of failure ϵ.

In early approaches, it is well known that there are identifiable problems in modeling competing risks data in terms of latent failure time (Tsiatis, 1975). The problems arise because we only observe the earliest of those latent failures T with failure type $\epsilon = k$.

Let $S(t_1, \ldots, t_k) = P(T_1 > t_1, \ldots, T_K > t_k)$ be the joint survival distribution of the time to the k different events and let $S_k(t) = P(T_k > t) = S(0, \ldots, 0, t, 0, \ldots, 0)$ be the marginal distribution. Tsiatis (1975) has proved that neither the joint survival distributions nor the marginal distributions of the latent failure times are identifiable from the observed data if the competing risks are dependent. The observed data (T, ϵ) cannot provide enough information to tell which one is the true underlying distribution among two different marginal survival functions since they reproduce the same cause specific subdistribution function. Moreover, it is not possible to test whether the assumption of independence of the marginal failure time distributions is valid.

Alternatively, the cause-specific hazard and the cumulative hazard function have been proposed to summarize the competing risks probabilities. These two functions completely specify the joint distribution of the failure time T and the failure cause ϵ (Lawless, 2003) and both are directly estimable from the competing risks data (Prentice et al., 1978). None of them makes any assumptions about the relationship between the competing risks such as independence.

Under competing risks data, the cause-specific hazard function $\lambda_k(t)$ for cause k is defined as

$$\lambda_k(t|Z) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, \epsilon = k | T \geq t, Z)}{\Delta t}, \quad k = 1, ..., K, \tag{1.1}$$

which represents the instantaneous failure rate from cause k at time t in the presence of all causes of K failure types, given covariates Z. The cumulative cause-specific hazard is defined by

$$\Lambda_k(t|Z) \;\; = \;\; \int_0^t \lambda_k(s|Z) ds. \tag{1.2}$$

Failure types are assumed to be distinct, one of $\{1, \ldots, K\}$. The overall hazard function conditional on a vector of covariates Z can be defined in terms of cause-specific hazard functions as

$$\lambda(t|Z) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t | T \ge t, Z)}{\Delta t} = \sum_{k=1}^{K} \lambda_k(t|Z).$$

An equivalent identifiable quantity is the cumulative incidence function of cause k given covariates Z is defined as

$$F_k(t|z) = P(T \le t, \epsilon = k|Z), \tag{1.3}$$

and it represents the probability of the subject failing from the cause k before given time t in the presence of all the competing risks. The total cumulative incidence function is

$$F(t|Z) = P(T \le t|Z) = \sum_{k=1}^{K} F_k(t|Z).$$

The overall survival function $S(t|Z)$ can be expressed in terms of the cause specific hazard function

$$S(t|Z) = P(T > t|Z) = \exp(-\int_0^t \sum_{k=1}^{K} \lambda_k(s|Z) ds) = \exp(-\sum_{k=1}^{K} \Lambda_k(s|Z)), \tag{1.4}$$

which is interpreted as the probability of not having failed from any cause at time t.

From the definition of cause-specific hazard function (1.1),

$$\lambda_k(t|Z)\Delta t \approx P(t \le T < t + \Delta t, \epsilon = k | T \ge t, Z).$$

This implies that

$$\frac{P(t \le T < t + \Delta t, \epsilon = k|Z)}{\Delta t} \approx P(T \ge t)\lambda_k(t|Z) \tag{1.5}$$

for an infinitesimal Δt. The left-hand side of (1.5) is approximately the probability density function for cause k when Δt goes to 0. Therefore, by integrating both side of (1.5), the cumulative probability of cause k in (1.3) can be expressed in terms of the cause specific hazards and the overall survival function as

$$F_k(t|Z) = \int_0^t \lambda_k(s|Z)S(s-)ds. \tag{1.6}$$

The cumulative incidence function is also called subdistribution function (Pintilie, 2007) because it is not a proper distribution function. More precisely, the cumulative probability of failing from cause k remains less than one, as $\lim_{t\to\infty} F_k(t) = P(\epsilon = k) < 1$. Other alternative names for this function are the cause-specific failure probability (Gaynor et al., 1993), the crude incidence curve (Korn and Dorey, 1992) and the absolute cause-specific risk (Benichou and Gail, 1990).

The standard survival analysis methods such as Kaplan-Meier estimator and Log-rank test have been often misused to analyze competing risks in medical literature. For example, the standard Kaplan-Meier estimator for the jth failure estimates $S_j(t|Z) = \exp(-\Lambda_j(t|Z))$. However, it cannot be interpreted as a marginal survival function of the jth failure time. This only makes sense when analyzing data with a single event of occurrence even in the case of independent censoring (Lawless, 2003). Furthermore, the complement of the standard Kaplan-Meier estimator, which is the probability of a subject failing from cause j before or at time t, is greater than or equal to the cumulative incidence function,

$$\begin{aligned} 1 - S_j(t|Z) &= \int_0^t \lambda_j(s|Z)\exp(-\Lambda_j(s))ds \\ &\geq \int_0^t \lambda_j(s|Z)\exp(-\sum_{k=1}^{K}\Lambda_k(s))ds = F_j(t|Z) \end{aligned} \tag{1.7}$$

with equality at time t if there is no competition, $\sum_{k=1,k\neq j}^{K} \Lambda_k(s) = 0$. This shows the bias of the standard Kaplan-Meier estimator if it is used to estimate $F_j(t|Z)$ (Putter et al., 2007).

These problems come from the violation of assumption on Kaplan-Meier estimator that the time to the competing events are independent to the time to the event of interest. In this situation, the standard Kaplan-Meier estimator overestimates the probability of failure.

1.1.2 Estimation for Cumulative Incidence Function

The cumulative incidence function $F_k(t)$ for cause k can be estimated using equation (1.6) in the presence of competing risk data. Let $0 < t_1 < t_2 < \cdots < t_N$ be the ordered distinct time points at which failures of any cause occur. Let d_{ki} be the number of failures from type k at time t_i, $i = 1 \dots N$. It is allowed for a different subject to fail from the same cause k at the same time t_i. Let n_i be the risk set at time t_i, which is the number of patients who were not censored and have not failed yet from any cause up to time t_i. The discretized version of the cause-specific hazard function (1.1) is

$$\lambda_k(t_i|Z) = P(T = t_i, \epsilon = k|T > t_{i-1}) \tag{1.8}$$

and it can be estimated by

$$\hat{\lambda}_k(t|Z) = \frac{d_{ki}}{n_i}, \tag{1.9}$$

which is the proportion of subjects at risk who fail from cause k. By using the standard Kaplan-Meier estimator, the overall survival probability at time t_i including all types of events defined in (1.4) can be estimated by

$$\begin{aligned} \hat{S}(t_i) &= \hat{S}(t_{i-1})\hat{S}(t_i|t_{i-1}) = \hat{S}(t_{i-1})(1-\hat{\lambda}(t_i)) \\ &= \prod_{i:t_i\le t}\left(1-\sum_{k=1}^{K}\hat{\lambda}_k(t_i)\right) \\ &= \prod_{i:t_i\le t}(1-\frac{d_i}{n_i}), \end{aligned} \tag{1.10}$$

where $d_i = \sum_{k=1}^{K} d_{ki}$ denotes the total number of failures from any cause at t_i. Thus, by using (1.9) and (1.10), the cumulative incidence function of cause k (1.3) can be estimated by

$$\widehat{F}_k(t|Z) = \sum_{i:t_i\le t}\hat{\lambda}_k(t_i|Z)\hat{S}(t_{i-1}). \tag{1.11}$$

In the presence of competing risks, the interrelation among failure types is one of the distinct problems in the analysis of failure times. The problem of testing the equality of two cause-specific hazard rates has been studied by Bagai and Kochar (89 a,b), Yip and Lam (1992), Neuhaus (1991), Sen (1979), Aras and Deshpandé (1992), Aly et al. (1994), Sun and Tiwari (1995), Sun (2001) and Gilbert et al. (2004) among others. Alternatively, Gray (1988), Benichou and Gail (1990), Fine and Gray (1999), and McKeague et al. (2001) and Scheike et al. (2008) considered the problems on the cumulative incidence function.

1.2 Literature Review

It is important to study the covariate effects on the cumulative incidence function of a particular failure since the cumulative incidence function is a proper summary statistic for analyzing competing risks data (Zhang et al., 2008).

In the past, hazard-based regression models have been studied by many authors including Prentice et al. (1978), Cheng et al. (1988), Shen and Cheng (1999), Lin and Ying (1994), and Scheike and Zhang (2002, 2003). They used to model all cause-specific hazard functions and then estimate the cumulative incidence function based on these cause-specific hazard functions. From these approaches, it is quite easy to obtain estimation of the cumulative incidence function if the cause specific hazards are correctly modeled. However, it is difficult to summarize the effect of covariates on the cumulative incidence function in the presence of competing risks data in a simple way.

In section 1.2.1, we review Cox proportional regression model as one of the most popular hazard-based regression models in the presence of competing risks. To overcome disadvantages of hazard-based competing risks models, the Fine and Gray (1999) model based on subdistribution hazard and the direct binomial regression model (Scheike et al., 2008) will be discussed in section 1.2.2.

1.2.1 Hazard Based Regression Model

Prentice et al. (1978) proposed the Cox regression model for the cause-specific hazard $\lambda_k(t|Z)$ with possibly time dependent covariate vector Z by

$$\lambda_k(t|Z) = \lambda_{k,0}(t) \exp(\beta_k^T Z), \tag{1.12}$$

where $\lambda_{k,0}(t)$ is an unspecified cause-specific baseline hazard rate and the vector β_k represents covariate effects on cause k. Since there is no structure on $\lambda_{k,0}(t)$, then there is no need to make any assumption on the distribution of the lifetimes of the baseline population. The covariate effects in (1.12) are proportional for the cause-specific hazards. The model (1.12) treats failures from the cause of interest k as events, and failures from causes other than k, as censored observations. Under independent censoring assumption, the cumulative incidence function for cause k given covariates Z in the presence of all competing risks is

$$F_k(t|Z) = \int_0^t \lambda_k(s|Z) \exp(-\sum_{k=1}^{K} \int_0^t \lambda_k(s|Z))ds. \tag{1.13}$$

The regression coefficients on the model (1.12) can be obtained by using standard likelihood methods.

However, the effects of the covariates on the cause-specific hazard rate cannot be translated directly to an effect on the cumulative incidence function (1.13) under the model (1.12). The reason is that the failures from competing events are ignored by treating them as censored observations on the analysis of competing risks data, but the cumulative incidence function (1.13) for cause k not only depends on the hazard of cause k, but also on the hazards of all other causes. For example, the important covariate effects for the cause-specific hazards will highly influence the cumulative incidence probability, but an effect on the cause specific hazard for a particular cause k may have an adverse effect on the overall survival. Therefore, the significant covariate for all cause-specific hazards may not affect the cumulative incidence probability for cause k. Also, it is possible that some covariates influence the cumulative incidence function, but it may not be significantly associated with any of the cause-specific hazards (Scheike and Zhang, 2008). Therefore, the effect of covariates through modeling the cause specific hazards can not be simply used to predict the cumulative incidence function.

1.2.2 Direct Modeling on Cumulative Incidence Function

To overcome these problems, as mentioned in section 1.2.1, some new regression approaches have been considered to evaluate the covariate effects on the cumulative incidence function directly.

Fine and Gray (1999) developed a regression model that directly links the regression coefficients with the cumulative incidence function by using subdistribution hazard introduced by Gray (1988). They consider a semiparametric regression model with transformation $g(u) = log\{-log(1-u)\}$ for the cumulative incidence function, which is

$$g\{F_k(t|Z)\} = h_{k,0}(t) + \beta_k^{\mathsf{T}} Z, \quad k = 1, ..., K, \tag{1.14}$$

where $h_{k,0}(\cdot)$ is a completely unspecified, invertible, and monotone increasing function and β_k is a $p \times 1$ parameter vector related to cause k. The transformation $g(u) = log\{-log(1-u)\}$ leads to a reasonable assumption that there is a constant difference between cumulative incidence functions for two individuals with covariate vectors Z_1 and Z_2. That is, $g\{F_k(t|Z_1)\} - g\{F_k(t|Z_2)\} = \{Z_1 - Z_2\}^{\mathsf{T}}\beta_k$ for all t.

To directly estimate the model (1.14), they used the subdistribution hazard (Gray, 1988) for cause k

$$\begin{aligned}
\lambda_k^*(t|z) &= \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t, \epsilon = k | T \ge t \cup \{T < t, \epsilon \ne k\})}{\Delta t} \\
&= \frac{dF_k(t|z)/dt}{1 - F_k(t|z)} \\
&= \frac{-dlog\{1 - F_k(t|z)\}}{dt}.
\end{aligned} \tag{1.15}$$

This subdistribution hazard function for cause k represents the probability of a subject to fail from cause k in a very small time interval Δt, given the subject experienced no event until time t or experienced an event other than k before time t. By the ordered distinct time points and definitions in subsection 1.1.2, it can be estimated at time t_i by

$$\hat{\lambda}_k^*(t_i|z) = \frac{d_{ki}}{n_i^*}, \tag{1.16}$$

where d_{ki} denotes the number of failures of cause k at time t_i and n_i^* is the modified risk set including all subjects who did not experience any event until time t_i and all subjects that failed before t_i from a cause other than k. Thus, the modified risk set has subjects having already experienced events other than cause k. Those subjects are always at future risks of the event of interest cause k.

By using a semiparametric proportional hazards model with time varying covariates $Z(t)$,

$$\lambda_k^*(t|Z) \quad = \quad \lambda_{k,0}^*(t|Z)\exp(Z^{\mathsf{T}}(t)\beta_k) \tag{1.17}$$

where $\lambda_{k,0}^*(t|z)$ is a completely unspecified, nonnegative function in time t and by applying $g(u) = log\{-log(1-u)\}$ transformation to (1.14) with $h_{k,0}(t) = log\{\int_0^t \lambda_{k,0}^*(t)\}$, they proposed the following cumulative incidence function for cause k

$$F_k(t|Z) \quad = \quad 1 - \exp\left[-\int_0^t \lambda_{k,0}^*(s)\exp\{\beta_k^T Z(s)\}ds\right], \tag{1.18}$$

which enables us to assess the effect of the covariates on the cumulative incidence function directly.

With complete data, the standard partial likelihood method with modified risk set can be applied (Fine and Gray, 1999). For right censored competing risks data, Fine and Gray (1999) utilized inverse probability of censoring weighted (ICPW) method (Robins and Rotnitzky, 1992) to construct an unbiased estimating function from the score function of the complete data partial likelihood (Fine and Gray, 1999). This approach can be analyzed by using the 'comprsk' package for R developed by Robert Gray.

However, the Fine and Gray model may not fit the data well even though it is easy to decide if covariates significantly affect the cumulative incidence function for a specific cause of failure (Scheike and Zhang, 2008). To remedy this problem, Scheike et al. (2008) have proposed a class of general models containing the Fine and Gray model.

Scheike et al. (2008) proposed a fully nonparametric model to evaluate covariate effects directly on the cumulative incidence function for cause k, that is

$$h\{F_k(t|X_i)\} = X_i^{\mathsf{T}}\eta(t),\ \ i = 1, ..., n \tag{1.19}$$

where $h(\cdot)$ is a known link function and $\eta(t)$ is the time varying effects of the covariate X_i. The $(p+1)-$ dimensional regression coefficient $\eta(t)$ is an unspecified function and $X_i = (1, X_{i1}, \ldots, X_{ip})^{\mathsf{T}}$ is a $(p+1)-$ dimensional covariate vector. The first component of X_i yields the time-dependent intercept. This model is very flexible since it allows covariates to have time varying effects. The model (1.19) contains the Aalen's generalized additive model by using log link function $h(x) = log(1-x)$. They proposed a direct binomial regression method for a regression analysis by using the inverse probability of censoring weighted response. They also proposed a class of general semiparametric models by

$$h\{F_j(t|X_i,Z_i)\} = \{X_i^{\mathsf{T}}\eta(t)\}g(\gamma,Z_i,t), \tag{1.20}$$

$$h\{F_j(t|X_i,Z_i)\} = X_i^{\mathsf{T}}\eta(t)+g(\gamma,Z_i,t) \tag{1.21}$$

where g is a known function, $\eta(t)$ is a $(p+1)-$ dimensional vector of time-dependent regression coefficients of $X_i = (1, X_{i1}, \ldots, X_{ip})^{\mathsf{T}}$ and γ is a $q-$dimensional vector of time-invariant coefficients of $Z_i = (Z_{i1}, \ldots, Z_{iq})^{\mathsf{T}}$.

With the log link function $h(x) = log(1-x)$ and $g(\gamma, Z_i, t) = \exp(\gamma^{\mathsf{T}} Z_i)$, the multiplicative model (1.20) reduces to a Cox-Aalen model which contains the Cox model and Aalen's additive model. When $x = 1$, the model (1.20) reduces to the Fine and Gray (1999) model. With the log link function $h(x) = log(1-x)$ and $g(\gamma, Z_i, t) = \gamma^{\mathsf{T}} Z_i t$, the model (1.21) generates a partially semiparametric additive model (McKeague and Sasieni, 1994). When $x = 1$, the model (1.21) reduces to the Lin and Ying (1994) special additive model.

Scheike and Zhang (2008) also proposed estimating equations to estimate $\eta(t)$ and γ simultaneously. They derived asymptotic results and studied the predicted cumulative incidence function for a given set of covariate values. Scheike and Zhang (2008) considered a new simple goodness-of-fit procedure for the proportional subdistribution hazards assumption.

One drawback for both direct binomial and subdistribution approaches is that one has to estimate the censoring distribution for each individual. Usually, the Kaplan-Meier estimator is utilized for the censoring distribution. This non-augmented inverse probability weighting technique was firstly proposed by Koul et al. (1981). By using the semiparametric efficiency theory in Bickel et al. (1993), Robins and Rotnitzky (1992) showed that regression modeling of the censoring distribution improves efficiency of the inverse probability weighting technique. This is because each censored observation carries information about the relationship between event time and covariates, even if the censoring is independent of the covariates.

We allow covariates to have missing values in the general semiparametric additive model (1.21) in subsection 1.3 and develop estimating equations to analyze the missing model in chapter 2 and chapter 3.

1.2.3 Case-Cohort/Two Phase Sampling

Epidemiologic cohort studies and disease prevention trials often necessitate the follow-up of several thousand subjects for a number of years and thus can be prohibitively expensive (Prentice, 1986). The assembly of covariate histories for all cohort members results in much cost and effort in such studies. Therefore, it is an important issue to reduce cost in those studies and achieve the same goals as a cohort study.

Among several study designs proposed to reduce cost, the case-cohort design has been widely used in those cohort studies to assess the effects of possibly time-dependent covariates on a failure time. The case-cohort design has been proposed by Prentice (1986). Under this design, the covariate histories are investigated only for the subjects who experience the event of interest during the follow-up period (the cases) and for a relatively small random sample (the subcohort) from the original cohort. This design is very useful where the occurrence of the failure event is rare in large cohort studies since it is unnecessary to investigate the covariates of event-free subjects (Sun et al., 2016). The case-cohort data is a biased sample and thus applying standard methods for randomly sampled data to the biased data may result in biased estimation (Sun et al., 2016).

The case-cohort design is also a form of two-phase sampling. At the first phase, the study cohort is randomly sampled from a general population. The phase one covariate data are observed on all of the subjects in the cohort. These covariates are treatment type, age and gender as examples. At the second phase, the subcohort is randomly selected from the study cohort. Covariate histories are completely assembled for the cases and the subcohort by collecting the second phase covariate including all of the expensive covariates which are not measured at the first phase. These covariates is called the phase two covariate data.

The Cox (1972, 1975) proportional hazard model has been widely used in analysis of case cohort data. Many authors have proposed statistical methods for case-cohort studies by modifying the full data partial likelihood score function for the Cox model, giving the inverses of true or estimated sampling probabilities to the score functions as weight functions, including Prentice (1986), Self and Prentice (1988), Kalbfleisch and Lawless (1988), Lin and Ying (1993), Barlow (1994), Chen and Lo (1999), Sørensen and P.K. (2000), Borgan et al. (2000), Chen (2001), Kulich and Lin (2004), and Samuelsen et al. (2007). The Cox model assumes that the hazard functions associated with different covariate values are proportional over time. This assumption may be too restrictive and the Cox model does not always fit data well in practice. Alternatively, the accelerated failure time model and the proportional odds model have been studied by Chen (2001) and Kong and Cai (2009), respectively.

The additive hazard model is another popular framework for the analysis of case-cohort data. This is because the risk differences between different treatment can be easily derived from the regression coefficients. It also gives a valuable public health interpretation. Kulich and Lin (2000) applied additive hazard models (Lin and Ying, 1993) to case-cohort study. Kang et al. (2013) recently proposed an estimation method for case-cohort data with the simple additive model of Lin and Ying (1994) that allows only constant covariate effects. Sun et al. (2016) proposed an estimation procedure for the semiparametric additive hazards model with case-cohort/two-phase sampling data, which allows the effects of some covariates to be time varying while specifying the effects of others to be constant. They used the inverse probability weighting of complete-case technique of Horvitz and Thompson (1952). With this approach, if a subject has a missing value for one covariate, then the observed values of other covariates together with the observed failure/censoring time of the same subject are not utilized. This leads to loss of efficiency. They also proposed an augmented estimating equation on the basis of the inverse probability weighting of complete cases by adapting the theory of Robins et al. (1994). By doing so, they showed the efficiency of the proposed estimators has been improved.

1.3 Model

Let T_i be the failure time and $\epsilon_i \in \{1, 2, \ldots, k\}$ denote the k different failure types for the ith subject. Assume that cause $\epsilon_i = 1$ is the primary cause of interest and $\epsilon_i \neq 1$ for other competing causes. Let $X_i = (1, X_{i1}, \ldots, X_{ip})^T$ and $Z_i = (Z_{i1}, \ldots, Z_{iq})^T$ be the $(p + 1)$- and q-dimensional possibly time-dependent covariate vectors, respectively. Let C_i be the right censoring time. Let $\Delta_i = I(T_i \leq C_i)$ be an indicator for uncensored failure time. The observed n competing risks data can be represented by $Y_i = (X_i, Z_i, \tilde{T}_i, \tilde{\epsilon}_i)$ for $i = 1, 2, \ldots, n$, where $\tilde{T}_i = \min(T_i, C_i)$ and $\tilde{\epsilon}_i = \epsilon_i \Delta_i$. The value $\tilde{\epsilon}_i = \epsilon_i$ indicates that the system failure time is observed at $\tilde{T}_i$ and the cause of failure is of type ϵ_i. Let $[a, \tau]$ be the time period from which data are collected. We assume that (T_i, ϵ_i) are independent of C_i given covariates (X_i, Z_i). In follow-up studies, the covariates of a subject i are only meaningful in the time interval when the subject is at-risk and still in the study, i.e., $t \leq \tilde{T}_i$.

Let $F_1(t; X_i, Z_i) = P(T_i \leq t, \epsilon_i = 1 | X_i, Z_i)$ be the conditional cumulative incidence function given covariates (X_i, Z_i) for each ith subject. We consider the following the additive semiparametric model for $F_1(t; X_i, Z_i)$:

$$F_1(t; X_i, Z_i) = h\{X_i^{\mathsf{T}} \eta(t) + g(\gamma, Z_i, t)\} \tag{1.22}$$

where $h(\cdot)$ is a known link function, $g(\cdot)$ is a known function of (γ, Z_i, t), $\eta(t)$ is a $(p+1)$-dimensional vector of time-dependent regression coefficients and γ is a q-dimensional vector of time-invariant coefficients. The first component of X_i yields the time-dependent intercept. Under model (1.22), the effects of the covariates X_i change with time while the effects of Z_i are time-invariant.

Suppose that X_i has two parts $(X_i^{(1)}, X_i^{(2)})$. The covariates $X_i^{(1)}$ and Z_i are observed for all the cohort members, but $X_i^{(2)}$ is only observed for a subset(subcohort, phase two sample) of the study subjects. Let ξ_i be the indicator of whether the subject i is selected into the phase-two sample. The subject i with $\xi_i = 1$ has fully observed X_i and Z_i. The subject i with $\xi_i = 0$ does not have the observed values for $X_i^{(2)}$. Let

$\mathcal{V}_i = \{\tilde{T}_i, \Delta_i, \tilde{\epsilon}_i, X_i^{(1)}, Z_i, A_i\}$ where A_i denotes possible auxilary variables that may be informative for selection of phase-two sample and / or phase-two covariates. We assume that the probability a subject is missing the phase-two covariates $X_i^{(2)}$ does not depend on the values of these covariates $P(\xi_i = 1|X_i^{(2)}, \mathcal{V}_i) = P(\xi_i = 1|\mathcal{V}_i)$. This assumption is the missing at random (MAR) assumption (Rubin, 1976). However, the selection probability may depend on any of the phase-one information $\mathcal{V}_i$.

Let $(\mathcal{V}_i, X_i^{(2)}, \xi_i)$, $i = 1, 2, \ldots, n$ be the competing risks data under two-phase sampling design. The observed case-cohort/two-phase sampling data are $(\mathcal{V}_i, \xi_i X_i^{(2)}, \xi_i)$. That is, $\{\tilde{T}_i, \Delta_i, \tilde{\epsilon}_i, X_i, Z_i, A_i\}$ are observed for a subject with $\xi_i = 1$, and $\{\tilde{T}_i, \Delta_i, \tilde{\epsilon}_i, X_i^{(1)}, Z_i, A_i\}$ are observed if $\xi_i = 0$. The selection probability, defined as the conditional probability that $X_i^{(2)}$ is observed, is $S_i = P(\xi_i = 1|\mathcal{V}_i)$. This selection probability S_i may depend on outcomes $\tilde{\epsilon}_i$ based on the competing causes of failure and the censoring indicator. Under the competing risks model, classical case-cohort design implies that $S_i = 1$ if $\tilde{\epsilon}_i = 1$ and $S_i = P(\xi_i = 1|\mathcal{V}_i, \tilde{\epsilon}_i \neq 1) < 1$ if $\tilde{\epsilon}_i \neq 1$. A subject is referred to as the case if the failure time is observed and the failure has cause 1; and the non-case, otherwise.

CHAPTER 2: ANALYSIS OF A SEMIPARAMETRIC CUMULATIVE INCIDENCE MODEL WITH MISSING COVARIATES USING INVERSE PROBABILITY WEIGHTED METHOD

In this chapter, we analysis the semiparametric model (1.22) by using inverse probability weighting of complete cases proposed by Horvitz and Thompson (1952). In section 2.1, the estimation procedure is studied based on direct binomial estimation (Scheike et al., 2008). We estimate the selection probability and censoring distribution. We also develop an estimating equation for the missing model (1.22). Asymptotic properties of the inverse probability weighting estimators will be discussed in section 2.2. Simulation studies are conducted under the subdistribution models to investigate the finite sample properties of the IPW method in section 2.3.

2.1 Estimation

2.1.1 Estimation Procedure

The selection probability S_i is unknown in practice, but it can be estimated based on a parametric model. Assume that $\varphi(\mathcal{V}_i, \theta)$ is the parametric model for the probability of complete-case $S_i = P(\xi_i = 1|\mathcal{V}_i)$, where θ is a finite dimensional vector of parameters. For example, one can assume that the logistic model is $\text{logit}(\varphi_1(\mathcal{V}_i, \theta_1)) = \theta_1^\mathsf{T}\mathcal{V}_i$ for those $\tilde{\epsilon}_i = 1$ and the different logistic model is $\text{logit}(\varphi_2(\mathcal{V}_i, \theta_2)) = \theta_2^\mathsf{T}\mathcal{V}_i$ for those $\tilde{\epsilon}_i \neq 1$. Let $\theta = (\theta_1^\mathsf{T}, \theta_2^\mathsf{T})^\mathsf{T}$. The parameter θ can be estimated by $\hat{\theta}$ as the maximizer of the observed data likelihood:

$$\begin{aligned} \mathrm{L}(\theta) &= \Pi_{i=1}^{n}\{\varphi_1(\mathcal{V}_i,\theta_1)\}^{\xi_i I(\tilde{\epsilon}_i=1)}\{1-\varphi_1(\mathcal{V}_i,\theta_1)\}^{(1-\xi_i)I(\tilde{\epsilon}_i=1)} \\ &\cdot\Pi_{i=1}^{n}\{\varphi_2(\mathcal{V}_i,\theta_2)\}^{\xi_i I(\tilde{\epsilon}_i\neq1)}\{1-\varphi_2(\mathcal{V}_i,\theta_2)\}^{(1-\xi_i)I(\tilde{\epsilon}_i\neq1)} \qquad (2.1) \end{aligned}$$

where $I(D)$ is the indicator function of the set D.

Let $N_i(t) = I(T_i \le t, \epsilon_i = 1)$ be the counting process associated with cause 1 and let $G(t|X_i, Z_i) = P(C_i > t|X_i, Z_i)$ be the conditional survival function of the censoring time for $0 \le t \le \tau$. Assume that there exists a positive number $0 < \delta \le 1$ such that $G(\tau|x, z) \ge \delta > 0$ for (x, z) in the range of (X_i, Z_i). Under conditional independence between C_i and (T_i, ϵ_i) given the covariates (X_i, Z_i), we have $E(\Delta_i|X_i, Z_i, T_i, \epsilon_i) = G(T_i|X_i, Z_i)$. It follows that

$$\begin{aligned} E\{\frac{\Delta_i N_i(t)}{G(T_i|X_i,Z_i)}|X_i,Z_i\} &= E[E\{\frac{\Delta_i N_i(t)}{G(T_i|X_i,Z_i)}|T_i,X_i,Z_i,\epsilon_i\}|X_i,Z_i] \\ &= E[\frac{1}{G(T_i|X_i,Z_i)}E\{\Delta_i N_i(t)|T_i,X_i,Z_i,\epsilon_i\}|X_i,Z_i] \\ &= E[N_i(t)|X_i,Z_i] \\ &= P(T_i \le t,\epsilon_i = 1|X_i,Z_i) = F_1(t|X_i,Z_i). \qquad (2.2) \end{aligned}$$

We consider the estimating equation based on the weighted response $\frac{\Delta_i N_i(t)}{G(T_i|X_i,Z_i)}$. In practice, the censoring distribution $G(t|x_i, z_i)$ is often unknown and can be estimated by the Kaplan-Meier estimator or by using a regression model for censoring times such as a Cox regression model or an additive Aalen regression model to improve efficiency. For simplicity, we use the Kaplan-Meier estimator $\widehat{G}(t)$ for $G(t)$.

Asymptotic results of the maximum likelihood estimator $\hat{\theta}$ and the estimator of censoring distribution $\widehat{G}(t)$ will be discussed in section 2.2.

2.1.2 Inverse Probability Weighted Complete-Case Estimation

Following Horvitz and Thompson (1952), the inverse probability weighting of the complete cases has been often used in missing data analysis.

Let $\boldsymbol{D}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma}) = \partial F_1(t;X_i,Z_i)/\partial\boldsymbol{\eta}(t)$ and $\boldsymbol{D}_{\boldsymbol{\gamma},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma}) = \partial F_1(t;X_i,Z_i)/\partial\boldsymbol{\gamma}$. Let $\psi_i(\theta) = \xi_i/\varphi(\mathcal{V}_i,\theta)$, where $\varphi(\mathcal{V}_i,\theta) = I(\tilde{\epsilon}_i = 1)\varphi_1(\mathcal{V}_i,\theta_1) + I(\tilde{\epsilon}_i \neq 1)\varphi_2(\mathcal{V}_i,\theta_2)$. By modifying the estimating equations of Scheike et al. (2008), the regression functions $\boldsymbol{\eta}(t)$ and parameters $\boldsymbol{\gamma}$ in model (1.22) can be estimated based on the following estimating functions:

$$\boldsymbol{U}_{\boldsymbol{\eta}}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\hat{\theta}) = \sum_{i=1}^{n}\psi_i(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})w_i(t)\left(\frac{\Delta_i\, N_i(t)}{\widehat{G}(T_i)} - F_1(t;\boldsymbol{X}_i,\boldsymbol{Z}_i)\right), \quad (2.3)$$

$$\boldsymbol{U}_{\boldsymbol{\gamma}}(\tau,\boldsymbol{\eta}(\cdot),\boldsymbol{\gamma},\hat{\theta}) = \sum_{i=1}^{n}\int_0^{\tau}\psi_i(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\gamma},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})w_i(t)\left(\frac{\Delta_i\, N_i(t)}{\widehat{G}(T_i)} - F_1(t;\boldsymbol{X}_i,\boldsymbol{Z}_i)\right)dt, \quad (2.4)$$

where $w_i(t)$ is a weight function. Under the MAR assumption, it is clear that $E\{\psi_i(\theta)|\mathcal{V}_i\} = 1$.

Let $\boldsymbol{W}(t) = \mathrm{diag}(w_i(t))$ and let $\boldsymbol{\Psi}(\theta) = \mathrm{diag}(\psi_i(\theta))$. Let $\boldsymbol{F}_1(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})$ be the $n\times 1$ vector of the model $F_1(t;\boldsymbol{X}_i,\boldsymbol{Z}_i)$ in (1.22) for $i = 1,\ldots,n$, denoting the ith element of the vector $\boldsymbol{F}_1(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})$ as $F_{1i}(t)$. Let $\boldsymbol{R}(t)$ be the $n\times 1$ vector of weighted responses $\Delta_i\, N_i(t)/\widehat{G}(T_i)$ and let $\boldsymbol{D}_{\boldsymbol{\eta}}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})$ and $\boldsymbol{D}_{\boldsymbol{\gamma}}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})$ be the $n\times(p+1)$ and $n\times q$ matrices with the ith rows equal to $\boldsymbol{D}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})$ and $\boldsymbol{D}_{\boldsymbol{\gamma},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})$, respectively.

The estimating equations given in (2.3) and (2.4) are equivalent to

$$\boldsymbol{U}_{\boldsymbol{\eta}}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\hat{\theta}) = (\boldsymbol{D}_{\boldsymbol{\eta}}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma}))^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\left\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})\right\}, \tag{2.5}$$

$$\boldsymbol{U}_{\boldsymbol{\gamma}}(\tau,\boldsymbol{\eta}(\cdot),\boldsymbol{\gamma},\hat{\theta}) = \int_0^{\tau}(\boldsymbol{D}_{\boldsymbol{\gamma}}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma}))^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\left\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})\right\}dt. \tag{2.6}$$

Let $l^{\infty}[0,\tau]$ be the set of uniformly bounded functions on $[0,\tau]$ and $\tilde{B} = (l^{\infty}[0,\tau])^{p+1}\times \mathbb{R}^q$. The estimating functions $\boldsymbol{U}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(\boldsymbol{\eta}(t),\boldsymbol{\gamma},\hat{\theta})=\{\boldsymbol{U}_{\boldsymbol{\eta}}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\hat{\theta}),\boldsymbol{U}_{\boldsymbol{\gamma}}(\tau,\boldsymbol{\eta}(\cdot),\boldsymbol{\gamma},\hat{\theta})\}$ are the mappings from $\tilde{B}$ to $\tilde{B}$. The inverse probability weighting of the complete-case estimators $\widehat{\boldsymbol{\eta}}(t)$ and $\widehat{\boldsymbol{\gamma}}$ of $\boldsymbol{\eta}(t)$ and $\boldsymbol{\gamma}$ solve the joint estimating equation $\boldsymbol{U}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(\boldsymbol{\eta}(t),\boldsymbol{\gamma},\hat{\theta})$, that is, $\boldsymbol{U}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(\widehat{\boldsymbol{\eta}}(t),\widehat{\boldsymbol{\gamma}},\hat{\theta}) = 0$.

By mimicking the procedure of Scheike et al. (2008), the estimating equations (2.5) and (2.6) can be solved by using an iterative algorithm. Consider the following Taylor expansion of $\boldsymbol{F}_1(t,\widehat{\boldsymbol{\eta}}(t),\widehat{\boldsymbol{\gamma}})$ around the values $(\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0)$:

$$\begin{aligned}\boldsymbol{F}_1(t,\widehat{\boldsymbol{\eta}}(t),\widehat{\boldsymbol{\gamma}}) &= \boldsymbol{F}_1(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0) + \boldsymbol{D}_{\boldsymbol{\eta}}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0)\left\{\widehat{\boldsymbol{\eta}}(t)-\boldsymbol{\eta}_0(t)\right\}\\ &+ \boldsymbol{D}_{\boldsymbol{\gamma}}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0)\left\{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\} + o_p(n^{-\frac{1}{2}}).\end{aligned} \tag{2.7}$$

Replacing (2.7) into the score equations (2.5) and (2.6), and denoting $\boldsymbol{D}_{\boldsymbol{\eta}}(t) = \boldsymbol{D}_{\boldsymbol{\eta}}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0)$, $\boldsymbol{D}_{\boldsymbol{\gamma}}(t) = \boldsymbol{D}_{\boldsymbol{\gamma}}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0)$ and $\boldsymbol{F}_1(t) = \boldsymbol{F}_1(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0)$, we have

$$\begin{aligned}\boldsymbol{U}_{\boldsymbol{\eta}}(t,\widehat{\boldsymbol{\eta}}(t),\widehat{\boldsymbol{\gamma}},\hat{\theta}) &= \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\left[\boldsymbol{R}(t)-\boldsymbol{F}_1(t)-\boldsymbol{D}_{\boldsymbol{\eta}}(t)\left\{\widehat{\boldsymbol{\eta}}(t)-\boldsymbol{\eta}_0(t)\right\}\right.\\ &\left.-\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\left\{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\}\right] + o_p(n^{\frac{1}{2}})\end{aligned} \tag{2.8}$$

$$\begin{aligned}\boldsymbol{U}_{\boldsymbol{\gamma}}(\tau,\widehat{\boldsymbol{\eta}}(\cdot),\widehat{\boldsymbol{\gamma}},\hat{\theta}) &= \int_0^{\tau}\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\left[\boldsymbol{R}(t)-\boldsymbol{F}_1(t)-\boldsymbol{D}_{\boldsymbol{\eta}}(t)\left\{\widehat{\boldsymbol{\eta}}(t)-\boldsymbol{\eta}_0(t)\right\}\right.\\ &\left.-\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\left\{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\}\right]dt + o_p(n^{\frac{1}{2}}).\end{aligned} \tag{2.9}$$

Solving equations (2.8) and (2.9) for $\widehat{\boldsymbol{\eta}}(t)$ and $\widehat{\boldsymbol{\gamma}}$, the estimators for $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}(t)$ are given by

$$\begin{aligned}\widehat{\boldsymbol{\gamma}} &= \boldsymbol{\gamma}_0 + \left\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})\right\}^{-1} \boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta}) + o_p(n^{-\frac{1}{2}}) \quad (2.10)\\ \widehat{\boldsymbol{\eta}}(t) &= \boldsymbol{\eta}_0(t)\\ &+\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\hat{\theta})\}^{-1}\{\boldsymbol{D}_{\boldsymbol{\eta}}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\Big\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t)-\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})\}^{-1}\boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta})\Big\}\\ &+o_p(n^{-\frac{1}{2}}), \quad (2.11)\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta) &= \int_0^{\tau} \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta)\boldsymbol{H}(t,\theta)\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\, dt,\\ \boldsymbol{B}_{\boldsymbol{\gamma}}(\theta) &= \int_0^{\tau} \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta)\boldsymbol{H}(t,\theta)\left\{\boldsymbol{R}(t) - \boldsymbol{F}_1(t)\right\}\, dt,\\ \boldsymbol{H}(t,\theta) &= I - \boldsymbol{D}_{\boldsymbol{\eta}}(t)\left[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta)\right]^{-1}\boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta),\\ \boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta) &= \boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta)\boldsymbol{D}_{\boldsymbol{\eta}}(t). \quad (2.12)\end{aligned}$$

The estimators $\widehat{\boldsymbol{\eta}}(t)$ and $\widehat{\boldsymbol{\gamma}}$ can be solved iteratively similar to Scheike et al. (2008) based on the equations (2.10) and (2.11). Specifically, the $(m+1)-$th iterative estimators are obtained by replacing $\widehat{\boldsymbol{\gamma}}^{(m)}$ and $\widehat{\boldsymbol{\eta}}^{(m)}(t)$ for $\boldsymbol{\gamma}_0$ and $\boldsymbol{\eta}_0(t)$ on the right side of (2.10) and (2.11) as the $m-$th step estimators, and replacing $\widehat{\boldsymbol{\gamma}}^{(m+1)}$ and $\widehat{\boldsymbol{\eta}}^{(m+1)}(t)$ for $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\eta}}(t)$ on the left side (2.10) and (2.11) as the $(m+1)-$th step estimators;

$$\widehat{\gamma}^{(m+1)} = \widehat{\gamma}^{(m)} + \left\{\mathcal{I}^{(m)}_{\boldsymbol{\gamma}}(\hat{\theta})\right\}^{-1} \boldsymbol{B}^{(m)}_{\boldsymbol{\gamma}}(\hat{\theta}) \tag{2.13}$$

$$\begin{aligned}\widehat{\boldsymbol{\eta}}^{(m+1)}(t) = {} & \widehat{\boldsymbol{\eta}}^{(m)}(t) \\ & +\{\mathcal{I}^{(m)}_{\boldsymbol{\eta}}(t,\hat{\theta})\}^{-1}\{\boldsymbol{D}^{(m)}_{\boldsymbol{\eta}}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\left\{\boldsymbol{R}(t)-\boldsymbol{F}^{(m)}_1(t)\right. \\ & \left.-\boldsymbol{D}^{(m)}_{\boldsymbol{\gamma}}(t)\{\mathcal{I}^{(m)}_{\boldsymbol{\gamma}}(\hat{\theta})\}^{-1}\boldsymbol{B}^{(m)}_{\boldsymbol{\gamma}}(\hat{\theta})\right\},\end{aligned} \tag{2.14}$$

where $\mathcal{I}^{(m)}_{\boldsymbol{\gamma}}(\hat{\theta})$, $\boldsymbol{B}^{(m)}_{\boldsymbol{\gamma}}(\hat{\theta})$, $\boldsymbol{H}^{(m)}(t,\hat{\theta})$ and $\mathcal{I}^{(m)}_{\boldsymbol{\eta}}(t,\hat{\theta})$ are $m-$th step estimators of $\mathcal{I}_{\boldsymbol{\gamma}}(\hat{\theta})$, $\boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta})$, $\boldsymbol{H}(t,\hat{\theta})$ and $\mathcal{I}_{\boldsymbol{\eta}}(t,\hat{\theta})$ obtained by plugging mth step estimators $(\widehat{\boldsymbol{\eta}}^{(m)}(t), \widehat{\gamma}^{(m)})$ of $(\boldsymbol{\eta}(t),\gamma)$ into $\boldsymbol{D}_{\boldsymbol{\eta}}(t,\boldsymbol{\eta}(t),\gamma)$, $\boldsymbol{D}_{\boldsymbol{\gamma}}(t,\boldsymbol{\eta}(t),\gamma)$ and $\boldsymbol{F}_1(t,\boldsymbol{\eta}(t),\gamma)$. This approach can be implemented by using 'timereg' package for R developed by Scheike and Zhang (2008).

2.2 Asymptotic Results

2.2.1 Asymptotic Results Concerning $\hat{\theta}$

The estimation method requires the following property on the maximal likelihood estimator $\hat{\theta}$ and the estimator of censoring distribution $\widehat{G}(t)$.

Proposition 1. *Let $\varphi_1(\mathcal{V}_i,\theta_1)$ and $\varphi_2(\mathcal{V}_i,\theta_2)$ be the parametric models for $P(\xi_i = 1|\mathcal{V}_i,\tilde{\epsilon}_i = 1)$ and $P(\xi_i = 1|\mathcal{V}_i,\tilde{\epsilon}_i \neq 1)$, respectively. Assume that the logistic model logit$(\varphi_1(\mathcal{V}_i,\theta_1)) = \theta_1^{\mathsf{T}}\mathcal{V}_i$ holds for those $\tilde{\epsilon}_i = 1$ and the different logistic model logit$(\varphi_2(\mathcal{V}_i,\theta_2)) = \theta_2^{\mathsf{T}}\mathcal{V}_i$ holds for those $\tilde{\epsilon}_i \neq 1$. Let $\theta = (\theta_1^{\mathsf{T}},\theta_2^{\mathsf{T}})^{\mathsf{T}}$ and let $\theta_0 = (\theta_{01}^{\mathsf{T}},\theta_{02}^{\mathsf{T}})^{\mathsf{T}}$ be the true value of θ. Then the maximum likelihood estimator $\hat{\theta}$ satisfies*

$$n^{\frac{1}{2}}(\hat{\theta}-\theta_0) = n^{-\frac{1}{2}}\left[J\left(\mathcal{V}_i,\theta_0\right)\right]^{-1}\sum_{i=1}^{n}U(\mathcal{V}_i,\theta_0) + o_p(1), \tag{2.15}$$

where the fisher information matrix

$$J(\mathcal{V}_i,\theta_0) \;=\; \begin{pmatrix} E_{\theta_0}\left[I(\tilde{\epsilon}_i=1)\frac{\exp\{\theta_{01}^{\mathcal{T}}\mathcal{V}_i\}\mathcal{V}_i\mathcal{V}_i^{\mathcal{T}}}{(1+\exp^{\theta_{01}^{\mathcal{T}}\mathcal{V}_i})^2}\right] & 0 \\ 0 & E_{\theta_0}\left[I(\tilde{\epsilon}_i\neq 1)\frac{\exp\{\theta_{02}^{\mathcal{T}}\mathcal{V}_i\}\mathcal{V}_i\mathcal{V}_i^{\mathcal{T}}}{(1+\exp^{\theta_{02}^{\mathcal{T}}\mathcal{V}_i})^2}\right] \end{pmatrix} \tag{2.16}$$

and

$$U(\mathcal{V}_i,\theta_0) \;=\; \begin{pmatrix} I(\tilde{\epsilon}_i=1)[\xi_i\mathcal{V}_i-\frac{\mathcal{V}_i\exp\{\theta_{01}^{\mathcal{T}}\mathcal{V}_i\}}{1+\exp^{\theta_{01}^{\mathcal{T}}\mathcal{V}_i}}]^{\mathcal{T}} \\ \\ I(\tilde{\epsilon}_i\neq 1)[\xi_i\mathcal{V}_i-\frac{\mathcal{V}_i\exp\{\theta_{02}^{\mathcal{T}}\mathcal{V}_i\}}{1+\exp^{\theta_{02}^{\mathcal{T}}\mathcal{V}_i}}]^{\mathcal{T}} \end{pmatrix}.$$

Proof of Proposition 1 is shown in section 4.1.

2.2.2 Asymptotic Results Concerning $\widehat{G}(t)$

Proposition 2. *The estimator $\widehat{G}(t)$ is asymptotically linear estimator of $G(t)$ with the following influence function such that*

(1) Assume that the censoring time is independent of the covariates. If $G(t) > 0$, then

$$n^{\frac{1}{2}}(\widehat{G}(t)-G(t)) = n^{-\frac{1}{2}}\left\{-G(t)\sum_{j=1}^{n}\int_0^{\tau} I(Y_{\bullet}(s)>0)\frac{dM_j{}^c(s)}{y(s)}\right\} + o_p(1), \tag{2.17}$$

where $n^{-1}Y_{\bullet}(s) \xrightarrow{p} y(s)$ as $n \xrightarrow{p} \infty$ with $Y_{\bullet}(s) = \sum_{i=1}^{n} Y_i(s) = \sum_{i=1}^{n}\mathcal{I}(\widetilde{T}_i \geq s)$, and $M_j{}^c(s) = \mathcal{I}(\widetilde{T}_j \leq s, \Delta_j = 0) - \int_0^s Y_j(u)\ d(-\log G(u))$ is the martingale associated with the censoring time.

(2) If the censoring time follows the Cox model depending on the covariates (X_i, Z_i)*.*

$$n^{\frac{1}{2}}(\widehat{G}(t|X_i,Z_i) - G(t|X_i,Z_i)) = n^{-\frac{1}{2}}\left\{-G(t|X_i,Z_i)\sum_{j=1}^{n}\int_0^{\tau} I(Y_\bullet(s)>0)\frac{dM_j{}^c(s)}{y(s)}\right\} + o_p(1), \quad (2.18)$$

where $n^{-1}Y_\bullet(s) \xrightarrow{p} y(s)$ *as* $n \xrightarrow{p} \infty$ *with* $Y_\bullet(s) = \sum_{i=1}^n Y_i(s)\exp\left(\beta_0^{\mathsf{T}} X_i(s) + \beta_1^{\mathsf{T}} Z_i(s)\right)$ *and* $M_j{}^c(s) = \mathcal{I}(\widetilde{T}_j \le s, \Delta_j = 0) - \int_0^s Y_j(u)\exp\left(\beta_0^{\mathsf{T}} X_i(u) + \beta_1^{\mathsf{T}} Z_i(u)\right)\Lambda_0(u)$ *with* $\Lambda_0(u) = \sum_{i=1}^n \int_0^u \frac{d\mathcal{I}(\widetilde{T}_j \le s, \Delta_j = 0)}{\sum_{i=1}^n Y_i(s)\exp\left(\beta_0^{\mathsf{T}} X_i(s)+\beta_1^{\mathsf{T}} Z_i(s)\right)}$ *is the martingale associated with the censoring time.*

Proof of Proposition 2 is shown in section 4.1.

2.2.3 Asymptotic Properties for IPW Estimator

We denote $F_1(t;\boldsymbol{X}_i,\boldsymbol{Z}_i)$, $\boldsymbol{D}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})$ and $\boldsymbol{D}_{\boldsymbol{\gamma},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})$ by $\boldsymbol{F}_{1i}(t)$, $\boldsymbol{D}_{\boldsymbol{\eta},i}(t)$ and $\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)$, respectively. Let

$$\begin{aligned}
\boldsymbol{A}_i(\theta) &= \partial\psi_i(\theta)/\partial\theta,\\
\boldsymbol{k}(t,\theta) &= E\left\{\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t)w_i(t)\boldsymbol{\psi}_i(\theta)\boldsymbol{D}_{\boldsymbol{\eta},i}(t)\right\}\left[E\left\{\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)w_i(t)\boldsymbol{\psi}_i(\theta)\boldsymbol{D}_{\boldsymbol{\eta},i}(t)\right\}\right]^{-1},\\
\boldsymbol{q}_{\boldsymbol{\gamma}}(s,t,\theta) &= \mathrm{E}\left[\left\{\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},j}(t) - \boldsymbol{k}(t,\theta)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},j}(t)\right\}w_j(t)\boldsymbol{\psi}_j(\theta)\frac{\Delta_j N_j(t)}{G(T_j)}\mathcal{I}(s\le \widetilde{T}_j \le t)\right],\\
g(\tau,\theta) &= E\left\{\int_0^{\tau}\left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t) - \boldsymbol{k}(t,\theta)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right]w_i(t)\boldsymbol{A}_i(\theta)\left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\}dt\right\},\\
\boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t,\theta) &= \left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t) - \boldsymbol{k}(t,\theta)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right]w_i(t)\boldsymbol{\psi}_i(\theta)\left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\},\\
M_i{}^c(t) &= \mathcal{I}(\widetilde{T}_i \le t, \Delta_i = 0) - \int_0^t Y_i(s)d(-\log G(s)),\ \text{where } Y_i(s) = \mathcal{I}(\widetilde{T}_i \ge s),\\
y(t) &= \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mathcal{I}(\widetilde{T}_i \ge t), \quad \text{uniformly} \quad t\in[0,\tau],\\
\boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t,\theta) &= \left\{\int_0^{\tau}\frac{\boldsymbol{q}_{\boldsymbol{\gamma}}(s,t,\theta)}{y(s)}dM_i^c(s)\right\}. \qquad (2.19)
\end{aligned}$$

Theorem 2.1. *Under Condition I in section 4.1,*

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}_\gamma) \tag{2.20}$$

where $\boldsymbol{\Sigma}_\gamma = \boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0)^{-1} E\{\boldsymbol{W}_{\gamma,i}(\tau,\theta_0)\}^{\otimes 2} \boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0)^{-1}$ *and where*

$$\begin{aligned}
\boldsymbol{W}_{\gamma,i}(\tau,\theta) &= \int_0^\tau \boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t,\theta)\,dt + \int_0^\tau \boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t,\theta)\,dt + g(\tau,\theta)\,[J(\mathcal{V}_i,\theta)]^{-1} U(\mathcal{V}_i,\theta),\\
\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta) &= E\left\{\int_a^\tau \left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t) - \boldsymbol{k}(t,\theta)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right] w_i(t)\boldsymbol{\psi}_i(\theta)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt\right\}.
\end{aligned}$$

The asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ can be consistently estimated by

$$\widehat{\boldsymbol{\Sigma}}_\gamma = \widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}^{-1}(\hat{\theta})\, n^{-1}\sum_{i=1}^n \left\{\widehat{\boldsymbol{W}}_{\gamma,i}(\tau,\hat{\theta})\right\}^{\otimes 2} \widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}^{-1}(\hat{\theta}), \tag{2.21}$$

where

$$\begin{aligned}
\widehat{\boldsymbol{W}}_{\gamma,i}(\tau,\theta) &= \int_0^\tau \widehat{\boldsymbol{\zeta}}_{\boldsymbol{\gamma},i}(t,\theta)\,dt + \int_0^\tau \widehat{\boldsymbol{\kappa}}_{\boldsymbol{\gamma},i}(t,\theta)\,dt + \hat{g}(\tau,\theta)\left[\widehat{J}(\mathcal{V}_i,\theta)\right]^{-1} U(\mathcal{V}_i,\theta),\\
\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}(\theta) &= n^{-1}\sum_{i=1}^n \int_0^\tau \left[\widehat{\boldsymbol{D}}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t) - \widehat{\boldsymbol{K}}(t,\theta)\widehat{\boldsymbol{D}}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right] w_i(t)\boldsymbol{\psi}_i(\theta)\widehat{\boldsymbol{D}}_{\boldsymbol{\gamma},i}(t)\,dt,
\end{aligned}$$

where $\widehat{\boldsymbol{\zeta}}_{\boldsymbol{\gamma},i}(t,\theta)$, $\widehat{\boldsymbol{\kappa}}_{\boldsymbol{\gamma},i}(t,\theta)$ and $\hat{g}(\tau,\theta)$, defined in (4.29) in section 4.1, are the estimators of $\boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t,\theta)$, $\boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t,\theta)$ and $g(\tau,\theta)$. Those estimators can be obtained by replacing $\psi_i(\theta)$ with $\psi_i(\hat{\theta}) = \xi_i/\varphi(\mathcal{V}_i,\hat{\theta})$ and by replacing $\boldsymbol{F}_{1i}(t)$, $\boldsymbol{D}_{\boldsymbol{\eta},i}(t)$,$\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)$,$\boldsymbol{k}(t,\theta)$, $M_i^c(t)$, $\boldsymbol{q}_{\boldsymbol{\gamma}}(s,t,\theta)$, $y(t)$ with $\widehat{\boldsymbol{F}}_{1i}(t)$, $\widehat{\boldsymbol{D}}_{\boldsymbol{\eta},i}(t)$, $\widehat{\boldsymbol{D}}_{\boldsymbol{\gamma},i}(t)$, $\widehat{\boldsymbol{K}}(t,\theta)$, $\widehat{\boldsymbol{q}}_{\boldsymbol{\gamma}}(s,t,\theta)$, $\hat{y}(t)$, $\widehat{M}_i^c(t)$ defined in (4.29), which can be estimated by inserting the estimators $\hat{\theta}$,$\widehat{\boldsymbol{\eta}}(t)$, $\widehat{\boldsymbol{\gamma}}$, $\widehat{G}(t)$. For the logistic regression models for the selection probabilities $\boldsymbol{A}_i(\theta)$ can be estimated as

$$\begin{aligned}
\boldsymbol{A}_i(\hat{\theta}) &= -\xi_i I(\tilde{\epsilon}_i = 1)\varphi_1'(\mathcal{V}_i,\hat{\theta}_1)/\varphi_1^2(\mathcal{V}_i,\hat{\theta}_1) - \xi_i I(\tilde{\epsilon}_i \neq 1)\varphi_2'(\mathcal{V}_i,\hat{\theta}_2)/\varphi_2^2(\mathcal{V}_i,\hat{\theta}_2)\\
&= -\xi_i\mathcal{V}_i\{I(\tilde{\epsilon}_i = 1)\exp(-\hat{\theta}_1\mathcal{V}_i) + I(\tilde{\epsilon}_i \neq 1)\exp(-\hat{\theta}_2\mathcal{V}_i)\},
\end{aligned}$$

where $\varphi_1'(\mathcal{V}_i,\theta_1) = d\varphi_1(\mathcal{V}_i,\theta_1)/d\theta_1$ and $\varphi_2'(\mathcal{V}_i,\theta_2) = d\varphi_2(\mathcal{V}_i,\theta_2)/d\theta_2$.

Under the logistic models for the probabilities of the complete case given in Proposition 1,

$$\widehat{J}(\mathcal{V}_i,\hat\theta) \;=\; \mathrm{diag}(\widehat{J}_1(\mathcal{V}_i,\hat\theta_1), \widehat{J}_2(\mathcal{V}_i,\hat\theta_2))$$

where $\widehat{J}_1(\mathcal{V}_i,\theta_1) = n^{-1}\sum_{i=1}^n I(\tilde\epsilon_i = 1)\frac{\exp(\theta_1^\intercal\mathcal{V}_i)\mathcal{V}_i\mathcal{V}_i^\intercal}{(1+\exp(\theta_1^\intercal\mathcal{V}_i))^2}$ and $\widehat{J}_2(\mathcal{V}_i,\theta_2) = n^{-1}\sum_{i=1}^n I(\tilde\epsilon_i \neq 1)\frac{\exp(\theta_2^\intercal\mathcal{V}_i)\mathcal{V}_i\mathcal{V}_i^\intercal}{(1+\exp(\theta_2^\intercal\mathcal{V}_i))^2}$, and

$$U(\mathcal{V}_i,\hat\theta) \;=\; \begin{pmatrix} U_1(\mathcal{V}_i,\hat\theta_1) \\ \\ U_2(\mathcal{V}_i,\hat\theta_2) \end{pmatrix} = \begin{pmatrix} I(\tilde\epsilon_i = 1)\left[\xi_i\mathcal{V}_i - \frac{\mathcal{V}_i\exp(\hat\theta_1^\intercal\mathcal{V}_i)}{1+\exp(\hat\theta_1^\intercal\mathcal{V}_i)}\right] \\ \\ I(\tilde\epsilon_i \neq 1)\left[\xi_i\mathcal{V}_i - \frac{\mathcal{V}_i\exp(\hat\theta_2^\intercal\mathcal{V}_i)}{1+\exp(\hat\theta_2^\intercal\mathcal{V}_i)}\right] \end{pmatrix}.$$

Let

$$\begin{aligned}
\boldsymbol{\zeta}_{\boldsymbol{\eta},i}(t,\theta) &= \boldsymbol{D}_{\boldsymbol{\eta},i}^\intercal(t)w_i(t)\boldsymbol{\psi}_i(\theta)\left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\}, \\
\boldsymbol{q}_{\boldsymbol{\eta}}(s,t,\theta) &= E\left\{\boldsymbol{D}_{\boldsymbol{\eta},j}^\intercal(t)w_j(t)\boldsymbol{\psi}_j(\theta)\frac{\Delta_j N_j(t)}{G(T_j)}\mathcal{I}(s \le \tilde T_j \le t)\right\}, \\
\boldsymbol{\kappa}_{\boldsymbol{\eta},i}(t,\theta) &= \left\{\int_0^\tau \frac{\boldsymbol{q}_{\boldsymbol{\eta}}(s,t,\theta)}{y(s)}dM_i^c(s)\right\}, \\
\boldsymbol{Q}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta) &= E\{\boldsymbol{D}_{\boldsymbol{\eta},i}^\intercal(t)w_i(t)\psi_i(\theta)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\}.
\end{aligned}$$

Theorem 2.2. *Under Condition I in section 4.1,*

$$\sqrt{n}(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t))) = \left\{\boldsymbol{Q}_{\boldsymbol{\eta}}(t,\theta_0)\right\}^{-1} n^{-\frac{1}{2}}\sum_{i=1}^n \boldsymbol{W}_{\boldsymbol{\eta},i}(t,\theta_0) + o_p(1) \qquad (2.22)$$

uniformly in $t \in [0, \tau]$, *where*

$$\boldsymbol{W}_{\boldsymbol{\eta},i}(t,\theta) = \left\{\boldsymbol{\zeta}_{\boldsymbol{\eta},i}(t,\theta) + \boldsymbol{\kappa}_{\boldsymbol{\eta},i}(t,\theta) - \boldsymbol{Q}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta)\left\{\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta)\right\}^{-1}\boldsymbol{W}_{\gamma,i}(\tau,\theta)\right\}, \quad (2.23)$$

$$\boldsymbol{Q}_{\boldsymbol{\eta}}(t,\theta) = E\{\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t)\psi_i(\theta)\boldsymbol{D}_{\boldsymbol{\eta},i}(t)\}. \quad (2.24)$$

Thus, $\sqrt{n}(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t))$ *converges weakly to a mean zero Gaussian process on* $t \in [0, \tau]$ *with the covariance matrix* $\boldsymbol{\Sigma}_{\boldsymbol{\eta}} = \boldsymbol{Q}^{-1}_{\boldsymbol{\eta}}(t,\theta_0)\, E\left\{\boldsymbol{W}_{\boldsymbol{\eta},i}(t,\theta_0)\right\}^{\otimes 2} \boldsymbol{Q}^{-1}_{\boldsymbol{\eta}}(t,\theta_0)$.

The asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t))$ can be consistently estimated by

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}} = \widehat{\boldsymbol{Q}}^{-1}_{\boldsymbol{\eta}}(t,\hat{\theta})\, n^{-1}\sum_{i=1}^{n}\left\{\widehat{\boldsymbol{W}}_{\boldsymbol{\eta},i}(t,\hat{\theta})\right\}^{\otimes 2} \widehat{\boldsymbol{Q}}^{-1}_{\boldsymbol{\eta}}(t,\hat{\theta}).$$

where

$$\widehat{\boldsymbol{W}}_{\boldsymbol{\eta},i}(t,\theta) = \left\{\widehat{\boldsymbol{\zeta}}_{\boldsymbol{\eta},i}(t,\theta) + \widehat{\boldsymbol{\kappa}}_{\boldsymbol{\eta},i}(t,\theta) - \widehat{\boldsymbol{Q}}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta)\left\{\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}(\theta)\right\}^{-1}\widehat{\boldsymbol{W}}_{\gamma,i}(\tau,\theta)\right\},$$

$$\widehat{\boldsymbol{Q}}_{\boldsymbol{\eta}}(t,\theta) = n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{D}}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t)\boldsymbol{\psi}_i(\theta)\widehat{\boldsymbol{D}}_{\boldsymbol{\eta},i}(t).$$

where $\widehat{\boldsymbol{\zeta}}_{\boldsymbol{\eta},i}(t,\theta)$, $\widehat{\boldsymbol{\kappa}}_{\boldsymbol{\eta},i}(t,\theta)$, $\widehat{\boldsymbol{Q}}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta)$, defined in (4.44) in section 4.1, are the estimators of $\boldsymbol{\zeta}_{\boldsymbol{\eta},i}(t,\theta)$, $\boldsymbol{\kappa}_{\boldsymbol{\eta},i}(t,\theta)$, $\boldsymbol{Q}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta)$. Similarly, those estimators can be obtained by replacing $\psi_i(\theta)$, $\boldsymbol{F}_{1i}(t)$, $\boldsymbol{D}_{\boldsymbol{\eta},i}(t)$,$\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)$, $\boldsymbol{q}_{\boldsymbol{\eta}}(s,t,\theta)$, $M_i^c(t)$, $y(s)$ with $\psi_i(\hat{\theta})$, $\widehat{\boldsymbol{F}}_{1i}(t)$, $\widehat{\boldsymbol{D}}_{\boldsymbol{\eta},i}(t)$, $\widehat{\boldsymbol{D}}_{\boldsymbol{\gamma},i}(t)$, $\widehat{\boldsymbol{q}}_{\boldsymbol{\eta}}(s,t,\theta)$, $\widehat{M_i}^c(t)$, $\hat{y}(s)$ defined in (4.44) in section 4.1, which can be estimated by inserting the estimators $\hat{\theta}$, $\widehat{\boldsymbol{\eta}}(t)$, $\widehat{\boldsymbol{\gamma}}$, $\widehat{G}(t)$. The estimators $\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}(\theta)$ and $\widehat{\boldsymbol{W}}_{\gamma,i}(\tau,\theta)$ are described in Theorem 2.1.

2.3 Simulations

In this section, simulation studies have been conducted to evaluate the finite sample properties of the inverse probability weighted estimators of $(\eta(t), \gamma)$. In the simulation study, the cumulative incidence function (1.22) has been considered with two different link functions. Let X_i and Z_i be Bernoulli random variables with $P(X_i = 1) = 0.5$ and $P(Z_i = 1) = 0.5X + 0.2$ for a subject i. The covariate X_i are always observed and the covariate Z_i can be missing. Let $\epsilon_i = k,\ k \in \{1, 2\}$ be the types of failure and let $k = 1$ be the event of interest. We consider the following semi-parametric models for the cumulative incidence function (1.22) with cause 1:

$$\log\{1 - F_{1i}(t; X_i, Z_i)\} = -\eta_0(t)\exp(\gamma_1 X_i + \gamma_2 Z_i), \tag{2.25}$$

$$\text{logit}\{F_{1i}(t; X_i, Z_i)\} = \eta_0(t) + \gamma_1 X_i + \gamma_2 Z_i, \tag{2.26}$$

for $0 \leq t \leq \tau$ and $\tau = 3$, where $\gamma_1 = -0.3$, $\gamma_2 = 0.3$, $\eta_0(t) = 0.2t$ for model (2.25) and $\eta_0(t) = \log(\frac{p(t)}{1-p(t)})$ with $p(t) = 0.01 + \beta\sqrt{t}$ and $\beta = 0.2$ for model (2.26).

Given $\eta_0(t), \gamma_1, \gamma_2, X_i, Z_i$, the conditional probabilities of observed failures for cause 1 for model (2.25) and (2.26) are

$$\begin{aligned} F_{1i}(\tau) &= 1 - \exp(-\eta_0(\tau)\exp({\gamma_1}^{\mathsf{T}} x_i + {\gamma_2}^{\mathsf{T}} z_i)), \\ F_{1i}(\tau) &= \frac{\exp(\eta_0(\tau) + {\gamma_1}^{\mathsf{T}} x_i + {\gamma_2}^{\mathsf{T}} z_i)}{1 + \exp(\eta_0(\tau) + {\gamma_1}^{\mathsf{T}} x_i + {\gamma_2}^{\mathsf{T}} z_i)}, \end{aligned}$$

respectively, where $0 < t \leq 3$ for individual i, $i = 1, 2, \ldots, n$. The types of failure ϵ_i have been determined by generating a Bernoulli random variable with the probability $F_1(\tau)$, that is, $P(\epsilon_i = 1) = F_{1i}(\tau)$, $i = 1, \ldots, n$. The failure time T_i is generated by the conditional probability

$$\tilde{F}_{1i}(t) = P(T_1 \le t | \epsilon_i = 1) = \frac{F_{1i}(t)}{P(\epsilon_i = 1)} = \frac{F_{1i}(t; x, z, \eta, \gamma)}{F_{1i}(\tau)} = \frac{F_{1i}(t)}{F_{1i}(\tau)}.$$

for ith individual and $\tau = 3$.

Let C_i^* follow a uniform distribution on [0,3] for both models (2.25) and (2.26). The censoring time C_i is generated by $C_i = \min(C_i^*, \tau)$. Let $\tilde{T}_i$ be the observed failure time defined as $\tilde{T}_i = \min(T_i, C_i)$. This gives us approximately 45% of the subjects who are censored before $\tau = 3$ for both models (2.25) and (2.26). Let $\tilde{\epsilon}_i = \epsilon_i \Delta_i$ where $\Delta_i = I(T_i \le C_i)$.

We consider three simulation scenarios I, II and III in terms of whether the missing probabilities depend on the outcome variables $\tilde{\epsilon}_i$ and how the phase-two covariate Z_i is missing. The first two scenarios are called phase-two sampling design: I. The first scenario is classical case cohort sampling design, where phase-two covariate Z_i is sampled for all cases $\tilde{\epsilon}_i = 1$ and the information of the covariate Z_i will be missing for the non-cases $\tilde{\epsilon}_i = 0$ or 2; II. The second scenario is generalized case-cohort sampling design, which allows the phase-two covariate Z_i to be missing for both cases and non-cases; III. The missing probability in the third scenario does not depend on $\tilde{\epsilon}_i$ and the phase-two covariate Z_i is a simple random sample from the phase-one covariates.

Let m_0 be the average of the total missing probabilities. We consider $m_0 = 0.3$ and 0.5 for each sampling scenario. Let m_1 and m_2 be the average missing probabilities for the cases and the non-cases, respectively. For model (2.25), to have $m_0 = 0.3$ for scenario I, the missing probability $\vartheta_{1_i} = P(\xi_i = 0 | \mathcal{V}_i, \tilde{\epsilon}_i = 1) = 0$ for the cases. We assume that the missing probability $\vartheta_{2_i} = P(\xi_i = 0 | \mathcal{V}_i, \tilde{\epsilon}_i \neq 1)$ follows the logistic regression model

$$\text{logit}(\varphi(\mathcal{V}_i, \theta_2)) = \theta_{20} + \theta_{21}\tilde{T}_i + \theta_{22}X_i + \theta_{23}I(\tilde{\epsilon}_i = 2) \quad (2.27)$$

for non-cases $\epsilon_i = 0$ or 2. We obtain the average missing probability $m_2 = 0.35$ by choosing $\theta_2 = (-2.0, 0.6, 0.8, 1.0)$ in model (2.27) based on phase-one covariates. Similarly, to have $m_0 = 0.5$, we have $\vartheta_{1i} = P(\xi_i = 0|\mathcal{V}_i, \tilde{\epsilon}_i = 1) = 0$ for the cases. We have approximately $m_2 = 0.65$ by choosing $\theta_2 = (1.0, 0.3, 0.1, 0.3)$ in (2.27) for the non-cases $\epsilon_i = 0$ or 2.

For scenario II, the missing probability $\vartheta_{k_i} = P(\xi_i = 0|\mathcal{V}_i, \tilde{\epsilon}_i = k)$ for both cases and non-cases can be obtained by the following logistic regression model

$$\text{logit}(\varphi(\mathcal{V}_i, \theta_k)) = \theta_{k0} + \theta_{k1}\tilde{T}_i + \theta_{k2}X_i + \theta_{k3}I(\tilde{\epsilon}_i = 1) + \theta_{k4}I(\tilde{\epsilon}_i = 2), \quad (2.28)$$

which gives approximately $m_1 = 0.15$ and $m_2 = 0.35$ when $\theta_k = (-1.0, 0.1, 0.3, -1.0, 0.5)$. Similarly, to have $m_0 = 0.5$, we have approximately $m_1 = 0.40$ and $m_2 = 0.60$ by choosing $\theta_2 = (1.0, -0.6, -0.4, -0.6, 0.3)$ in (2.28).

For scenario III, we use the following logistic model for the missing probability $\vartheta_i = P(\xi_i = 0|\mathcal{V}_i)$:

$$\text{logit}(\varphi(\mathcal{V}_i, \theta)) = \theta_0 + \theta_1 X_i. \quad (2.29)$$

For $m_0 = 0.3$, we choose $\theta = (-1.0, 0.1)$ in (2.29), yielding $\vartheta_i = 0.3$. For $m_0 = 0.5$, $\theta_3 = (-0.1, 0.3)$ is chosen in (2.29), yielding $\vartheta_i = 0.5$.

Similarly, these missing probabilities can be set up for the model (2.26). To have the average of the total missing probabilities $m_0 = 0.3$, for scenario I, we have approximately $m_2 = 0.40$ by choosing $\theta_2 = (-1.5, 0.4, 0.2, 0.4)$. For scenario II, we have approximately $m_1 = 0.15$ and $m_2 = 0.30$ when $\theta_k = (-1.5, 0.2, 0.4, -0.5, 0.5)$. For scenario III, we can get $m_0 = 0.3$ by choosing $\theta = (-1.0, 0.2)$. Similarly, we consider $m_0 = 0.5$. For scenario I, we have approximately $m_2 = 0.65$ by choosing $\theta_2 = (-0.5, 0.3, 0.5, 1.5)$. For scenario II, we have approximately $m_1 = 0.60$ and $m_2 = 0.50$ when $\theta_k = (0.5, -0.3, -0.1, 0.1, -0.3)$. For III, we can get $m_0 = 0.5$ by choosing $\theta = (0.3, -0.5)$.

We denote the full estimators as Full when all the values of the phase-two covariate Z_i are fully observed and denote the complete-case estimators as CC obtained by removing subjects having missing covariate Z_i. The performances of the proposed IPW estimators for γ_1, γ_2 and $\eta_0(t)$ over $t \in [0, 3]$ are summarized by the bias (Bias), the empirical standard error (SSE), the average of the estimated standard error (ESE), and the empirical coverage probability (CP) of 95% confidence interval. We take sample size $n = 550, 700, 900$ and obtain the average of the total missing probabilities as $m_0 = 0.3$, and 0.5 by choosing different missing probabilities for the cases m_1 and the non-cases m_2. Each entry of the tables is estimated based on 1000 simulations runs.

Table 1 and 2 summarize the Bias, SEE, ESE, and CP of the proposed IPW estimator for γ_1 and γ_2 under the three sampling designs I, II, and III for models (2.25) and (2.26). Those tables show that the IPW is an unbiased estimator. As the sample size increases, the empirical standard errors tend to decrease and the averages of the estimated standard errors tend to be closer to the empirical standard errors. The coverage probabilities are close to the 0.95 nominal level. With the higher average of total missing probabilities $m_0 = 0.5$ compared to $m_0 = 0.3$, the empirical standard errors of IPW estimators are much larger.

Table 3 and 4 compare the Bias, SSE and ESE of IPW estimators and those of complete case (CC) estimators for γ_1 and γ_2 under model (2.25) and (2.26), respectively. The Full estimator is presented as a gold standard. Table 3 shows that the IPW estimators have smaller biases than the CC estimators. The CC estimators have much larger biases, which means that CC estimators are inconsistent. The CC estimators have the largest biases when the missing probabilities for the cases and for the non-cases differ the most. Table 4 compares SSE and ESE of IPW estimators and CC estimators. The empirical standard errors of the IPW estimator are larger than those of the Full estimator. Those differences are much larger at $m_0 = 0.5$ than at $m_0 = 0.3$. This shows the IPW estimator is not efficient compared to the Full estimator. This is because we still discard the information of subjects when some of their covariates are missing. The efficiency is getting worse when there are more subjects that have missing covariates. The empirical standard errors of the CC estimators are larger than or similar to those of the IPW estimator. The coverage probability of CC estimator is worst when the missing probabilities for the cases and for the non-cases differ the most, showing that the variance of the CC estimator is inappropriate.

Figures 1-6 show that the comparison of the Full, IPW and CC estimators for the baseline cumulative time varying regression coefficient $\eta_0(t),\ t \in [0, 3]$. Figures 1-3 are plots under model (2.25) and Figures 4-6 are plots under model (2.26) in the sampling scenarios I, II and III, respectively.

We take the sample size $n = 700$. In each Figure, (a)-(c) compare the Bias, SSE, ESE for the Full, IPW and CC estimators of $\eta_0(t)$ and (d) compares the coverage probabilities for Full and IPW estimators of $\eta_0(t)$ for $m_0 = 0.3$ and 0.5 over $t \in [0, 3]$. The plots show that the IPW estimator is unbiased, comparable to Full estimators as if all the values of the covariate Z_i were observed. The biases of complete-case (CC) estimator of $\eta_0(t)$ are much larger, meaning that those estimators are inconsistent. However, under the simple random sampling III, the biases of the CC estimator are as small as those of the Full and IPW estimators, since the design of III does not depend on outcomes $\tilde{\epsilon}_i$, giving small biases for all estimators. The ESE of $\eta_0(t)$ are close to the SSE of $\eta_0(t)$. The SSE of $\eta_0(t)$ with $m_0 = 0.5$ is much larger than that of $\eta_0(t)$ with $m_0 = 0.3$. The coverage probabilities of the IPW estimators are close to the 0.95 nominal level with $m_0 = 0.3$ and 0.5. The coverage probabilities of the CC estimators for $m_0 = 0.3$ and 0.5 are not shown in (d) in each scenario since those are very far away from the 0.95 nominal level.

2.4 Application

The RV144 vaccine efficacy trial randomized 16,394 HIV negative volunteers to the vaccine ($n = 8198$) and placebo ($n = 8196$) groups. We apply the proposed estimating procedures for IPW method to the vaccine group. Subjects enrolled in the RV144 trial were vaccinated at weeks 0, 4, 12 and 24. They were monitored for 42 months for the occurrence of the primary end point of HIV infection after their immune responses were measured at week 26, which turned out that 43 out of the 8198 vaccine recipients acquired HIV infection. Vaccine recipients were distributed in the Low, Medium, and High baseline behavioral risk scores, defined as in (Rerks-Ngarm et al., 2009) with 3863 Low, 2370 Medium, and 1965 High.

The vaccine can be constructed by inserting three HIV gp120 sequences; 92TH023 in the ALVAC canarypox vector prime component; and A244 and MN in the AIDSVAX protein boost component. The 92TH023 and A244 are subtype E HIVs whereas MN is subtype B. The subtype E vaccine-insert sequences are genetically much closer to the infecting sequences (or regional circulating sequences) than MN. We expect subtype E HIVs are prone to making protective immune responses actively. Thus, our analyses focus on the 92TH023 and A244 insert sequences. The observed failure time $\tilde{T}_i$ is the time to HIV infection diagnosis, which is minimum of failure time or right-censoring time.

Many authors (Haynes et al., 2012; Yates et al., 2014; Zolla-Pazner et al., 2014) pointed out that vaccine recipients with higher levels of antibodies binding to the V1V2 portion of the HIV envelope protein had a significantly lower rate of HIV infection. Therefore, it would be meaningful to study marks defined based on the genetic distance of an infecting HIV V1V2 sequence to the corresponding V1V2 sequence in the vaccine construct (using a multiple sequence alignment). The way of measuring the genetic distances (marks) is described in Nickle et al. (2007).

We consider two marks for the analyses. Let V be the two marks based on the 92TH023 and A244 vaccine construct sequences. Let V_{1_i} be the genetic distance marks 92TH023V1V2 and let V_{2_i} be the genetic distance marks A244V1V2 for a subject i. These mark variables $V = (V_{1_i}, V_{2_i})$ were re-scaled to take values between 0 and 1.

Two causes of failure are constructed by two marks $V = (V_{1_i}, V_{2_i})$. Let M_1 be the median of the observed marks V_{1_i}. We define $\epsilon_{1_i} = 1$ for a uncensored subject i if V_{1_i}

is less than M_1; otherwise $\epsilon_{1_i} = 2$. Let M_2 be the median of the observed marks V_{2_i}. Similarly, two causes of failure $\epsilon_{2_i} = 1$ and 2 for the analysis with mark V_{2_i} can be constructed by using M_2. If subjects are censored, then $\epsilon_{j_i} = 0$, where $j = 1, 2$.

Following the analysis in Yang et al. (2016), we study IgG and IgG3 biomarkers as correlates of 92TH023V1V2 and A244V1V2 mark-specific HIV infection for the cumulative incidence model based on competing risks data under two-phase sampling. In particular, paired to the 92TH023V1V2 mark variable, we study the two biomarkers Week 26 IgG and IgG3 binding antibodies to 92TH023V1V2, namely IgG-92TH023V1V2 and IgG3-92TH023V1V2; and, paired to the A244V1V2 mark variable, we study Week 26 IgG and IgG3 binding antibodies to A244V1V2, namely IgG-A244V1V2 and IgG3-A244V1V2. Therefore, we consider four different immune responses IgG-92TH023V1V2, IgG3-92TH023V1V2, IgG-A244V1V2 and IgG3-A244V1 V2 for each analysis. The immune response biomarkers were measured for 34 of 43 HIV infected vaccine recipients with HIV V1V2 sequence data and 212 of 8155 uninfected vaccine recipients at the Week 26 visit post entry. These observed biomarkers were each standardized to have mean 0 and variance 1 for each analysis.

Let R_i be the four different immune responses and define R_{11_i} = IgG-92TH023V1V2, R_{12_i} = IgG3-92TH023V1V2, R_{21_i} = IgG-A244V1V2, and R_{22_i} = IgG3-A244V1V2. For the mark V_{1_i}, each of R_{11_i} and R_{12_i} is analyzed, and, for the mark V_{2_i}, each of R_{21_i} and R_{22_i} is analyzed. Let δ_i be the infection status, whose value is 1 if a subject is infected the HIV; and 0 if a subject i is censored over a follow-up period of 42 months. We consider two causes of failure $\epsilon_{1_i} = k$ for immune responses R_{11_i} and R_{12_i}, respectively, and $\epsilon_{2_i} = k$ for immune responses R_{21_i} and R_{22_i}, respectively, where $k = 1, 2$. Let B_{1_i} and B_{2_i} be the dummy variables for baseline behavioral risk score groups B_i (High=1, Low=2, Medium=3), where $B_{1_i} = 1$ if a subject i is in the low risk score group; 0 otherwise, $B_{2_i} = 1$ if a subject i is in the medium risk score group; 0 otherwise and $B_{1_i} = B_{2_i} = 0$ if a subject i is in the high risk group.

The immune response R_i can be missing for both case and non-case subjects, and hence are phase two covariates. The baseline behavioral risk scores B_i are measured for all subjects, and hence are phase one covariates.

We consider the following semiparametric additive model for the cumulative incidence function by using the log link function $h(x) = 1 - \exp(-x)$ in (1.22);

$$F_{ki}(t; X_i, Z_i) = 1 - \exp(-\{\eta_0(t) + \eta_1(t)R_i + \gamma_1 B_{1_i} t + \gamma_2 B_{2_i} t\}) \tag{2.30}$$

for $k = 1, 2$.

Let ξ_i be the missing indicator of the immune response data, whose value is $\xi_i = 1$ if each of the four immune response data R_i is measured as a phase-two covariate; otherwise $\xi_i = 0$. Let $\vartheta_i = P(\xi_i = 1 | \mathcal{V}_i, \delta_i)$ be the selection probability for a subject i. To predict the probability of observing the immune response R_i, we consider the following logistic model

$$\text{logit}(\vartheta_i) = \theta_0 + \theta_1 \delta_i \tag{2.31}$$

The estimated selection probabilities $\hat{\vartheta}_i = P(\xi_i = 1 | \mathcal{V}_i, \delta_i)$ are given by $\hat{\theta} = (-3.6235,$ $4.9526)$ with standard errors (0.0696, 0.3813) of coefficients $\hat{\theta}$ in (2.31). The weights are estimated by $\psi(\hat{\theta}_i) = \xi_i / \hat{\vartheta}_i$.

We analyze the semiparametric additive model (2.30) with four different settings:

(S1). The model (2.30) is analyzed with immune response $R_i = R_{11_i}$ for $\epsilon_{1_i} = 1$ and 2, respectively. For $\epsilon_{1_i} = 1$, the IPW estimates of baseline behavioral risk score for B_{1i} and B_{2i} are $(\hat{\gamma}_1, \hat{\gamma}_2) = (-0.00115, -0.00122)$ with standard errors with (0.00074, 0.00074), yielding p-values (0.11874, 0.09653). Similarly, for $\epsilon_{1_i} = 2$, we have the estimates $(\hat{\gamma}_1, \hat{\gamma}_2) = (-0.00015, 0.00065)$ with standard errors with (0.00036, 0.00058), yielding p-values (0.66948, 0.25593);

(S2). The model (2.30) can be analyzed with immune response $R_i = R_{12_i}$ for $\epsilon_{1_i} = 1$ and 2, respectively. For $\epsilon_{1_i} = 1$, the IPW estimates of B_{1_i} and B_{2_i} are $(\hat{\gamma}_1, \hat{\gamma}_2) = (-0.00108, -0.00123)$ with standard errors (0.00075, 0.00074), yielding p-values (0.15117, 0.09575). For $\epsilon_{1_i} = 2$, we have the estimates $(\hat{\gamma}_1, \hat{\gamma}_2) = (-0.00017,$ $0.00064)$ with standard errors (0.00034, 0.00057), yielding p-values (0.62405, 0.25967);

(S3). The model (2.30) can be analyzed with immune response $R_i = R_{21_i}$ for $\epsilon_{2_i} = 1$ and 2, respectively. For $\epsilon_{2_i} = 1$, the IPW estimates of B_{1i} and B_{2i} are $(\hat{\gamma}_1, \hat{\gamma}_2) = (-0.00036, -0.00038)$ with standard errors (0.00059, 0.00058), yielding p-values (0.54059, 0.51436). For $\epsilon_{2_i} = 2$, we have the estimates $(\hat{\gamma}_1, \hat{\gamma}_2) = (-0.00086,$ $-0.00010)$ with standard errors (0.00050, 0.00067), yielding p-values (0.08665, 0.88050);

(S4). The model (2.30) can be analyzed with immune response $R_i = R_{22_i}$ for $\epsilon_{2_i} = 1$ and 2, respectively. For $\epsilon_{2_i} = 1$, the IPW estimates of B_{1i} and B_{2i} are $(\hat{\gamma}_1, \hat{\gamma}_2) = (-0.00039, -0.00042)$ with standard errors (0.00061, 0.00059), yielding p-values (0.52576, 0.47455). For $\epsilon_{2_i} = 2$, we have the estimates $(\hat{\gamma}_1, \hat{\gamma}_2) = (-0.00082,$ $-0.00009)$ with standard errors (0.00050, 0.00067), yielding p-values (0.09711, 0.89253).

Figures 7-10 compare IPW estimates of baseline cumulative coefficients $\eta_0(t)$ and cumulative coefficients $\eta_1(t)$ with 95% pointwise confidence intervals for the four different immune responses of R_i for $\epsilon_{j_i} = 1$ and $\epsilon_{j_i} = 2$, respectively, $j = 1, 2$. In Figures 7-8, the analyses with R_{11_i} and R_{12_i} show that the IPW estimates of baseline cumulative coefficients $\eta_0(t)$ for $\epsilon_{1_i} = 1$ are larger than those of $\eta_0(t)$ for $\epsilon_{1_i} = 2$. Similarly, the analyses with R_{21_i} and R_{22_i} show that the IPW estimates of $\eta_0(t)$ for $\epsilon_{2_i} = 1$ are similar to those of $\eta_0(t)$ for $\epsilon_{2_i} = 2$. While the immune responses R_{11_i}(IgG-92TH023V1V2), R_{21_i}(IgG-A244V1V2) and R_{22_i}(IgG3-A244V1V2) have negative effects on the cumulative incidence functions with $\epsilon_{j_i} = 1$, $j = 1, 2$ as seen in Figures 7, 9 and 10 for the IPW estimates of cumulative coefficients $\eta_1(t)$, the effects of immune responses R_{12_i}(IgG3-92TH023V1V2) are very close to zero and slightly positive in the upper range with $\epsilon_{1_i} = 1$ in Figure 8. On the other hand, the analyses with four immune responses R_i for $\epsilon_{1_i} = 2$ and $\epsilon_{2_i} = 2$, respectively, show that the cumulative coefficients $\eta_1(t)$ of each immune response are around zero over study time. By comparing Figure 7 to 9 and comparing Figure 8 to 10, IgG and IgG3 binding antibodies responding to A244V1V2 have significantly negative effects on the cumulative incidence function than those responding to 92TH023V1V2, i.e A244 would be more relevant for protection of HIV infection with V1V2 sequences.

The relationship between the behavioral risk scores and cumulative incidence of HIV infection can be summarized in Figures 11-14. For the four different immune responses R_i, these four figures show that the cumulative incidence functions $F_k(t)$ are estimated at the first, second and third quartiles Q_1, Q_2 and Q_3 of the observed immune responses R_i at each level of the behavioral risk score groups (Low, Medium and High). In each figure, for $\epsilon_{1_i} = 1$ and $\epsilon_{2_i} = 1$, the high risk score group tends to have higher probability of getting infected by HIV with V1V2 sequences than the low and medium risk score groups. However, these relationships are less noticeable for the immune responses IgG-92TH023V1V2 and IgG3-92TH023V1V2 with $\epsilon_{1_i} = 2$ in Figures 11-12.

The relationship between the immune responses and cumulative incidence of HIV infection can be also summarized in Figures 11-14. The estimated cumulative incidence functions for IgG-92TH023V1V2 with $\epsilon_{1_i} = 1$, IgG-A244V1V2 and IgG3-A244V1V2 with $\epsilon_{2_i} = 1$ show that subjects in the third quartile group (Q_3) have lower cumulative incidence of HIV infection than those in the second quartile group (Q_2), which have in turn lower cumulative incidence of HIV infection than those in the first quartile group (Q_1) at each behavioral risk group (a), (b) and (c) in Figures 11, 13, and 14. This tendency is clearly noticeable for immune responses IgG-A244V1V2 and IgG3-A244V1V2 with $\epsilon_{2_i} = 1$, meaning that one with a higher immune response may be strongly against HIV infection with the V1V2 sequences closer to A244 ($\epsilon_{2_i} = 1$).

Figures 11-14 show the relationship between the genetic distance (mark variable) and the cumulative incidence function $F_k(t)$. Note that the competing marks $\epsilon_{j_i} = 1$ and $\epsilon_{j_i} = 2$, where $\epsilon_{j_i} = 1$ is less than the median mark and $\epsilon_{j_i} = 2$ is greater than the median mark, where $j = 1, 2$. We expect there to be a lower probability of getting infected by HIV with V1V2 sequences closer to 92TH023 or A244 ($\epsilon_{j_i} = 1$, $j = 1, 2$) and expect a higher probability of getting infected by HIV with V1V2 sequences far away from 92TH023 or A244 ($\epsilon_{j_i} = 2$, $j = 1, 2$). However, for immune responses IgG-92TH023V1V2 and IgG3-92TH023V1V2, it is not always true that the marks less than the median mark lead to lower probabilities of HIV infection since these analyses depend on the prevalence of circulating HIV strain with genetic distance (specific mark). For example, (a) and (c) in Figure 11 show that the cumulative incidence of HIV infection with $\epsilon_{1_i} = 1$ is larger than that with $\epsilon_{1_i} = 2$. This shows that the HIV infection with shorter marks than the median mark may be more prevalent and exposed to more people than HIV infection with the farther marks.

These analyses imply that IgG3 antibody to 92TH023V1V2 does not have effect on cumulative incidence of HIV infection in Figure 8 and other IgG subclasses besides type 3 induced by 92TH023 would have negative effects on the cumulative incidence function. These analyses also imply that A244 was more important than 92TH023 for induction of protective IgG3 antibodies. For A244 vaccine insert sequences, marks shorter than the median of observed marks $\epsilon_i = 1$ are more important for protection against the HIV infection than mark distances larger than the median marker. These analyses support the hypothesis that IgG and IgG3 binding antibodies protect less against exposures to HIV V1V2 sequences with farther distance away from the A244V1V2 vaccine sequences and strongly against HIV infecting sequences with small V1V2 distance. Therefore, vaccine recipients exposed to HIVs with V1V2 sequences close to A244 (less marks than the median mark) may be more likely to be protected by antibodies than vaccine recipients exposed to HIVs with V1V2 sequences with farther marks than the median mark.

Table 1: Bias, empirical standard error(SSE), average of the estimated standard error(ESE), and empirical coverage probability (CP) of 95% confidence intervals for the IPW estimators of γ_1, γ_2 under model (2.25) with the average of the total missing probabilities $m_0 = 0.3,\ 0.5$ and about 45% censoring percentage based on 1000 simulations for sampling scenarios I, II and III, where m_1 and m_2 are average of the missing probabilities for the cases and the non-cases, respectively.

				γ_1				γ_2			
sampling	m_0	(m_1, m_2)	n	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
I	0.3	(0,0.35)	550	-0.0091	0.2519	0.2473	0.947	0.0123	0.2536	0.2468	0.952
			700	-0.0108	0.2090	0.2178	0.959	0.0144	0.2142	0.2174	0.953
			900	-0.0033	0.1884	0.1904	0.959	0.0035	0.2028	0.1899	0.934
II		(0.15,0.35)	550	-0.0047	0.2592	0.2529	0.956	0.0012	0.2545	0.2527	0.957
			700	-0.0076	0.2211	0.2233	0.950	0.0072	0.2171	0.2229	0.955
			900	-0.0048	0.1962	0.1954	0.955	0.0081	0.2053	0.1950	0.944
III			550	-0.0076	0.2701	0.2620	0.950	0.0074	0.2618	0.2618	0.955
			700	-0.0077	0.2363	0.2321	0.948	0.0065	0.2276	0.1318	0.959
			900	-0.0074	0.2045	0.2027	0.961	0.0061	0.2077	0.2025	0.940
I	0.5	(0,0.65)	550	0.0035	0.2511	0.2728	0.978	-0.0112	0.2870	0.2731	0.938
			700	-0.0088	0.2160	0.2409	0.971	0.0094	0.2486	0.2411	0.955
			900	-0.0090	0.1903	0.2117	0.968	0.0030	0.2191	0.2118	0.948
II		(0.40,0.60)	550	0.0015	0.3092	0.3037	0.949	0.0013	0.3106	0.3060	0.956
			700	-0.0205	0.2812	0.2686	0.943	0.0099	0.2862	0.2708	0.950
			900	-0.0047	0.2367	0.2341	0.948	0.0020	0.2456	0.2360	0.947
III			550	-0.0121	0.3308	0.3224	0.962	0.0082	0.3173	0.3210	0.962
			700	-0.0099	0.2885	0.2833	0.957	0.0013	0.2861	0.2821	0.954
			900	0.0044	0.2616	0.2476	0.941	-0.0013	0.2617	0.2464	0.944

Table 2: Bias, empirical standard error (SSE), average of the estimated standard error (ESE), and empirical coverage probability (CP) of 95% confidence intervals for the IPW estimators of γ_1, γ_2 under model (2.26) with the average of the total missing probabilities $m_0 = 0.3,\ 0.5$ and approximately 45% censoring percentage based on 1000 simulations for sampling scenarios I, II and III, where m_1 and m_2 are average of the missing probabilities for the cases and the non-cases, respectively.

				γ_1				γ_2			
sampling	m_0	(m_1, m_2)	n	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
I	0.3	(0,0.40)	550	-0.0112	0.2846	0.2804	0.953	0.0142	0.2990	0.2812	0.938
			700	-0.0021	0.2414	0.2455	0.958	-0.0034	0.2431	0.2462	0.956
			900	-0.0105	0.2091	0.2161	0.953	-0.0010	0.2162	0.2167	0.954
II		(0.15,0.30)	550	-0.0051	0.3020	0.2880	0.944	0.0113	0.3007	0.2882	0.950
			700	-0.0023	0.2576	0.2535	0.949	-0.0090	0.2518	0.2536	0.957
			900	-0.0073	0.2282	0.2230	0.948	-0.0019	0.2225	0.2231	0.959
III			550	-0.0055	0.3150	0.3006	0.940	0.0105	0.3084	0.3009	0.950
			700	-0.0031	0.2761	0.2644	0.941	-0.0054	0.2630	0.2646	0.952
			900	-0.0068	0.2392	0.2329	0.944	-0.0023	0.2337	0.2331	0.958
I	0.5	(0,0.65)	550	-0.0171	0.3404	0.3408	0.958	0.0350	0.3607	0.3414	0.953
			700	-0.0035	0.2879	0.2987	0.963	-0.0014	0.3078	0.2993	0.957
			900	-0.0180	0.2347	0.2623	0.970	0.0074	0.2574	0.2627	0.963
II		(0.60,0.50)	550	-0.0077	0.4220	0.3868	0.944	0.0136	0.4008	0.3877	0.951
			700	-0.0130	0.3445	0.3392	0.953	0.0024	0.3455	0.3409	0.948
			900	-0.0144	0.2937	0.2976	0.953	-0.0009	0.3126	0.2987	0.948
III			550	-0.0196	0.4023	0.3756	0.943	0.0278	0.4015	0.3778	0.953
			700	-0.0065	0.3263	0.3263	0.938	0.0062	0.3283	0.3283	0.952
			900	-0.0136	0.2944	0.2859	0.950	-0.0035	0.2923	0.2879	0.955

Table 3: The bias (Bias) for Full, IPW and CC estimators of γ_1 and γ_2 under model (2.25) and (2.26) with the average of the total missing probabilities $m_0 = 0.3$, 0.5 and about 45% censoring percentage based on 1000 simulations for sampling scenarios I, II and III, where m_1 and m_2 are average of the missing probabilities for the cases and the non-cases, respectively.

					Bias(γ_1)			Bias(γ_2)		
Model	Sample	m_0	(m_1, m_2)	n	Full	IPW	CC	Full	IPW	CC
(2.25)	I	0.3	(0,0.35)	550	-0.0042	-0.0091	0.2548	0.0030	0.0123	-0.0025
				700	-0.0054	-0.0108	0.2523	0.0070	0.0144	-0.0024
				900	-0.0002	-0.0033	0.2627	0.0018	0.0035	-0.0108
	II		(0.15,0.35)	550		-0.0047	0.0626		0.0012	-0.0095
				700		-0.0076	0.0612		0.0072	-0.0034
				900		-0.0048	0.0628		0.0081	-0.0019
	III			550		-0.0076	-0.0073		0.0074	0.0069
				700		-0.0077	-0.0081		0.0065	0.0066
				900		-0.0074	-0.0078		0.0061	0.0063
	I	0.5	(0,0.65)	550		0.0035	0.0841		-0.0112	-0.0436
				700		-0.0088	0.0755		0.0094	-0.0248
				900		-0.0090	0.0752		0.0030	-0.0313
	II		(0.40,0.60)	550		0.0015	-0.0424		0.0013	-0.0082
				700		-0.0205	-0.0533		0.0099	-0.0069
				900		-0.0047	-0.0453		0.0020	-0.0100
	III			550		-0.0121	-0.0160		0.0082	0.0092
				700		-0.0099	-0.0117		0.0013	0.0022
				900		0.0044	0.0020		-0.0013	0.0002
(2.26)	I	0.3	(0,0.40)	550	-0.0078	-0.0112	0.0752	0.0119	0.0142	0.0180
				700	0.0005	-0.0021	0.0759	-0.0073	-0.0034	0.0044
				900	-0.0091	-0.0105	0.0708	-0.0011	-0.0010	0.0066
	II		(0.15,0.30)	550		-0.0051	0.0489		0.0113	0.0118
				700		-0.0023	0.0534		-0.0090	-0.0085
				900		-0.0073	0.0477		-0.0019	-0.0015
	III			550		-0.0055	-0.0063		0.0105	0.0109
				700		-0.0031	-0.0036		-0.0054	-0.0059
				900		-0.0068	-0.0072		-0.0023	-0.0021
	I	0.5	(0,0.65)	550		-0.0171	0.3192		0.0350	0.0448
				700		-0.0035	0.3081		-0.0014	0.0114
				900		-0.0180	0.3109		0.0074	0.0215
	II		(0.60,0.50)	550		-0.0077	0.0057		0.0136	0.0102
				700		-0.0130	-0.0021		0.0024	-0.0003
				900		-0.0144	-0.0004		-0.0009	-0.0021
	III			550		-0.0196	-0.0124		0.0278	0.0236
				700		-0.0065	-0.0028		0.0062	0.0061
				900		-0.0136	-0.0108		-0.0035	-0.0042

Table 4: The empirical standard error (SSE), average of the estimated standard error (ESE), and empirical coverage probability (CP) of 95% confidence intervals forfor Full, IPW and CC estimators of γ_1 and γ_2 under (2.25) and (2.26) with the average of the total missing probabilities $m_0 = 0.3,\ 0.5$ and approximately 45% censoring percentage based on 1000 simulations for sampling scenarios I, II and III, where m_1 and m_2 are average of missing probabilities for the cases and the non-cases, respectively.

					SSE(γ_1)			SSE(γ_2)			ESE(γ_1)			ESE(γ_2)			CP(γ_1)			CP(γ_2)		
Model	Sample	m_0	(m_1, m_2)	n	Full	IPW	CC	Full	IPW	CC	Full	IPW	CC	Full	IPW	CC	Full	IPW	CC	Full	IPW	CC
(2.25)	I	0.3	(0,0.35)	550	0.2329	0.2519	0.2528	0.2239	0.2536	0.2475	0.2213	0.2473	0.2415	0.2213	0.2468	0.2422	0.946	0.947	0.803	0.952	0.952	0.948
				700	0.1988	0.2090	0.2193	0.1949	0.2142	0.2113	0.1957	0.2178	0.2127	0.1956	0.2174	0.2132	0.947	0.959	0.776	0.951	0.953	0.944
				900	0.1776	0.1884	0.1886	0.1835	0.2028	0.1972	0.1711	0.1904	0.1861	0.1711	0.1899	0.1866	0.950	0.959	0.970	0.930	0.934	0.932
	II		(0.15,0.35)	550		0.2592	0.2541		0.2545	0.2493		0.2529	0.2477		0.2527	0.2478		0.956	0.934		0.957	0.959
				700		0.2211	0.2204		0.2171	0.2126		0.2233	0.2188		0.2229	0.2187		0.950	0.938		0.955	0.955
				900		0.1962	0.1941		0.2053	0.2004		0.1954	0.1914		0.1950	0.1914		0.955	0.943		0.944	0.947
	III			550		0.2701	0.2699		0.2618	0.2618		0.2620	0.2618		0.2618	0.2617		0.950	0.951		0.955	0.955
				700		0.2363	0.2357		0.2276	0.2275		0.2321	0.2321		0.2318	0.2320		0.948	0.949		0.959	0.958
				900		0.2045	0.2045		0.2077	0.2077		0.2027	0.2027		0.2025	0.2026		0.961	0.962		0.940	0.938
	I	0.5	(0,0.65)	550		0.2511	0.2306		0.2870	0.2298		0.2728	0.2217		0.2731	0.2218		0.978	0.924		0.938	0.933
				700		0.2160	0.1951		0.2486	0.1940		0.2409	0.1954		0.2411	0.1956		0.971	0.942		0.955	0.951
				900		0.1903	0.1709		0.2191	0.1740		0.2117	0.1723		0.2118	0.1724		0.968	0.936		0.948	0.943
	II		(0.40,0.60)	550		0.3092	0.3018		0.3106	0.3004		0.3037	0.2963		0.3060	0.2058		0.949	0.950		0.956	0.960
				700		0.2812	0.2717		0.2862	0.2702		0.2686	0.2617		0.2708	0.2614		0.943	0.940		0.950	0.955
				900		0.2367	0.2318		0.2456	0.2352		0.2341	0.2238		0.2360	0.2280		0.948	0.947		0.947	0.948
	III			550		0.3308	0.3314		0.3173	0.3169		0.3224	0.3229		0.3210	0.3214		0.962	0.961		0.962	0.966
				700		0.2885	0.2885		0.2861	0.2865		0.2833	0.2838		0.2821	0.2828		0.957	0.953		0.954	0.955
				900		0.2616	0.2606		0.2617	0.2601		0.2476	0.2479		0.2464	0.2468		0.941	0.944		0.944	0.946
(2.26)	I	0.3	(0,0.40)	550	0.2666	0.2846	0.3023	0.2630	0.2990	0.3026	0.2523	0.2804	0.2856	0.2527	0.2812	0.2866	0.938	0.953	0.931	0.947	0.938	0.939
				700	0.2275	0.2414	0.2614	0.2188	0.2431	0.2468	0.2218	0.2455	0.2509	0.2221	0.2462	0.2517	0.944	0.958	0.931	0.950	0.956	0.955
				900	0.1995	0.2091	0.2223	0.1951	0.2162	0.2205	0.1954	0.2161	0.2209	0.1958	0.2167	0.2216	0.944	0.953	0.939	0.953	0.954	0.960
	II		(0.15,0.30)	550		0.3020	0.3050		0.3007	0.2994		0.2880	0.2879		0.2882	0.2883		0.944	0.936		0.950	0.949
				700		0.2576	0.2593		0.2518	0.2526		0.2535	0.2537		0.2536	0.2541		0.949	0.943		0.957	0.954
				900		0.2282	0.2313		0.2225	0.2226		0.2230	0.2232		0.2231	0.2235		0.948	0.926		0.959	0.959
	III			550		0.3150	0.3149		0.3084	0.3088		0.3006	0.3008		0.3009	0.3012		0.940	0.937		0.950	0.955
				700		0.2761	0.2765		0.2630	0.2639		0.2644	0.2647		0.2646	0.2650		0.941	0.939		0.952	0.949
				900		0.2392	0.2396		0.2337	0.2344		0.2329	0.2332		0.2331	0.2335		0.944	0.943		0.958	0.959
	I	0.5	(0,0.65)	550		0.3404	0.3464		0.3607	0.3498		0.3408	0.3407		0.3414	0.3423		0.958	0.833		0.953	0.952
				700		0.2879	0.3092		0.3078	0.3045		0.2987	0.3008		0.2993	0.3021		0.963	0.802		0.957	0.949
				900		0.2460	0.2701		0.2574	0.2609		0.2623	0.2641		0.2627	0.2652		0.970	0.774		0.963	0.964
	II		(0.60,0.50)	550		0.4220	0.4297		0.4008	0.4056		0.3868	0.3874		0.3877	0.3875		0.944	0.941		0.951	0.952
				700		0.3445	0.3475		0.3455	0.3463		0.3397	0.3400		0.3409	0.3404		0.953	0.956		0.948	0.952
				900		0.2937	0.2946		0.3126	0.3127		0.2976	0.2979		0.2987	0.2982		0.953	0.956		0.948	0.952
	III			550		0.4023	0.3985		0.4015	0.3939		0.3756	0.3708		0.3778	0.3690		0.943	0.946		0.953	0.950
				700		0.3263	0.3229		0.3283	0.3237		0.3263	0.3229		0.3283	0.3214		0.938	0.939		0.952	0.958
				900		0.2944	0.2930		0.2923	0.2847		0.2859	0.2831		0.2879	0.2819		0.950	0.950		0.955	0.954

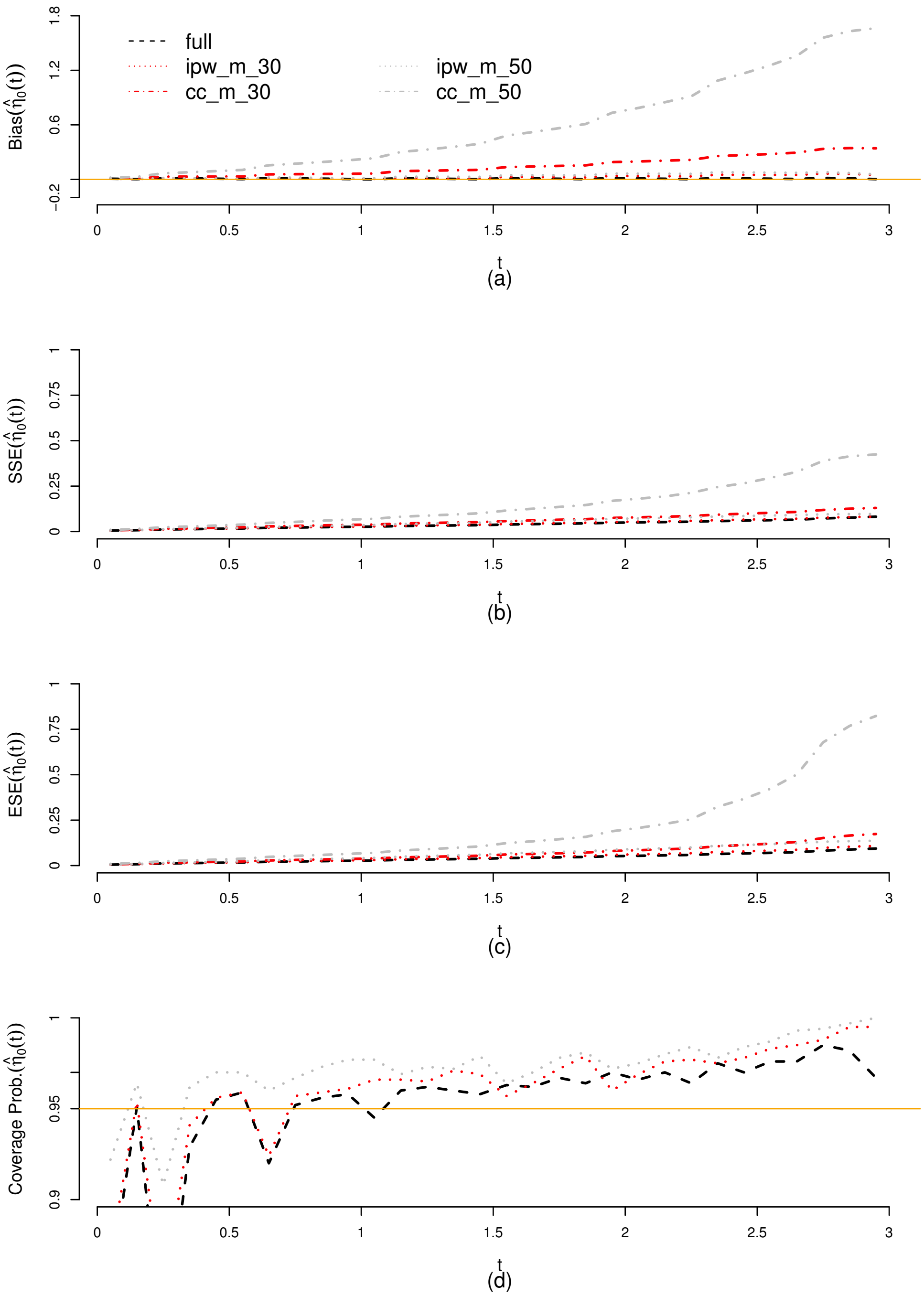


Figure 1: Comparison of Full, IPW, CC estimators for the baseline cumulative coefficient $\eta_0(t)$ under (2.25) with $m_0 = 0.3,\ 0.5$, $n = 700$ and approximately 45% censoring percentage based on 1000 simulations for sampling design I. (a): The plots of the biases of the estimates. (b): The plots of the empirical standard errors of the estimates. (c): The plots of the average of the estimated standard errors of the estimates. (d): The plots of the coverage probabilities of the Full and IPW estimators.

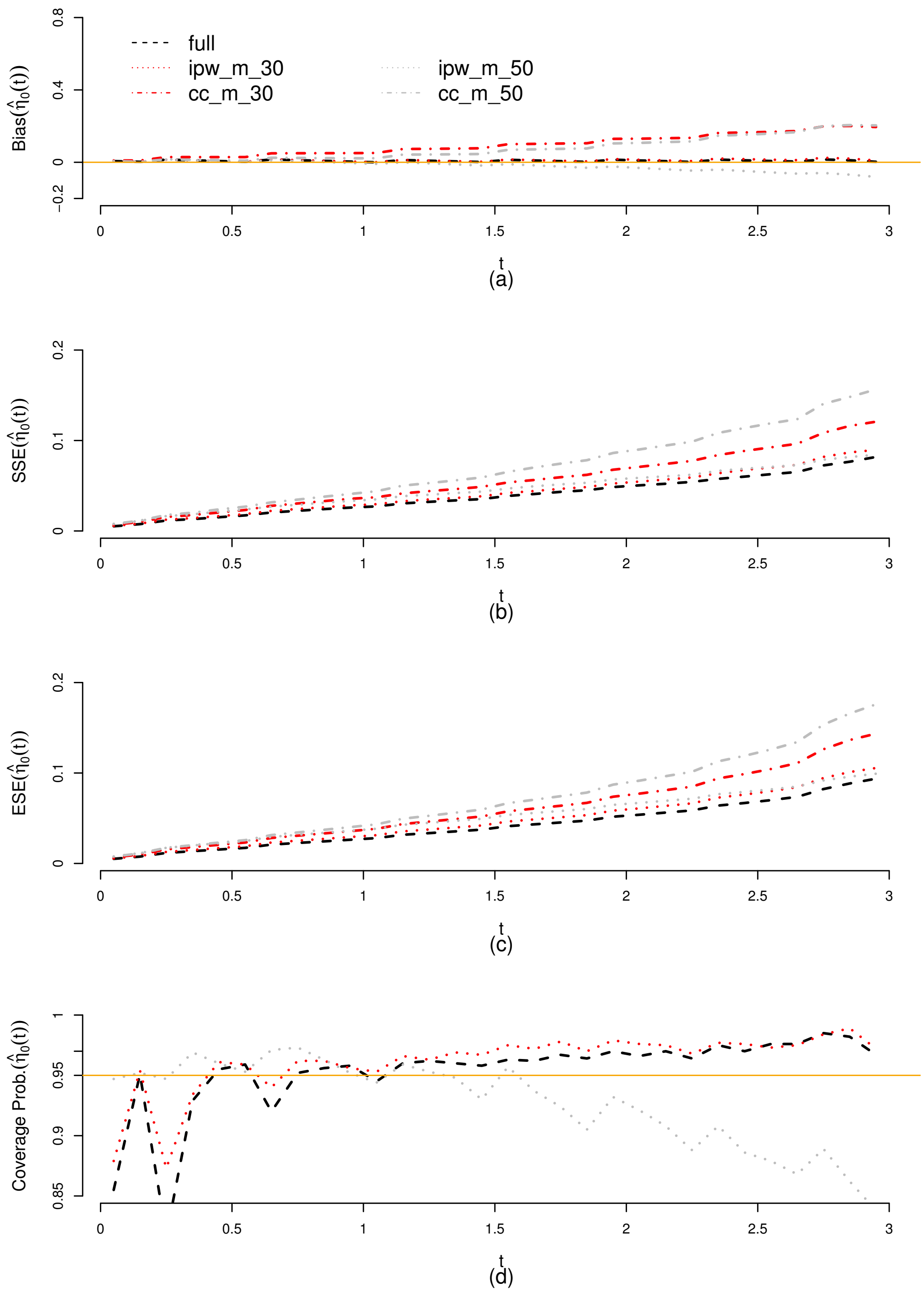


Figure 2: Comparison of Full, IPW, CC estimators for the baseline cumulative coefficient $\eta_0(t)$ under (2.25) with $m_0 = 0.3,\ 0.5$, $n = 700$ and approximately 45% censoring percentage based on 1000 simulations for sampling design II. (a): The plots of the biases of the estimates. (b): The plots of the empirical standard errors of the estimates. (c): The plots of the average of the estimated standard errors of the estimates. (d): The plots of the coverage probabilities of the Full and IPW estimators.

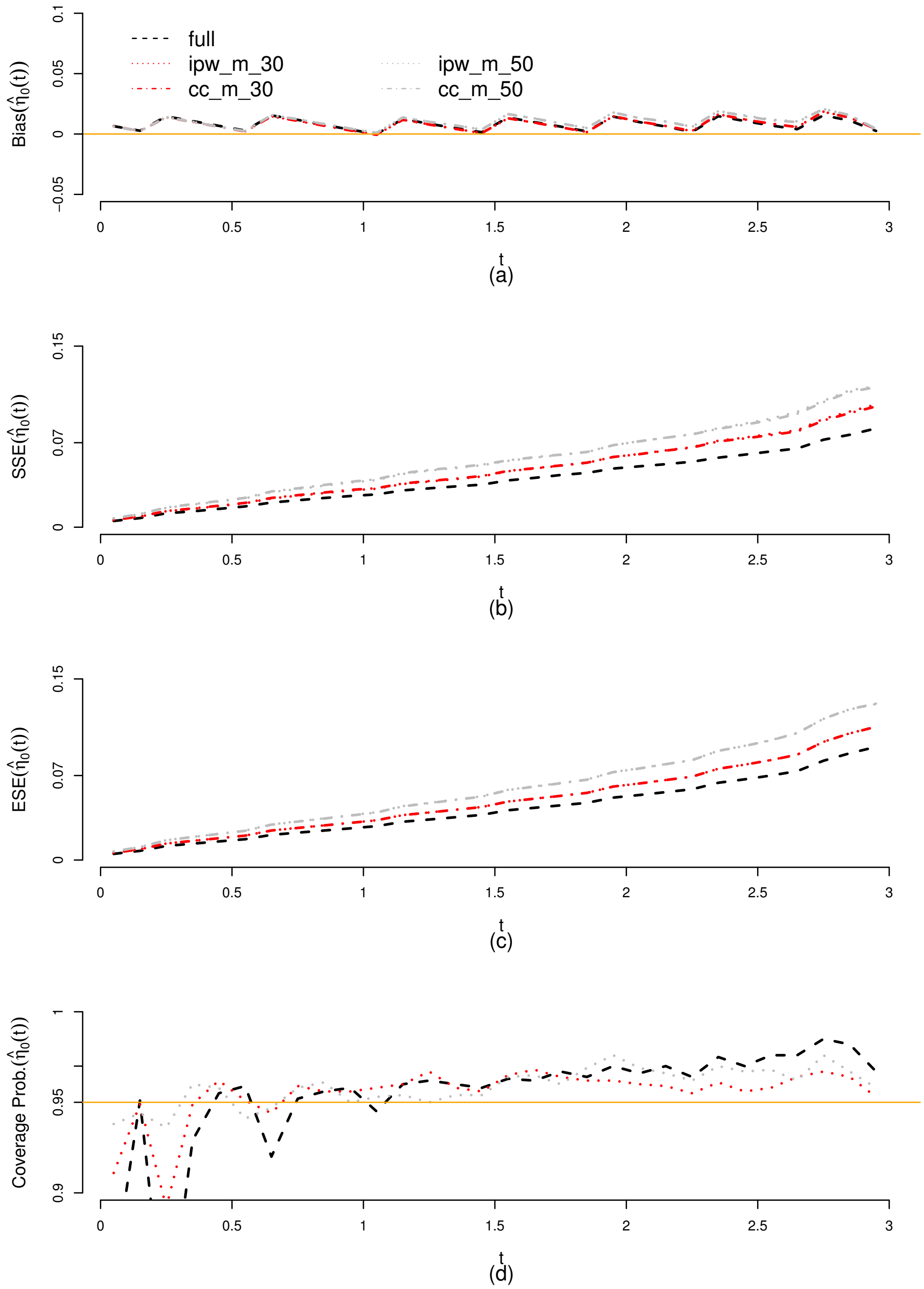


Figure 3: Comparison of Full, IPW, CC estimators for the baseline cumulative coefficient $\eta_0(t)$ under (2.25) with $m_0 = 0.3,\ 0.5$, $n = 700$ and approximately 45% censoring percentage based on 1000 simulations for sampling design III. (a): The plots of the biases of the estimates. (b): The plots of the empirical standard errors of the estimates. (c): The plots of the average of the estimated standard errors of the estimates. (d): The plots of the coverage probabilities of the Full and IPW estimators.

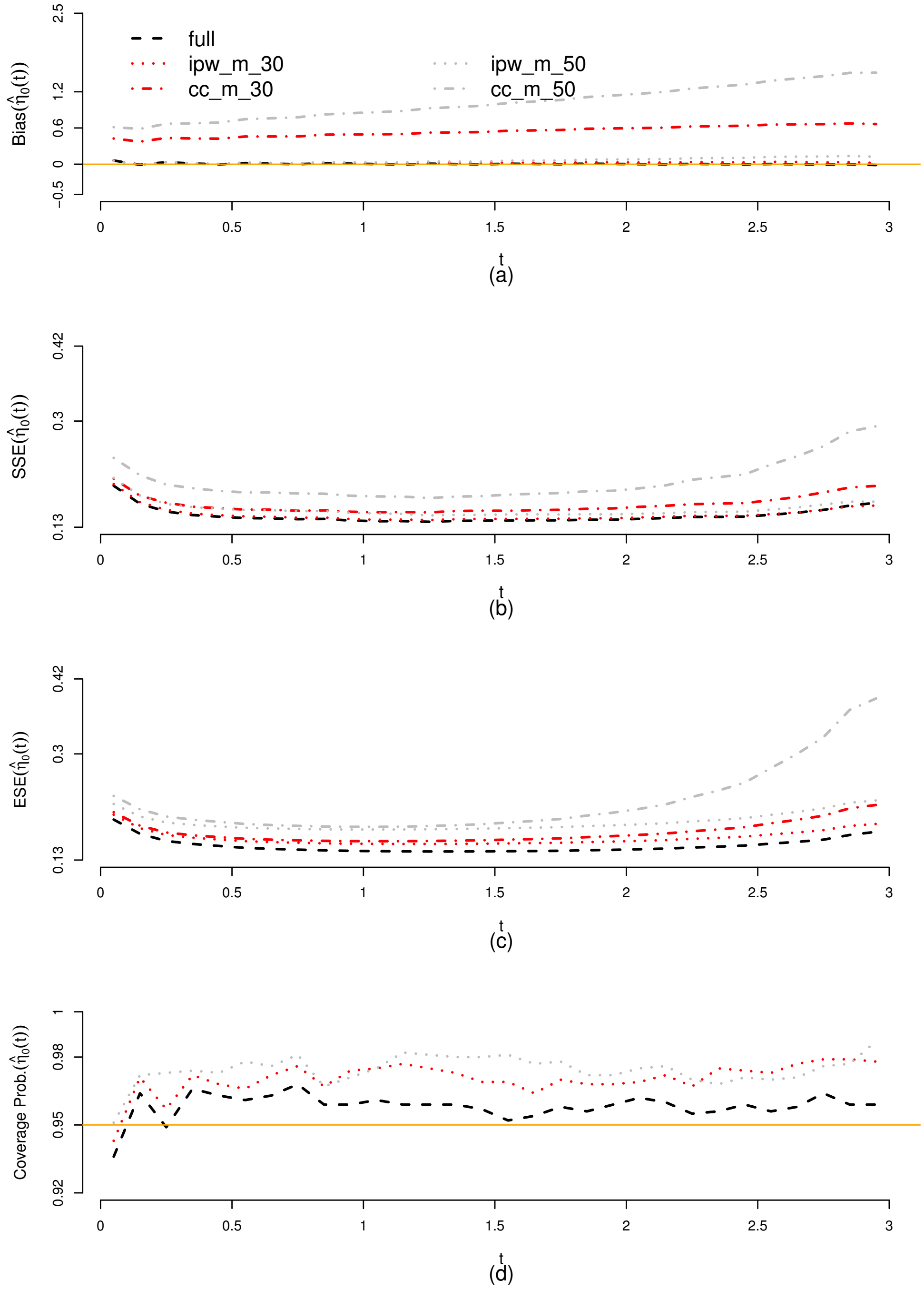


Figure 4: Comparison of Full, IPW, CC estimators for the baseline cumulative coefficient $\eta_0(t)$ under (2.26) with $m_0 = 0.3,\ 0.5$, $n = 700$ and approximately 45% censoring percentage based on 1000 simulations for sampling design I.(a): The plots of the biases of the estimates. (b): The plots of the empirical standard errors of the estimates. (c): The plots of the average of the estimated standard errors of the estimates. (d): The plots of the coverage probabilities of the Full and IPW estimators.

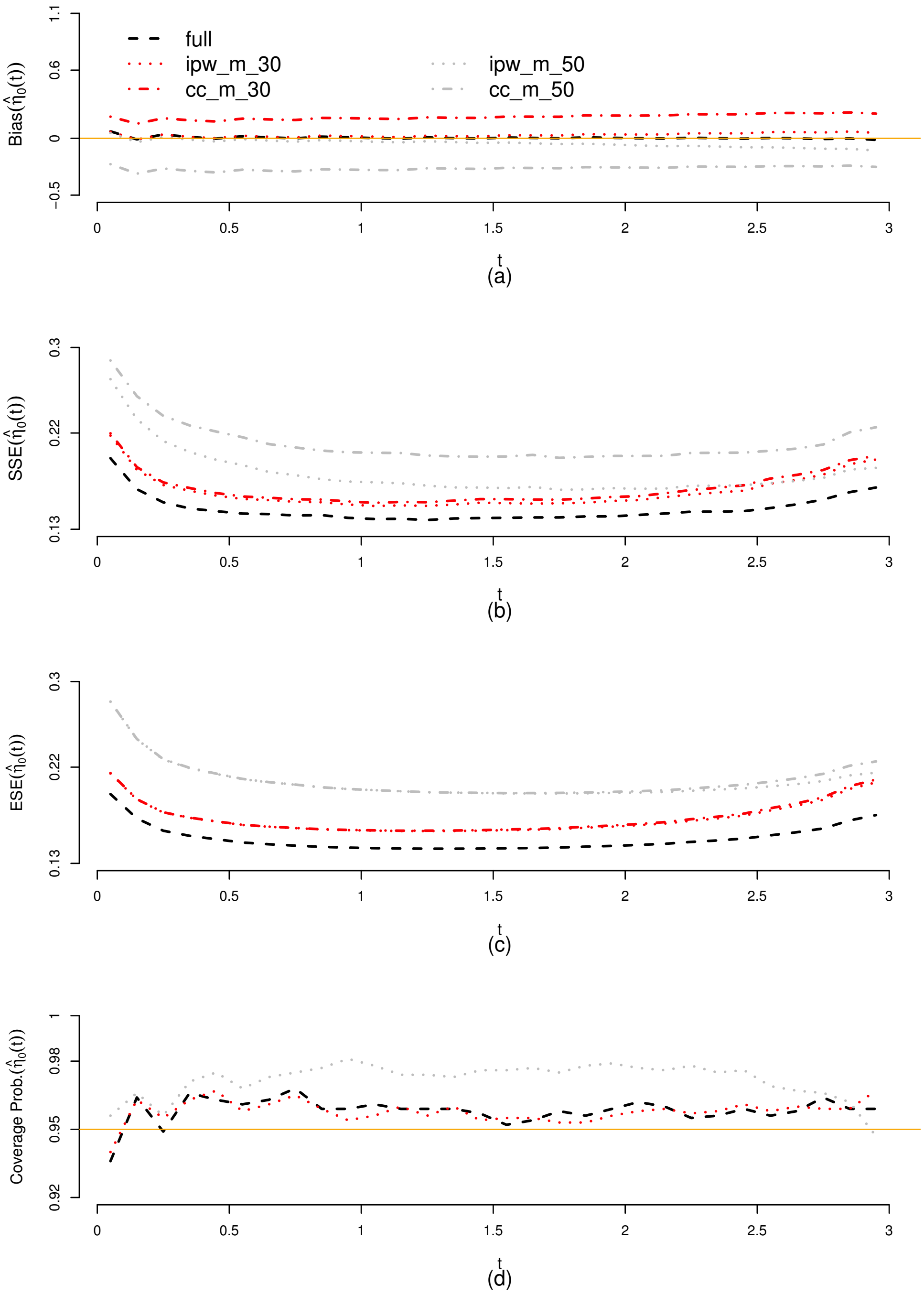


Figure 5: Comparison of Full, IPW, CC estimators for the baseline cumulative coefficient $\eta_0(t)$ under (2.26) with $m_0 = 0.3,\ 0.5,\ n = 700$ and approximately 45% censoring percentage based on 1000 simulations for sampling design II.(a): The plots of the biases of the estimates. (b): The plots of the empirical standard errors of the estimates. (c): The plots of the average of the estimated standard errors of the estimates. (d): The plots of the coverage probabilities of the Full and IPW estimators.

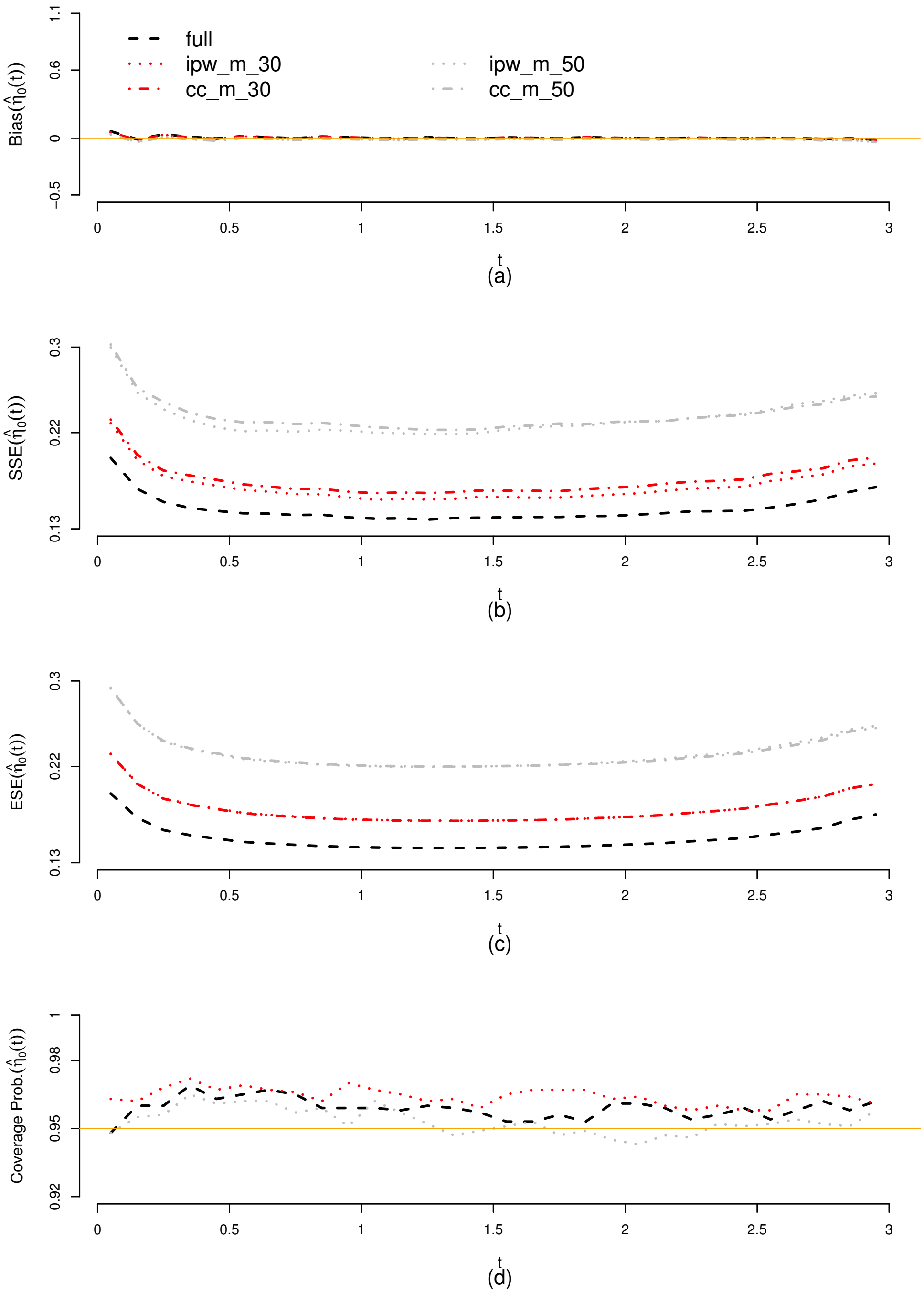


Figure 6: Comparison of Full, IPW, CC estimators for the baseline cumulative coefficient $\eta_0(t)$ under (2.26) with $m_0 = 0.3,\ 0.5$, $n = 700$ and approximately 45% censoring percentage based on 1000 simulations for sampling design III.(a): The plots of the biases of the estimates. (b): The plots of the empirical standard errors of the estimates. (c): The plots of the average of the estimated standard errors of the estimates. (d): The plots of the coverage probabilities of the Full and IPW estimators.

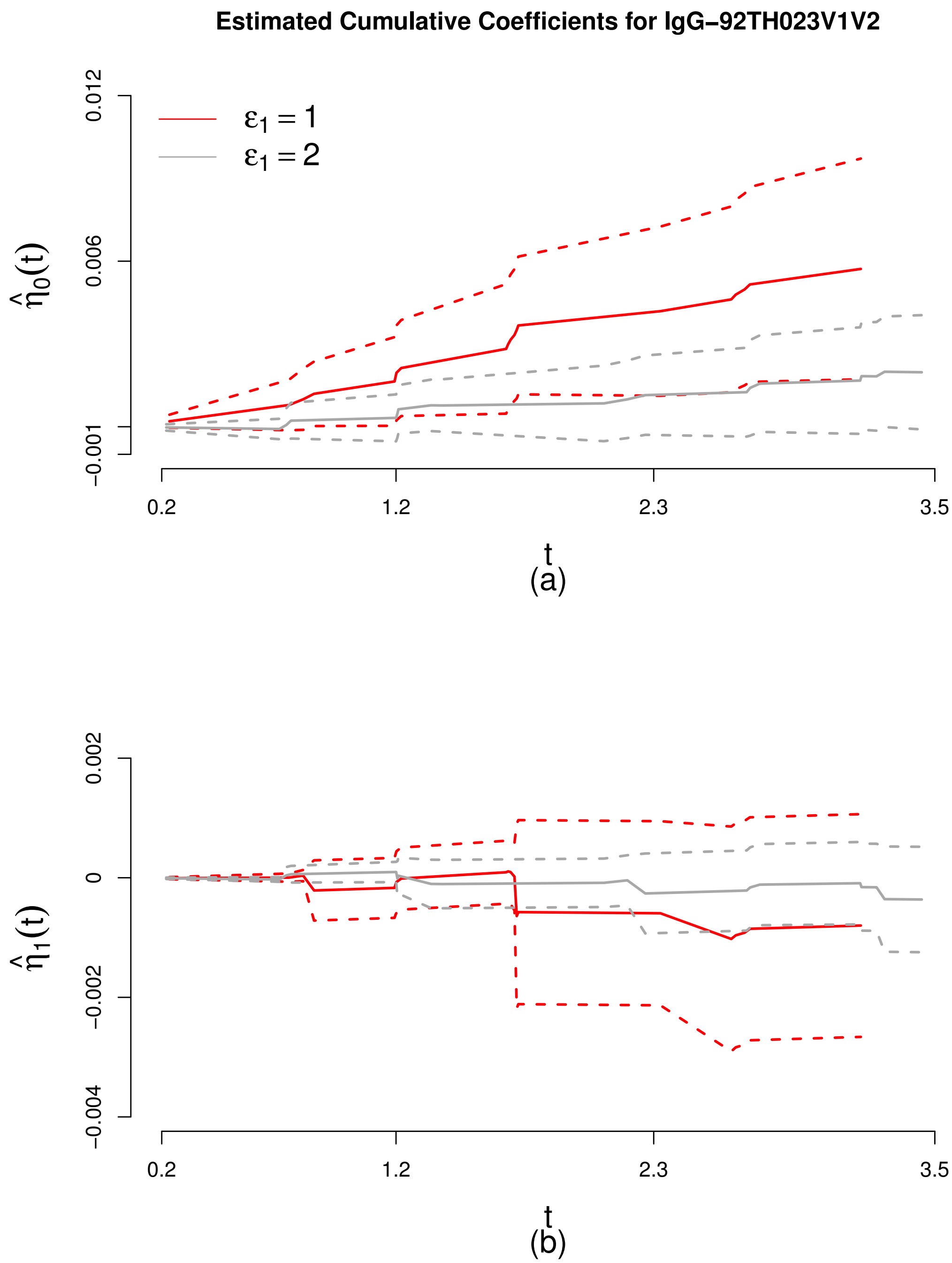


Figure 7: (a) and (b) show the comparison of the IPW estimates of the baseline cumulative coefficients $\eta_0(t)$ and the cumulative coefficients $\eta_1(t)$ with 95% pointwise confidence intervals for the immune response R_i (IgG-92TH023V1V2) in model (2.30) for $\epsilon_{1_i} = 1$ (red) and $\epsilon_{1_i} = 2$ (grey), respectively .

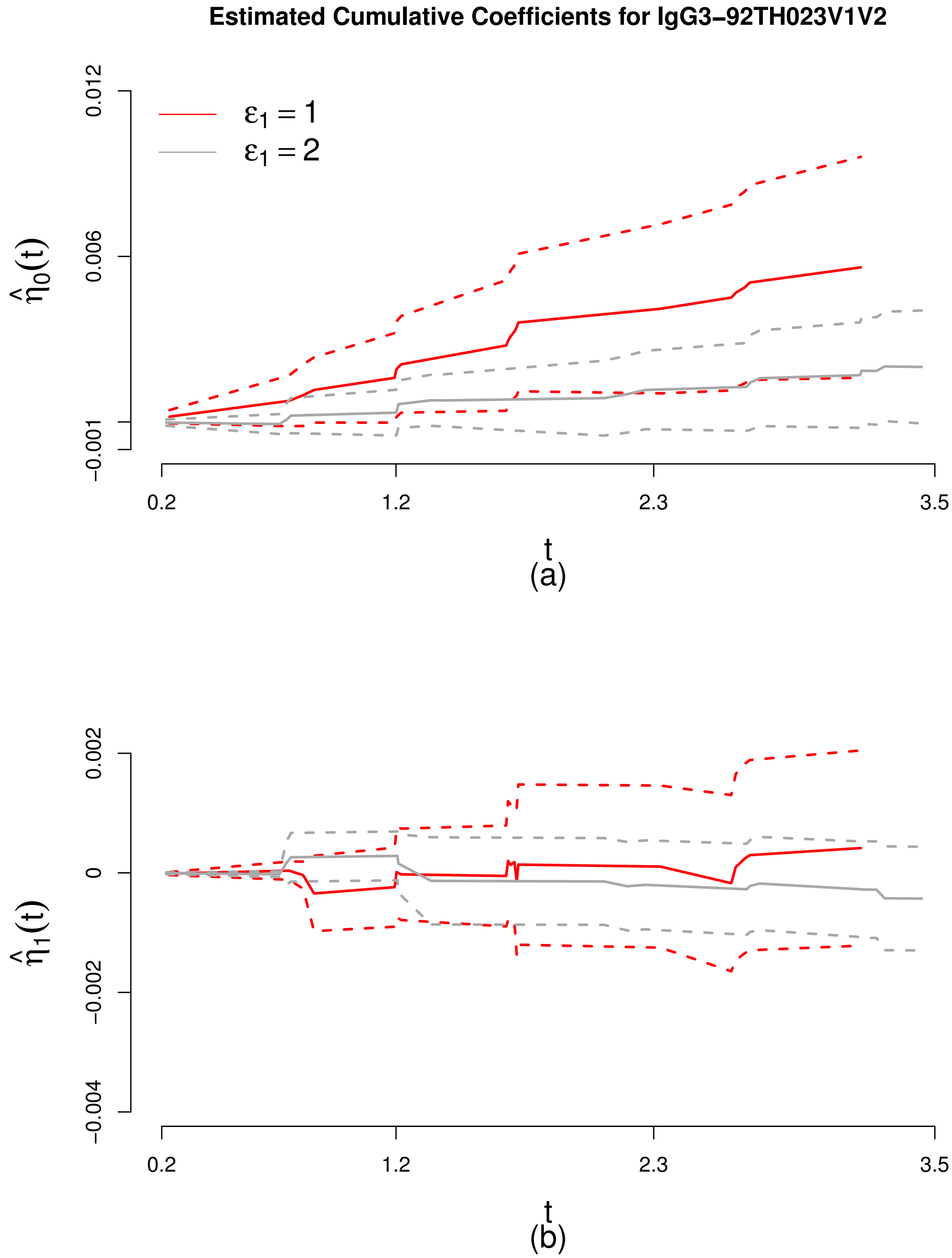


Figure 8: (a) and (b) show the comparison of the IPW estimates of the baseline cumulative coefficients $\eta_0(t)$ and the cumulative coefficients $\eta_1(t)$ with 95% pointwise confidence intervals for the immune response R_i (IgG3-92TH023V1V2) in model (2.30) for $\epsilon_{1_i} = 1$ (red) and $\epsilon_{1_i} = 2$ (grey), respectively.

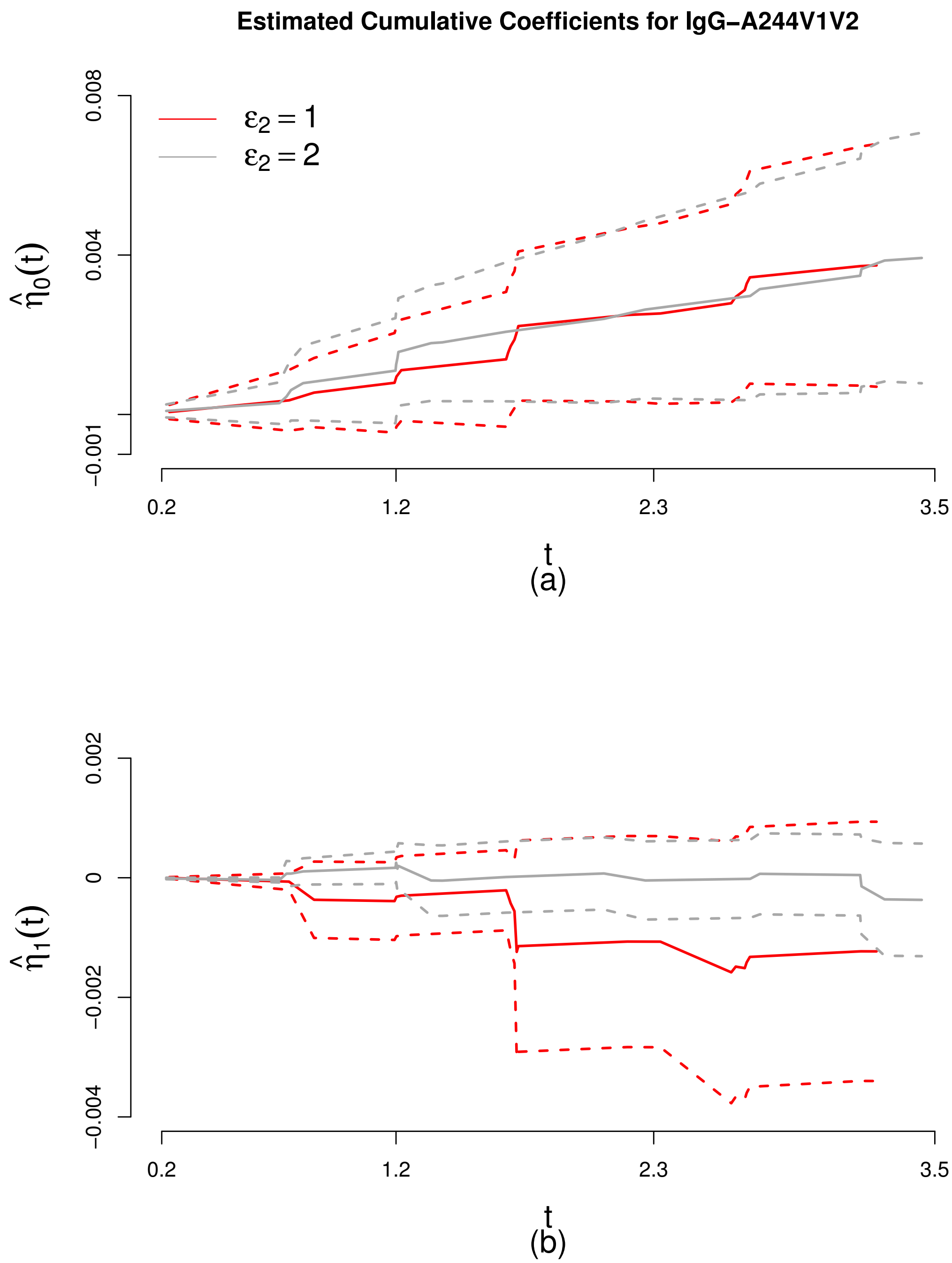


Figure 9: (a) and (b) show the comparison of the IPW estimates of the baseline cumulative coefficients $\eta_0(t)$ and the cumulative coefficients $\eta_1(t)$ with 95% pointwise confidence intervals for the immune response R_i (IgG-A244V1V2) in model (2.30) for $\epsilon_{2_i} = 1$ (red) and $\epsilon_{2_i} = 2$ (grey), respectively.

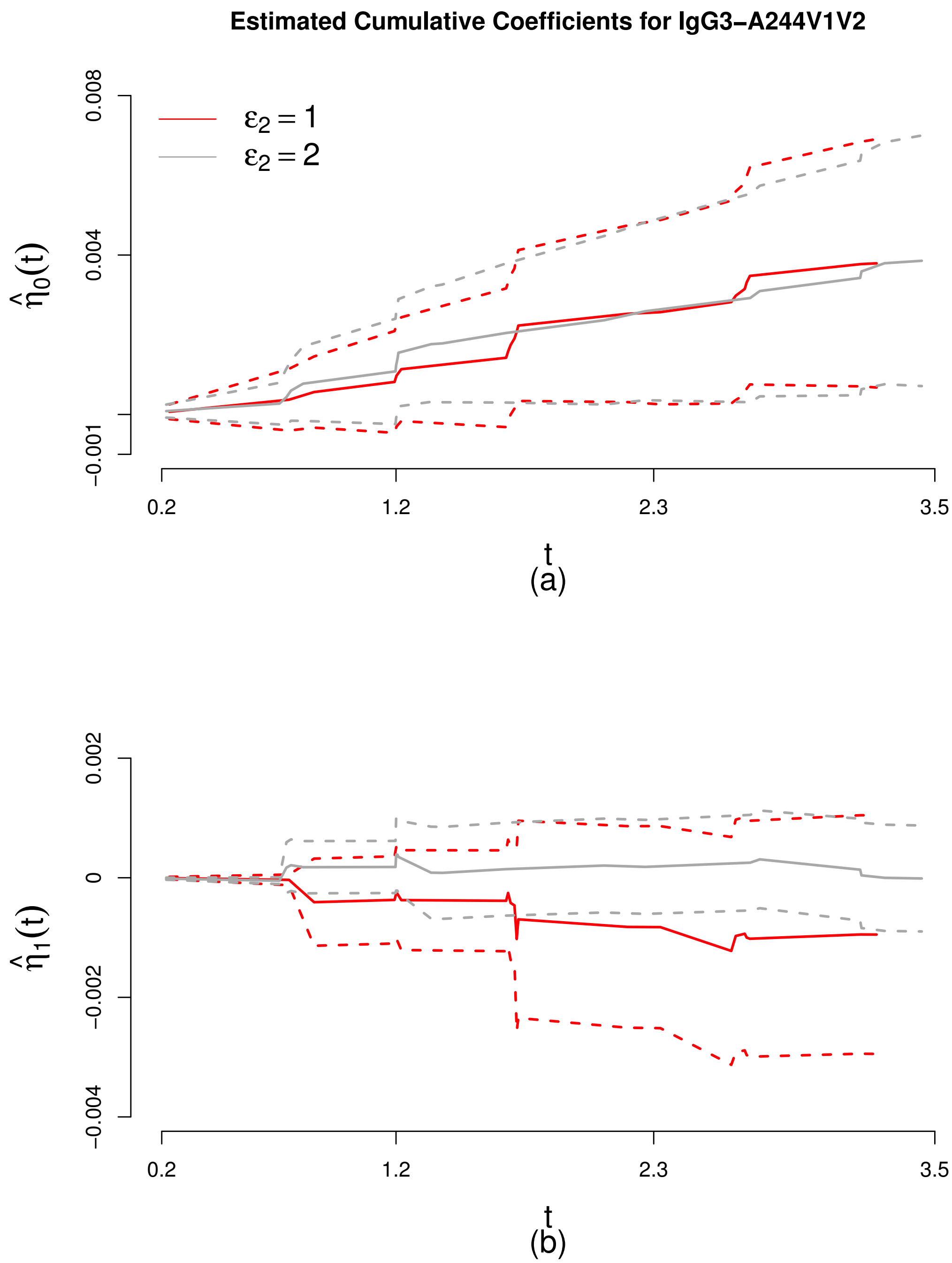


Figure 10: (a) and (b) show the comparison of the IPW estimates of the baseline cumulative coefficients $\eta_0(t)$ and the cumulative coefficients $\eta_1(t)$ with 95% pointwise confidence intervals for the observed immune response R_i (IgG3-A244V1V2) in model (2.30) for $\epsilon_{2_i} = 1$ (red) and $\epsilon_{2_i} = 2$ (grey), respectively.

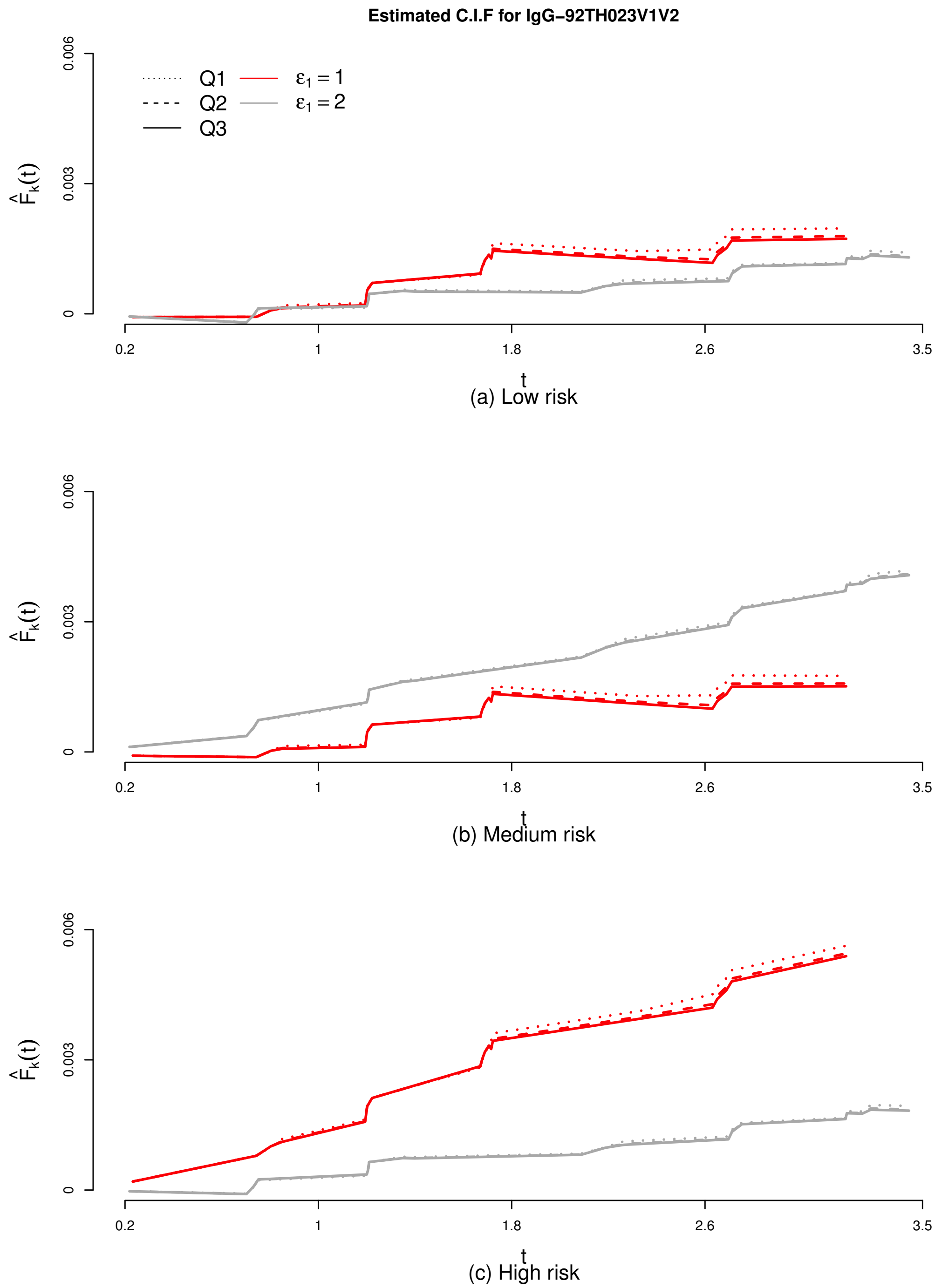


Figure 11: Q_1, Q_2, and Q_3 are the first, second and third quartiles of the observed immune response $R_i = R_{11_i}$ (IgG-92TH023V1V2), where $Q_1 = 0.09027, Q_2 = 0.31310$ and $Q_3 = 0.39230$. (a), (b) and (c) show the estimated cumulative incidence function $\hat{F}_k$ for $\epsilon_{1_i} = 1$ (red) and $\epsilon_{1_i} = 2$ (grey) at each level of behavioral risk score group(low, medium, high), respectively, based on the model (2.30).

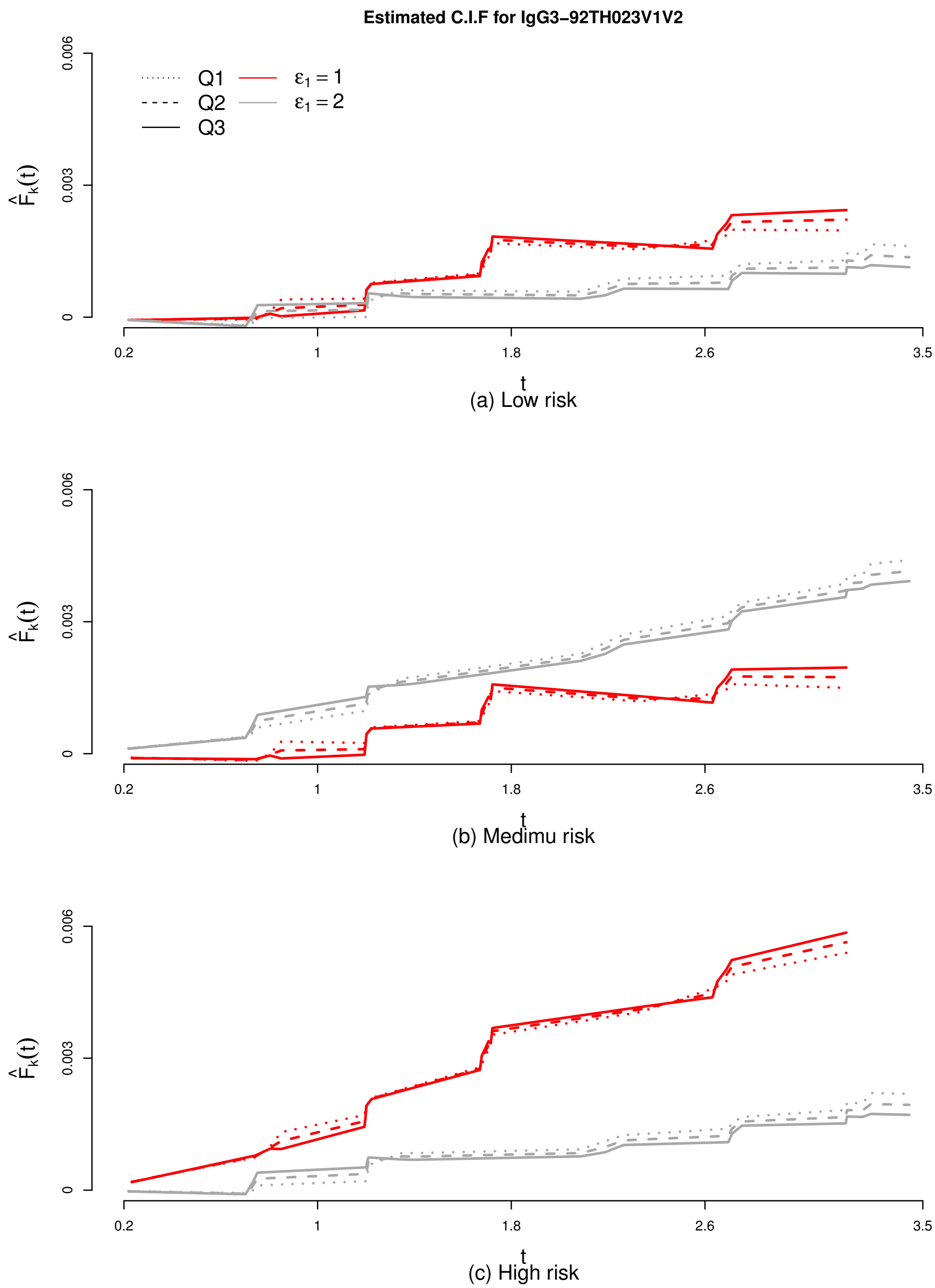


Figure 12: Q_1, Q_2, and Q_3 are the first, second and third quartiles of the observed immune response $R_i = R_{12_i}$ (IgG3-92TH023V1V2), where $Q_1 = -0.4677, Q_2 = 0.1196$ and $Q_3 = 0.6484$. (a), (b) and (c) show the estimated cumulative incidence function $\hat{F}_k$ for $\epsilon_{1_i} = 1$ (red) and $\epsilon_{1_i} = 2$ (grey) at each level of behavioral risk score group(low, medium, high), respectively, based on the model (2.30).

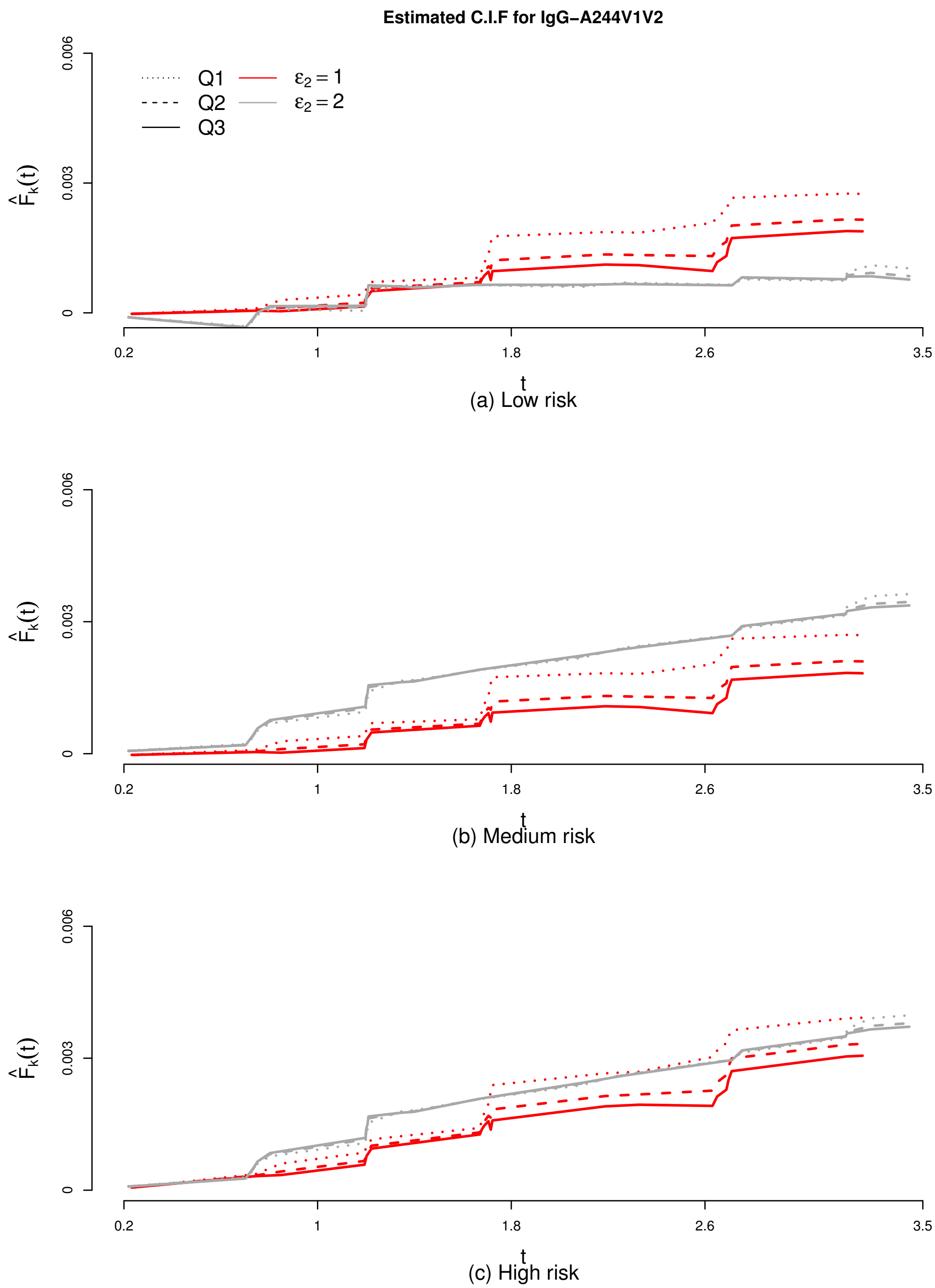


Figure 13: Q_1, Q_2, and Q_3 are the first, second and third quartiles of the observed immune response $R_i = R_{21_i}$ (IgG-A244V1V2), where $Q_1 = -0.1530, Q_2 = 0.3321$ and $Q_3 = 0.5514$. (a), (b) and (c) show the estimated cumulative incidence function $\hat{F}_k$ for $\epsilon_{2_i} = 1$ (red) and $\epsilon_{2_i} = 2$ (grey) at each level of behavioral risk score group(low, medium, high), respectively, based on the model (2.30).

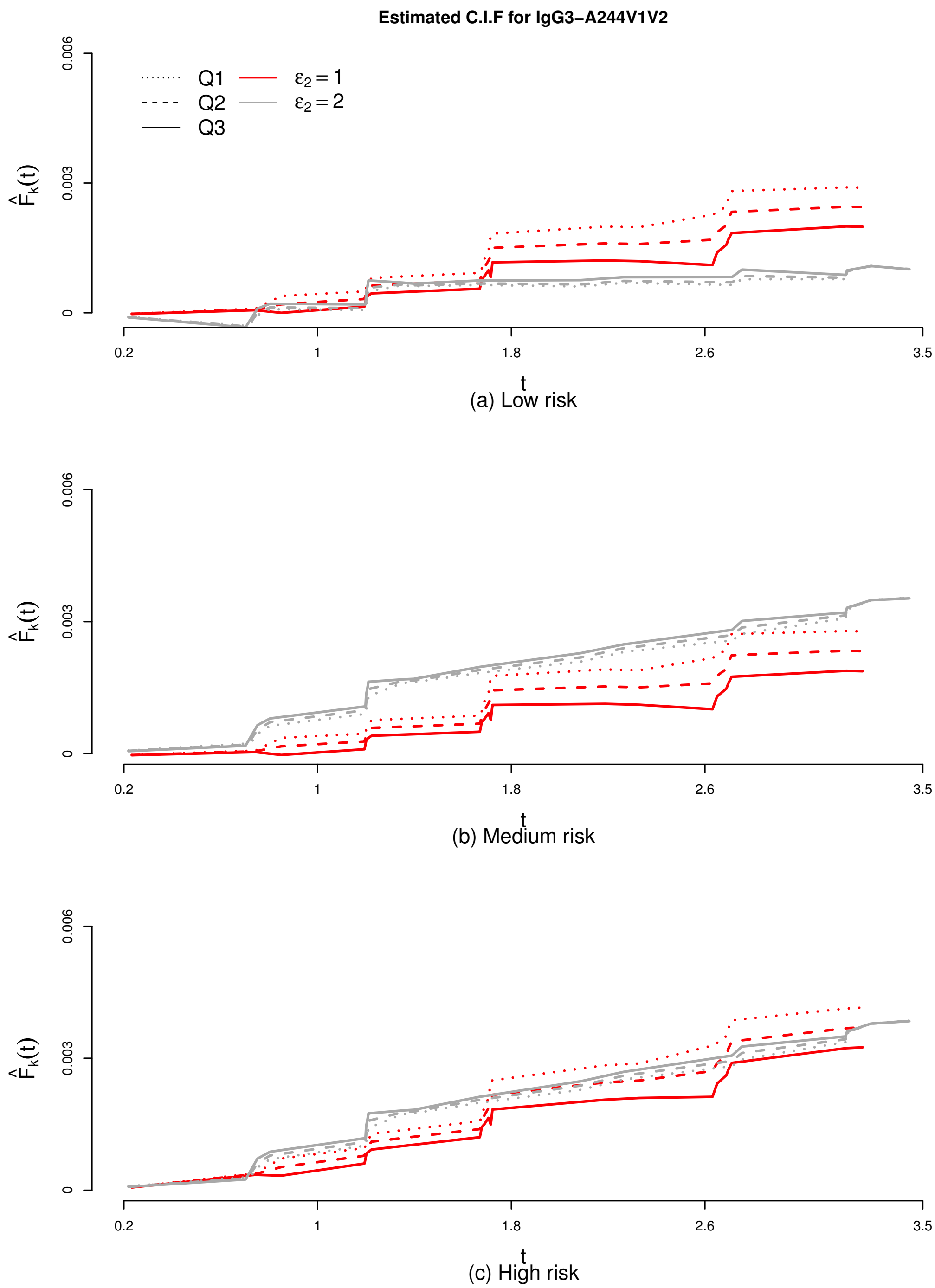


Figure 14: Q_1, Q_2, and Q_3 are the first, second and third quartiles of the observed immune response $R_i = R_{22_i}$ (IgG3-A244V1V2), where $Q_1 = -0.3851, Q_2 = 0.08807$ and $Q_3 = 0.5680$. (a), (b) and (c) show the estimated cumulative incidence function $\hat{F}_k$ for $\epsilon_{2_i} = 1$ (red) and $\epsilon_{2_i} = 2$ (grey) at each level of behavioral risk score group(low, medium, high), respectively, based on the model (2.30).

CHAPTER 3: ANALYSIS A SEMIPARAMETRIC ADDITIVE MODEL WITH MISSING COVARIATE USING AUGMENTED INVERSE PROBABILITY WEIGHTED COMPLETE CASE METHOD

In this chapter, we propose an improved estimating equation by adapting the theory of Robins, Rotnizky and Zhao (1994). In section 3.1, augmented IPW of complete case estimating equations have been developed for a semiparametric additive model with identity link function in model (1.22). The estimation procedure is given to obtain the augmented IPW estimators. In section 3.2, asymptotic results have been investigated. Simulation results for the AIPW estimator have been discussed in section 3.3, showing that this estimator improves efficiency.

3.1 Augmented IPW of Complete Case Estimating Equation for a Semiparametric Additive Model

We assume that the selection probability S_i, the conditional expectations $E\{X_i^{(2)}|\mathcal{V}_i\}$ and $E\{X_i^{(2)}(X_i^{(2)})^\mathsf{T}|\mathcal{V}_i\}$ are known for those with missing covariates $X_i^{(2)}$.

Let

$$\begin{aligned} e_{i,\eta(t)}(t) &= E\left[\boldsymbol{D}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})w_i(t)\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - F_1(t;\boldsymbol{X}_i,\boldsymbol{Z}_i)\right\}|\mathcal{V}_i\right], \\ e_{i,\gamma}(t) &= E\left[\boldsymbol{D}_{\boldsymbol{\gamma},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})w_i(t)\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - F_1(t;\boldsymbol{X}_i,\boldsymbol{Z}_i)\right\}|\mathcal{V}_i\right], \end{aligned}$$

where $\mathcal{V}_i = \{\tilde{T}_i, \Delta_i, \tilde{\epsilon}_i, X_i^{(1)}, Z_i, A_i\}$ is the observed phase-one data. Following the augmentation theory of Robins, Rotnizky and Zhao (1994), we consider the following augmented IPW estimating equations for $(\boldsymbol{\eta}(\cdot), \boldsymbol{\gamma}, \hat{\theta})$:

$$\widetilde{\boldsymbol{U}}_{\boldsymbol{\eta}}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \hat{\theta}) = \sum_{i=1}^{n} \left[\psi_i(\hat{\theta}) \boldsymbol{D}_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}) w_i(t) \left\{ \frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - F_1(t; \boldsymbol{X}_i, \boldsymbol{Z}_i) \right\} + (1 - \psi_i(\hat{\theta})) e_{i,\eta(t)}(t) \right], \tag{3.1}$$

$$\widetilde{\boldsymbol{U}}_{\boldsymbol{\gamma}}(\tau, \boldsymbol{\eta}(\cdot), \boldsymbol{\gamma}, \hat{\theta}) = \sum_{i=1}^{n} \int_0^{\tau} \left[\psi_i(\hat{\theta}) \boldsymbol{D}_{\boldsymbol{\gamma},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}) w_i(t) \left\{ \frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - F_1(t; \boldsymbol{X}_i, \boldsymbol{Z}_i) \right\} + (1 - \psi_i(\hat{\theta})) e_{i,\gamma}(t) \right] dt. \tag{3.2}$$

In equation (3.1), the first part of the contribution, $\psi_i(\hat{\theta}) \boldsymbol{D}_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}) w_i(t)$ $\left\{ \frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - F_1(t; \boldsymbol{X}_i, \boldsymbol{Z}_i) \right\}$, represents the inverse probability weighting of complete case. The second part of the contribution, $(1 - \psi_i(\hat{\theta})) e_{i,\eta(t)}(t)$, is the augmentation to the first part with the knowledge of the conditional expectations $E\{X_i^{(2)} | \mathcal{V}_i\}$ and $E\{X_i^{(2)} (X_i^{(2)})^{\mathsf{T}} | \mathcal{V}_i\}$ for the missing covariates. The contribution from subject i with $\xi_i = 0$ only involves the conditional expectation $e_{i,\eta(t)}(t)$. A similar interpretation applies to the equation (3.2).

The estimating functions given in (3.1) and (3.2) are equivalent to

$$\begin{aligned} \widetilde{\boldsymbol{U}}_{\boldsymbol{\eta}}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \hat{\theta}) = & (\boldsymbol{D}_{\boldsymbol{\eta}}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}))^{\mathsf{T}} \boldsymbol{W}(t) \boldsymbol{\Psi}(\hat{\theta}) \{ \boldsymbol{R}(t) - \boldsymbol{F}_1(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}) \} \\ & + E\Big[(\boldsymbol{D}_{\boldsymbol{\eta}}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}))^{\mathsf{T}} \boldsymbol{W}(t) (I - \boldsymbol{\Psi}(\hat{\theta})) \\ & \{ \boldsymbol{R}(t) - \boldsymbol{F}_1(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}) \} | \boldsymbol{V} \Big], \end{aligned} \tag{3.3}$$

$$\begin{aligned} \widetilde{\boldsymbol{U}}_{\boldsymbol{\gamma}}(\tau, \boldsymbol{\eta}(\cdot), \boldsymbol{\gamma}, \hat{\theta}) = & \int_0^{\tau} (\boldsymbol{D}_{\boldsymbol{\gamma}}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}))^{\mathsf{T}} \boldsymbol{W}(t) \boldsymbol{\Psi}(\hat{\theta}) \{ \boldsymbol{R}(t) - \boldsymbol{F}_1(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}) \} \, dt \\ + & \int_0^{\tau} E\Big[(\boldsymbol{D}_{\boldsymbol{\gamma}}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}))^{\mathsf{T}} \boldsymbol{W}(t) (I - \boldsymbol{\Psi}(\hat{\theta})) \\ & \{ \boldsymbol{R}(t) - \boldsymbol{F}_1(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}) \} | \boldsymbol{V} \Big] \, dt. \end{aligned} \tag{3.4}$$

Consider the following semiparametric additive model by using identity link function $h(x) = x$ in (1.22):

$$F_{1i}(t; X_i, Z_i) = X_i^\mathsf{T}\eta(t) + g(\gamma, Z_i, t). \tag{3.5}$$

Under (3.5), $\boldsymbol{D}_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma})$ is the ith row vector of $X_i^\mathsf{T} = (1, X_{i1}, \cdots, X_{ip})$ and $\boldsymbol{D}_{\gamma,i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma})$is the ith row vector of $\partial g(\gamma, Z_i, t)/\partial\gamma$ where $g(\gamma, Z_i, t)$ is known function.

Note that

$$\begin{aligned} e_{i,\eta(t)}(t) &= E\left[X_i^\mathsf{T}|\mathcal{V}_i\right] w_i(t)\left\{\frac{\Delta_i\, N_i(t)}{\widehat{G}(T_i)} - g(\gamma, Z_i, t)\right\} - \{\eta(t)\}^\mathsf{T} E\left[X_i X_i^\mathsf{T}|\mathcal{V}_i\right] w_i(t), \\ e_{i,\gamma}(t) &= \{\partial g(\gamma, Z_i, t)/\partial\gamma\}^\mathsf{T} w_i(t)\left\{\frac{\Delta_i\, N_i(t)}{\widehat{G}(T_i)} - g(\gamma, Z_i, t) - E\left[X_i^\mathsf{T}|\mathcal{V}_i\right]\eta(t)\right\}. \end{aligned}$$

Let $X = (X_1, \cdots, X_n)^\mathsf{T}$ and $\partial g(\gamma, Z, t)/\partial\gamma$ be the $n \times (p+1)$ and $n \times q$ matrices, respectively. Let $\mathcal{V} = (\mathcal{V}_1, \ldots, \mathcal{V}_n)$, $V_x = (E\{X_1|\mathcal{V}_1\}, \cdots, E\{X_n|\mathcal{V}_n\})^\mathsf{T}$, and $V_{xx}(\theta) = \sum_{i=1}^n (1 - \psi_i(\theta)) w_i(t) E\left[X_i X_i^\mathsf{T}|\mathcal{V}_i\right]$, where

$$\begin{aligned} E\{X_i|\mathcal{V}_i\} &= \begin{pmatrix} X_i^{(1)} \\ \\ E\{X_i^{(2)}|\mathcal{V}_i\} \end{pmatrix}, \\ E\left\{X_i X_i^\mathsf{T}|\mathcal{V}_i\right\} &= \begin{pmatrix} X_i^{(1)}(X_i^{(1)})^\mathsf{T} & X_i^{(1)} E\{(X_i^{(2)})^\mathsf{T}|\mathcal{V}_i\} \\ \\ E\{X_i^{(2)}|\mathcal{V}_i\}(X_i^{(1)})^\mathsf{T} & E\{X_i^{(2)}(X_i^{(2)})^\mathsf{T}|\mathcal{V}_i\} \end{pmatrix} \text{ for each } i. \end{aligned} \tag{3.6}$$

Let

$$a_\eta(t,\eta(t),\gamma,\theta) = V_x^{\mathsf{T}} W(t)(I-\mathbf{\Psi}(\theta))\{\boldsymbol{R}(t)-g(\gamma,Z,t)\} - V_{xx}(\theta)\eta(t), \tag{3.7}$$

$$a_\gamma(\tau,\eta(\cdot),\gamma,\theta) = \int_0^\tau \{\frac{\partial g(\gamma,Z,t)}{\partial\gamma}\}^{\mathsf{T}} \boldsymbol{W}(t)(I-\mathbf{\Psi}(\theta))\{\boldsymbol{R}(t)-V_x\eta(t)-g(\gamma,Z,t)\}\,dt, \tag{3.8}$$

where $V_{xx}(\theta) = E(X^T W(t)(I-\mathbf{\Psi}(\theta))X|\mathcal{V})$.

Under (3.5), the estimating equations are followed by (3.3) and (3.4) that

$$\widetilde{\boldsymbol{U}}_{\boldsymbol{\eta}}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\hat{\theta}) = X^{\mathsf{T}}\boldsymbol{W}(t)\mathbf{\Psi}(\hat{\theta})\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})\} + a_\eta(t,\eta(t),\gamma,\hat{\theta}), \tag{3.9}$$

$$\widetilde{\boldsymbol{U}}_{\boldsymbol{\gamma}}(\tau,\boldsymbol{\eta}(\cdot),\boldsymbol{\gamma},\hat{\theta}) = \int_0^\tau \{\frac{\partial g(\gamma,Z,t)}{\partial\gamma}\}^{\mathsf{T}}\boldsymbol{W}(t)\mathbf{\Psi}(\hat{\theta})\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})\}\,dt + a_\gamma(\tau,\eta(\cdot),\gamma,\hat{\theta}). \tag{3.10}$$

3.1.1 Estimation Procedure

Let $\mu_1(\mathcal{V}_i,\alpha_1)$ and $\mu_2(\mathcal{V}_i,\alpha_2)$ be the parametric models for $E\{X_i^{(2)}|\mathcal{V}_i\}$ and $E\{X_i^{(2)}(X_i^{(2)})^{\mathsf{T}}|\mathcal{V}_i\}$, respectively, where α_1 and α_2 are r_1 and r_2 dimensional vectors of parameters belonging to some compact sets, and $\mu_1(\cdot,\alpha_1)$ and $\mu_2(\cdot,\alpha_2)$ are some smooth functions. For example, $\mu_1(\cdot,\alpha_1)$ and $\mu_2(\cdot,\alpha_2)$ can be approximated by the first order or second order linear functions of the variables in $\mathcal{V}_i$ or their transformations. In this case the parameters α_1 and α_2 can be estimated by $\hat{\alpha}_1$ and $\hat{\alpha}_2$ using the least square regressions of $X_i^{(2)}$ on $\mathcal{V}_i$ and $X_i^{(2)}(X_i^{(2)})^{\mathsf{T}}$ on $\mathcal{V}_i$, respectively, based on the observations with $\xi_i = 1$.

The conditional expectations $E\{X_i^{(2)}|\mathcal{V}_i\}$ and $E\{X_i^{(2)}(X_i^{(2)})^{\mathsf{T}}|\mathcal{V}_i\}$ are replaced with $\mu_1(\mathcal{V}_i,\hat{\alpha}_1)$ and $\mu_2(\mathcal{V}_i,\hat{\alpha}_2)$, respectively, meaning that V_x and $V_{xx}(\hat{\theta})$ are replaced by $\widehat{V}_x$ and $\widehat{V}_{xx}(\hat{\theta})$ in $a_\gamma(\tau, \eta(\cdot), \gamma, \hat{\theta})$ and $a_\gamma(\tau, \eta(\cdot), \gamma, \hat{\theta})$ defined in (3.7) and (3.8), respectively. It follows that the estimators $\hat{a}_\eta(t, \eta(t), \gamma, \hat{\theta})$ and $\hat{a}_\gamma(\tau, \eta(\cdot), \gamma, \hat{\theta})$ of $a_\eta(t, \eta(t), \gamma, \hat{\theta})$ and $a_\gamma(\tau, \eta(\cdot), \gamma, \hat{\theta})$ are

$$\hat{a}_\eta(t, \eta(t), \gamma, \hat{\theta}) = \widehat{V}_x^\top W(t)(I - \mathbf{\Psi}(\hat{\theta}))\left\{\boldsymbol{R}(t) - g(\gamma, Z, t)\right\} - \widehat{V}_{xx}(\hat{\theta})\eta(t), \quad (3.11)$$

$$\hat{a}_\gamma(\tau, \eta(\cdot), \gamma, \hat{\theta}) = \int_0^\tau \{\frac{\partial g(\gamma, Z, t)}{\partial \gamma}\}^\top \boldsymbol{W}(t)(I - \mathbf{\Psi}(\hat{\theta}))\left\{\boldsymbol{R}(t) - \widehat{V}_x\eta(t) - g(\gamma, Z, t)\right\} dt. \quad (3.12)$$

By replacing (3.11) and (3.12) into the score functions (3.9) and (3.10), we obtain the following augmented IPW estimating equation for $\boldsymbol{\eta}(t)$ and $\boldsymbol{\gamma}$:

$$\widehat{\widetilde{\boldsymbol{U}}}_{\boldsymbol{\eta}}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \hat{\theta}) = X^\top \boldsymbol{W}(t)\mathbf{\Psi}(\hat{\theta})\left\{\boldsymbol{R}(t) - \boldsymbol{F}_1(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma})\right\} + \hat{a}_\eta(t, \eta(t), \gamma, \hat{\theta}), \quad (3.13)$$

$$\widehat{\widetilde{\boldsymbol{U}}}_{\boldsymbol{\gamma}}(\tau, \boldsymbol{\eta}(\cdot), \boldsymbol{\gamma}, \hat{\theta}) = \int_0^\tau \{\frac{\partial g(\gamma, Z, t)}{\partial \gamma}\}^\top \boldsymbol{W}(t)\mathbf{\Psi}(\hat{\theta})\left\{\boldsymbol{R}(t) - \boldsymbol{F}_1(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma})\right\} dt + \hat{a}_\gamma(\tau, \eta(\cdot), \gamma, \hat{\theta}). \quad (3.14)$$

The augmented inverse probability weighted of the complete-case estimators $\widehat{\boldsymbol{\eta}}(t)$ and $\widehat{\boldsymbol{\gamma}}$ of $\boldsymbol{\eta}(t)$ and $\boldsymbol{\gamma}$ solve the equation $\widehat{\widetilde{\boldsymbol{U}}}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\gamma}}, \hat{\theta}) = 0$, where $\widehat{\widetilde{\boldsymbol{U}}}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(\boldsymbol{\eta}, \boldsymbol{\gamma}, \hat{\theta}) = \{\widehat{\widetilde{\boldsymbol{U}}}_{\boldsymbol{\eta}}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \hat{\theta}), \widehat{\widetilde{\boldsymbol{U}}}_{\boldsymbol{\gamma}}(\tau, \boldsymbol{\eta}(\cdot), \boldsymbol{\gamma}, \hat{\theta})\}$.

Similar to numerical algorithm in section 2.1.2, the estimating equations (3.13) and (3.14) can be solved by using an iterative algorithm.

[Computational Algorithm] The estimators of $\boldsymbol{\eta}(t)$ and $\boldsymbol{\gamma}$ can be obtained through the following algorithm.

1. Given inverse probability weighting estimators $\boldsymbol{\eta}^{(0)}(t)$ and $\boldsymbol{\gamma}^{(0)}$ as initial values.

2. Estimate V_x and $V_{xx}(\hat{\theta})$ by $\widehat{V}_x$ and $\widehat{V}_{xx}(\hat{\theta})$.

3. Using Taylor expansion of $\boldsymbol{F}_1(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma})$ around the values $(\widehat{\boldsymbol{\eta}}^{(i)}(t), \widehat{\boldsymbol{\gamma}}^{(i)})$ at ith iteration, we have

$$
\begin{aligned}
\boldsymbol{F}_1(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}) &\approx \boldsymbol{F}_1(t, \widehat{\boldsymbol{\eta}}^{(i)}(t), \widehat{\boldsymbol{\gamma}}^{(i)}) + \boldsymbol{D}_{\boldsymbol{\eta}}(t, \widehat{\boldsymbol{\eta}}^{(i)}(t), \widehat{\boldsymbol{\gamma}}^{(i)}) \left\{ \boldsymbol{\eta}(t) - \widehat{\boldsymbol{\eta}}^{(i)}(t) \right\} \\
&+ \boldsymbol{D}_{\boldsymbol{\gamma}}(t, \widehat{\boldsymbol{\eta}}^{(i)}(t), \widehat{\boldsymbol{\gamma}}^{(i)}) \left\{ \boldsymbol{\gamma} - \widehat{\boldsymbol{\gamma}}^{(i)} \right\}. \qquad (3.15)
\end{aligned}
$$

4. Using (3.7) and (3.8), $a_\eta(t, \widehat{\boldsymbol{\eta}}^{(i)}(t), \widehat{\boldsymbol{\gamma}}^{(i)}, \hat{\theta})$ and $a_\gamma(\tau, \widehat{\boldsymbol{\eta}}^{(i)}(\cdot), \widehat{\boldsymbol{\gamma}}^{(i)}, \hat{\theta})$ are estimated by

$$
\begin{aligned}
\hat{a}_\eta(t, \widehat{\boldsymbol{\eta}}^{(i)}(t), \widehat{\boldsymbol{\gamma}}^{(i)}, \hat{\theta}) =& \widehat{V}_x^{\mathsf{T}} W(t)(I - \boldsymbol{\Psi}(\hat{\theta})) \left\{ \boldsymbol{R}(t) - g(\widehat{\boldsymbol{\gamma}}^{(i)}, Z, t) \right\} \\
& - \widehat{V}_{xx}(\hat{\theta}) \widehat{\boldsymbol{\eta}}^{(i)}(t), \qquad (3.16) \\
\hat{a}_\gamma(\tau, \widehat{\boldsymbol{\eta}}^{(i)}(\cdot), \widehat{\boldsymbol{\gamma}}^{(i)}, \hat{\theta}) =& \int_0^\tau \{ \frac{\partial g(\widehat{\boldsymbol{\gamma}}^{(i)}, Z, t)}{\partial \widehat{\boldsymbol{\gamma}}^{(i)}} \}^{\mathsf{T}} \boldsymbol{W}(t)(I - \boldsymbol{\Psi}(\hat{\theta})) \left\{ \boldsymbol{R}(t) - \widehat{V}_x \widehat{\boldsymbol{\eta}}^{(i)}(t) \right. \\
& \left. - g(\widehat{\boldsymbol{\gamma}}^{(i)}, Z, t) \right\} dt. \qquad (3.17)
\end{aligned}
$$

and we denote $\hat{a}_\eta^{(i)}(t, \hat{\theta}) = \hat{a}_\eta(t, \widehat{\boldsymbol{\eta}}^{(i)}(t), \widehat{\boldsymbol{\gamma}}^{(i)}, \hat{\theta})$ and $\hat{a}_\gamma^{(i)}(\hat{\theta}) = \hat{a}_\gamma(\tau, \widehat{\boldsymbol{\eta}}^{(i)}(\cdot), \widehat{\boldsymbol{\gamma}}^{(i)}, \hat{\theta})$.

5. Plugging (3.15), (3.16), and (3.17) into (3.13) and (3.14), respectively, to get the approximate estimating equations

$$
\begin{aligned}
\widehat{\widehat{\boldsymbol{U}}}_{\boldsymbol{\eta}}(t, \widehat{\boldsymbol{\eta}}^{(i)}(t), \widehat{\boldsymbol{\gamma}}^{(i)}, \hat{\theta}) &\approx \{ \boldsymbol{D}_{\boldsymbol{\eta}}^{(i)}(t) \}^{\mathsf{T}} \boldsymbol{W}(t) \boldsymbol{\Psi}(\hat{\theta}) \left[\boldsymbol{R}(t) - \boldsymbol{F}_1^{(i)}(t) \right. \\
& \left. - \boldsymbol{D}_{\boldsymbol{\eta}}^{(i)}(t) \left\{ \boldsymbol{\eta}(t) - \widehat{\boldsymbol{\eta}}^{(i)}(t) \right\} - \boldsymbol{D}_{\boldsymbol{\gamma}}^{(i)}(t) \left\{ \boldsymbol{\gamma} - \widehat{\boldsymbol{\gamma}}^{(i)} \right\} \right] \\
& + \hat{a}_\eta^{(i)}(t, \hat{\theta}) = 0, \qquad (3.18) \\
\widehat{\widehat{\boldsymbol{U}}}_{\boldsymbol{\gamma}}(\tau, \widehat{\boldsymbol{\eta}}^{(i)}(\cdot), \widehat{\boldsymbol{\gamma}}^{(i)}, \hat{\theta}) &\approx \int_0^\tau \{ \boldsymbol{D}_{\boldsymbol{\gamma}}^{(i)}(t) \}^{\mathsf{T}} \boldsymbol{W}(t) \boldsymbol{\Psi}(\hat{\theta}) \left[\boldsymbol{R}(t) - \boldsymbol{F}_1^{(i)}(t) \right. \\
& \left. - \boldsymbol{D}_{\boldsymbol{\eta}}^{(i)}(t) \left\{ \boldsymbol{\eta}(t) - \widehat{\boldsymbol{\eta}}^{(i)}(t) \right\} - \boldsymbol{D}_{\boldsymbol{\gamma}}^{(i)}(t) \left\{ \boldsymbol{\gamma} - \widehat{\boldsymbol{\gamma}}^{(i)} \right\} \right] dt \\
& + \hat{a}_\gamma^{(i)}(\hat{\theta}) = 0, \qquad (3.19)
\end{aligned}
$$

where $\boldsymbol{D}_{\boldsymbol{\eta}}^{(i)}(t) = \boldsymbol{D}_{\boldsymbol{\eta}}^{(i)}(t, \widehat{\boldsymbol{\eta}}^{(i)}(t), \widehat{\boldsymbol{\gamma}}^{(i)})$, $\boldsymbol{D}_{\boldsymbol{\gamma}}^{(i)}(t) = \boldsymbol{D}_{\boldsymbol{\gamma}}^{(i)}(t, \widehat{\boldsymbol{\eta}}^{(i)}(t), \widehat{\boldsymbol{\gamma}}^{(i)})$, and $\boldsymbol{F}_1^{(i)}(t) = \boldsymbol{F}_1(t, \widehat{\boldsymbol{\eta}}^{(i)}(t), \widehat{\boldsymbol{\gamma}}^{(i)})$.

6. Solving equation (3.18) for $\boldsymbol{\eta}(t)$ to get

$$\begin{aligned} \boldsymbol{\eta}(t) &= \widehat{\boldsymbol{\eta}}^{(i)}(t) + \{\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}^{(i)}(t, \hat{\theta})\}^{-1}\{\boldsymbol{D}_{\boldsymbol{\eta}}^{(i)}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\Big\{\boldsymbol{R}(t) - \boldsymbol{F}_1^{(i)}(t) \\ &\quad -\boldsymbol{D}_{\boldsymbol{\gamma}}^{(i)}(t)\Big\{\boldsymbol{\gamma} - \widehat{\boldsymbol{\gamma}}^{(i)}\Big\}\Big\} + \{\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}^{(i)}(t, \hat{\theta})\}^{-1}\hat{a}_{\eta}^{(i)}(t, \hat{\theta}). \end{aligned} \tag{3.20}$$

7. Plugging (3.20) into (3.19) and then solving (3.19) for $\boldsymbol{\gamma} - \widehat{\boldsymbol{\gamma}}^{(i)}$. Then the resulting estimate of $\boldsymbol{\gamma}$ is $\widehat{\boldsymbol{\gamma}}^{(i+1)}$ at $(i+1)$th step estimation. Specially, the $(i+1)$th step estimate for $\boldsymbol{\gamma}$ is

$$\widehat{\boldsymbol{\gamma}}^{(i+1)} = \widehat{\boldsymbol{\gamma}}^{(i)} + \Big\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}^{(i)}(\hat{\theta})\Big\}^{-1}\Big\{\boldsymbol{B}_{\boldsymbol{\gamma}}^{(i)}(\hat{\theta}) + \hat{a}_{\gamma}^{(i)}(\hat{\theta}) - \boldsymbol{A}_{\eta}^{(i)}(\hat{\theta})\Big\}, \tag{3.21}$$

where

$$\begin{aligned} \boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}^{(i)}(\hat{\theta}) &= \int_0^{\tau}\{\boldsymbol{D}_{\boldsymbol{\gamma}}^{(i)}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\boldsymbol{H}^{(i)}(t, \hat{\theta})\boldsymbol{D}_{\boldsymbol{\gamma}}^{(i)}(t)\, dt, \\ \boldsymbol{B}_{\boldsymbol{\gamma}}^{(i)}(\hat{\theta}) &= \int_0^{\tau}\{\boldsymbol{D}_{\boldsymbol{\gamma}}^{(i)}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\boldsymbol{H}^{(i)}(t, \hat{\theta})\Big\{\boldsymbol{R}(t) - \boldsymbol{F}_1^{(i)}(t)\Big\}\, dt, \\ \boldsymbol{H}^{(i)}(t, \hat{\theta}) &= I - \boldsymbol{D}_{\boldsymbol{\eta}}^{(i)}(t)\Big[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}^{(i)}(t, \hat{\theta})\Big]^{-1}\{\boldsymbol{D}_{\boldsymbol{\eta}}^{(i)}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta}), \\ \boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}^{(i)}(t, \hat{\theta}) &= \{\boldsymbol{D}_{\boldsymbol{\eta}}^{(i)}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\eta}}^{(i)}(t), \\ \boldsymbol{A}_{\eta}^{(i)}(\hat{\theta}) &= \int_0^{\tau}\boldsymbol{K}^{(i)}(t, \hat{\theta})\hat{a}_{\eta}^{(i)}(t, \hat{\theta})\, dt, \\ \boldsymbol{K}^{(i)}(t, \hat{\theta}) &= \{\boldsymbol{D}_{\boldsymbol{\gamma}}^{(i)}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\eta}}^{(i)}(t)\Big[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}^{(i)}(t, \hat{\theta})\Big]^{-1}. \end{aligned} \tag{3.22}$$

8. The estimate of $\boldsymbol{\eta}(t)$ at $(i+1)$th iteration is obtained by plugging $\widehat{\boldsymbol{\gamma}}^{(i+1)}$ into

(3.20). Then the $(i+1)$th step estimator for $\boldsymbol{\eta}(t)$ is

$$\begin{aligned} \widehat{\boldsymbol{\eta}}^{(i+1)}(t) &= \widehat{\boldsymbol{\eta}}^{(i)}(t) + \{\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}^{(i)}(t,\hat{\theta})\}^{-1}\{\boldsymbol{D}_{\boldsymbol{\eta}}^{(i)}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\Big\{\boldsymbol{R}(t)-\boldsymbol{F}_1^{(i)}(t) \\ &\quad -\boldsymbol{D}_{\boldsymbol{\gamma}}^{(i)}(t)\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}^{(i)}(\hat{\theta})\}^{-1}\Big\{\boldsymbol{B}_{\boldsymbol{\gamma}}^{(i)}(\hat{\theta}) + \hat{a}_{\gamma}^{(i)}(\hat{\theta}) - \boldsymbol{A}_{\eta}^{(i)}(\hat{\theta})\Big\}\Big\} \\ &\quad +\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}^{(i)}(t,\hat{\theta})\}^{-1}\hat{a}_{\eta}^{(i)}(t,\hat{\theta}). \end{aligned} \tag{3.23}$$

9. Repeat steps 7 and 8 until $\widehat{\boldsymbol{\gamma}}$ converges. We use the criteria of $||\widehat{\boldsymbol{\gamma}}^{(i+1)} - \widehat{\boldsymbol{\gamma}}^{(i)}|| < 10^{-4}$.

3.2 Asymptotic Properties

We derive the expressions for the proposed AIPW estimators and study asymptotic results for those estimators.

Theorem 3.1. *Assume that the models for the selection probability* $P(\xi_i = 1|\mathcal{V}_i)$ *and both the conditional expectations* $E\{X_i^{(2)}|\mathcal{V}_i\}$ *and* $E\{X_i^{(2)}(X_i^{(2)})^{\mathsf{T}}|\mathcal{V}_i\}$ *of the phase-two covariates are correctly specified. The estimators of* $\boldsymbol{\gamma}$ *and* $\boldsymbol{\eta}(t)$ *obtained by solving equations (3.13) and (3.14) have the following expressions:*

$$\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 = \left\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0)\right\}^{-1}\left\{\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0) + \tilde{a}_{\gamma}(\theta_0) - \boldsymbol{A}_{\eta}(\theta_0)\right\} + o_p(n^{-\frac{1}{2}}), \tag{3.24}$$

$$\begin{aligned} \widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t) &= \{\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\}^{-1}\{\boldsymbol{D}_{\boldsymbol{\eta}}(t)\}^{T}\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\left\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t)-\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0)\}^{-1}\right. \\ &\quad \left.\left\{\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0) + \tilde{a}_{\gamma}(\theta_0) - \boldsymbol{A}_{\eta}(\theta_0)\right\}\right\} + \{\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\}^{-1}\tilde{a}_{\eta}(t,\theta_0) \\ &\quad + o_p(n^{-\frac{1}{2}}), \end{aligned} \tag{3.25}$$

where

$$
\begin{aligned}
\boldsymbol{A}_{\eta}(\theta) &= \int_0^{\tau} \boldsymbol{K}(t,\theta)\tilde{a}_{\eta}(t,\theta)\ dt, \\
\boldsymbol{K}(t,\theta) &= \boldsymbol{D}_{\boldsymbol{\gamma}}^{T}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta)\boldsymbol{D}_{\boldsymbol{\eta}}(t)\left[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta)\right]^{-1}, \\
\tilde{a}_{\eta}(t,\theta) &= V_x^T W(t)(I-\boldsymbol{\Psi}(\theta))\left\{\boldsymbol{R}(t)-g(\hat{\gamma},Z,t)\right\}-V_{xx}(\theta)\hat{\eta}(t), \\
\tilde{a}_{\gamma}(\theta) &= \int_0^{\tau}\{\frac{\partial g(\gamma_0,Z,t)}{\partial\gamma_0}\}^{\mathsf{T}}\boldsymbol{W}(t)(I-\boldsymbol{\Psi}(\hat{\theta}))\left\{\boldsymbol{R}(t)-V_x\hat{\eta}(t)-g(\hat{\gamma},Z,t)\right\}\ dt,
\end{aligned}
\tag{3.26}
$$

and where $\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta)$, $\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta)$ *and* $\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta)$ *are defined in (2.12).*

Proof of Theorem 3.1 is given in section 4.1.

Let

$$
\begin{aligned}
\boldsymbol{q}^*_{\boldsymbol{\gamma}}(s,t,\boldsymbol{\gamma},\theta) &= n^{-1}\sum_{i=1}^{n}\{\frac{\partial g(\gamma,Z_i,t)}{\partial\gamma}\}^{\mathsf{T}}w_i(t)(1-\psi_i(\theta))\frac{\Delta_i N_i(t)}{G(T_i)}I(s\le\tilde{T}_i\le t), \\
y(t) &= \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\mathcal{I}(\widetilde{T}_i\ge s), \quad \text{uniformly} \quad t\in[0,\tau], \\
M_j{}^c(t) &= \mathcal{I}(\widetilde{T}_j\le t,\Delta_j=0)-\int_0^t\mathcal{I}(\widetilde{T}_i\ge s)d(-\log G(s)), \\
\boldsymbol{\kappa}^*_{\boldsymbol{\gamma},i}(t,\boldsymbol{\gamma},\theta) &= \int_0^{\tau}\frac{\boldsymbol{q}^*_{\boldsymbol{\gamma}}(s,t,\boldsymbol{\gamma},\theta)}{y(s)}dM_i^c(s), \\
\boldsymbol{\zeta}^*_{\boldsymbol{\gamma},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta) &= \sum_{i=1}^{n}\int_0^{\tau}\{\frac{\partial g(\gamma,Z_i,t)}{\partial\gamma}\}^{\mathsf{T}}w_i(t)(1-\psi_i(\theta))\left\{\frac{\Delta_i N_i(t)}{G(T_i)}-V_{x,i}\eta(t)\right. \\
&\quad \left.-g(\gamma,Z_i,t)\right\}, \\
\boldsymbol{q}^*_{\boldsymbol{\eta}}(s,t,\theta) &= n^{-1}\sum_{i=1}^{n}V_{x,i}^{\mathsf{T}}w_i(t)(1-\psi_i(\theta))\frac{\Delta_i N_i(t)}{G(T_i)}\mathcal{I}(s\le\widetilde{T}_i\le t), \\
\boldsymbol{\kappa}^*_{\boldsymbol{\eta},i}(t,\theta) &= \int_0^{\tau}\frac{\boldsymbol{q}^*_{\boldsymbol{\eta}}(s,t,\theta)}{y(s)}dM_i{}^c(s), \\
\boldsymbol{\zeta}^*_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta) &= n^{-1}\sum_{i=1}^{n}\left\{V_{x,i}^{\mathsf{T}}w_i(t)(1-\psi_i(\theta))\left\{\frac{\Delta_i N_i(t)}{G(T_i)}-g_i(\gamma,Z_i,t)\right\}\right. \\
&\quad \left.-V_{xx,i}\eta(t)\right\}.
\end{aligned}
$$

Theorem 3.2. *Under Condition I in Chapter 4.1, if the selection probability* $P(\xi_i = 1|\mathcal{V}_i)$ *or both the conditional expectations* $E\{X_i^{(2)}|\mathcal{V}_i\}$ *and* $E\{X_i^{(2)}(X_i^{(2)})^\mathsf{T}|\mathcal{V}_i\}$ *are correctly specified, then*

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{d} N(0, \Sigma^*_{\boldsymbol{\gamma}}),$$

where covariance matrix $\Sigma^*_{\boldsymbol{\gamma}} = \boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0)^{-1}E\{\boldsymbol{W}^*_{\boldsymbol{\gamma},i}(\tau, \boldsymbol{\eta}_0(\cdot), \boldsymbol{\gamma}_0, \theta_0)\}^{\otimes 2}\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0)^{-1}$*, where* $\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta)$ *is defined in Theorem 2.1 and where*

$$\begin{aligned}
\boldsymbol{W}^*_{\boldsymbol{\gamma},i}(\tau, \boldsymbol{\eta}(\cdot), \boldsymbol{\gamma}, \theta) \;\; = \;\; & \int_0^\tau \boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t, \theta))\, dt + \int_0^\tau \boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t, \theta)\, dt - \int_0^\tau \boldsymbol{\kappa}^*_{\boldsymbol{\gamma},i}(t, \boldsymbol{\gamma}, \theta) dt \\
& + \int_0^\tau \boldsymbol{\zeta}^*_{\boldsymbol{\gamma},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \theta)\, dt \\
& - \int_0^\tau \boldsymbol{k}(t, \theta) \left\{ \boldsymbol{\kappa}^*_{\boldsymbol{\eta},i}(t, \theta) - \boldsymbol{\zeta}^*_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \theta) \right\} dt. \qquad (3.27)
\end{aligned}$$

The asymptotic covariance matrix of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ *can be consistently estimated by*

$$\widehat{\boldsymbol{\Sigma}}^*_{\boldsymbol{\gamma}} = \widehat{\boldsymbol{Q}}^{-1}_{\boldsymbol{\gamma}}(\hat{\theta})\; n^{-1} \sum_{i=1}^n \left\{ \widehat{\boldsymbol{W}}^*_{\boldsymbol{\gamma},i}(\tau, \widehat{\boldsymbol{\eta}}(\cdot), \widehat{\boldsymbol{\gamma}}, \hat{\theta}) \right\}^{\otimes 2} \widehat{\boldsymbol{Q}}^{-1}_{\boldsymbol{\gamma}}(\hat{\theta}),$$

where

$$\begin{aligned}
\widehat{\boldsymbol{W}}^*_{\boldsymbol{\gamma},i}(\tau, \boldsymbol{\eta}(\cdot), \boldsymbol{\gamma}, \theta) \;\; = \;\; & \int_0^\tau \widehat{\boldsymbol{\zeta}}_{\boldsymbol{\gamma},i}(t, \theta)\, dt + \int_0^\tau \widehat{\boldsymbol{\kappa}}_{\boldsymbol{\gamma},i}(t, \theta)\, dt - \int_0^\tau \widehat{\boldsymbol{\kappa}}^*_{\boldsymbol{\gamma},i}(t, \boldsymbol{\gamma}, \theta) dt \\
& + \int_0^\tau \widehat{\boldsymbol{\zeta}}^*_{\boldsymbol{\gamma},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \theta)\, dt \\
& - \int_0^\tau \widehat{\boldsymbol{K}}(t, \theta) \left\{ \widehat{\boldsymbol{\kappa}}^*_{\boldsymbol{\eta},i}(t, \theta) - \widehat{\boldsymbol{\zeta}}^*_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \theta) \right\} dt,
\end{aligned}$$

where $\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}(\theta)$, $\widehat{\boldsymbol{\zeta}}_{\boldsymbol{\gamma},i}(t, \theta)$, $\widehat{\boldsymbol{\kappa}}_{\boldsymbol{\gamma},i}(t, \theta)$ *are described in Theorem 2.1, and where* $\widehat{\boldsymbol{\kappa}}^*_{\boldsymbol{\gamma},i}(t, \boldsymbol{\gamma}, \theta)$, $\widehat{\boldsymbol{\zeta}}^*_{\boldsymbol{\gamma},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \theta)$, $\widehat{\boldsymbol{\kappa}}^*_{\boldsymbol{\eta},i}(t, \theta)$ *and* $\widehat{\boldsymbol{\zeta}}^*_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \theta)$ *are the estimators of* $\boldsymbol{\kappa}^*_{\boldsymbol{\gamma},i}(t, \boldsymbol{\gamma}, \theta)$, $\boldsymbol{\zeta}^*_{\boldsymbol{\gamma},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \theta)$, $\boldsymbol{\kappa}^*_{\boldsymbol{\eta},i}(t, \theta)$ *and* $\boldsymbol{\zeta}^*_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \theta)$*. With similar arguments to Theorem 2.1, those estimators can be obtained by using definition in (4.78) in section*

4.2 and by replacing V_x *and* V_{xx} *with* $\widehat{V}_x$ *and* $\widehat{V}_{xx}$*, where the unknown conditional expectations* $E(X_i^{(2)}|\mathcal{V}_i)$ *and* $E(X_i^{(2)}(X_i^{(2)})^\mathsf{T}|\mathcal{V}_i)$ *of* $E(X_i|\mathcal{V}_i)$ *and* $E(X_i(X_i)^\mathsf{T}|\mathcal{V}_i)$ *can be obtained by* $\mu_1(\mathcal{V}_i, \hat{\alpha}_1)$ *and* $\mu_2(\mathcal{V}_i, \hat{\alpha}_2)$*, respectively.*

Theorem 3.3. *Under Condition I in section 4.1, if the selection probability* $P(\xi_i = 1|\mathcal{V}_i)$ *or both the conditional expectations* $E\{X_i^{(2)}|\mathcal{V}_i\}$ *and* $E\{X_i^{(2)}(X_i^{(2)})^\mathsf{T}|\mathcal{V}_i\}$ *are correctly specified, then*

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t)) = \{\boldsymbol{Q}_{\boldsymbol{\eta}}(t, \theta_0)\}^{-1}\; n^{-\frac{1}{2}} \sum_{i=1}^{n} \boldsymbol{W}^*_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}_0(t), \boldsymbol{\gamma}_0, \theta_0) + o_p(1). \quad (3.28)$$

where $\boldsymbol{Q}_{\boldsymbol{\eta}}(t,\theta)$ *is defined in Theorem 2.2, and where*

$$\begin{aligned} \boldsymbol{W}^*_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \theta) \;=\;& \boldsymbol{\zeta}_{\boldsymbol{\eta},i}(t,\theta) + \boldsymbol{\kappa}_{\boldsymbol{\eta},i}(t,\theta) \\ & - \boldsymbol{Q}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta) \left\{\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta)\right\}^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} \boldsymbol{W}^*_{\boldsymbol{\gamma},i}(\tau, \boldsymbol{\eta}(\cdot), \boldsymbol{\gamma}, \theta) \\ & + \boldsymbol{\kappa}^*_{\boldsymbol{\eta},i}(t,\theta) - \boldsymbol{\zeta}^*_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \theta). \end{aligned} \quad (3.29)$$

By using lemma 1 of Sun and Wu (2005), $n^{\frac{1}{2}}(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t))$ *converges weakly to a mean-zero Gaussian process on* $t \in [0, \tau]$ *with the covariance matrix*

$$\boldsymbol{\Sigma}^*_{\boldsymbol{\eta}} \;=\; \boldsymbol{Q}_{\boldsymbol{\eta}}(t,\theta_0)^{-1} E\{\boldsymbol{W}^*_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}_0(t), \boldsymbol{\gamma}_0, \theta_0)\}^{\otimes 2} \boldsymbol{Q}_{\boldsymbol{\eta}}(t,\theta_0)^{-1},$$

which can be consistently estimated by

$$\widehat{\boldsymbol{\Sigma}}^*_{\boldsymbol{\eta}} \;=\; \widehat{\boldsymbol{Q}}_{\boldsymbol{\eta}}^{-1}(t,\hat{\theta})\, n^{-1} \sum_{i=1}^{n} \left\{\widehat{\boldsymbol{W}}^*_{\boldsymbol{\eta},i}(t, \widehat{\boldsymbol{\eta}}(t), \widehat{\gamma}, \hat{\theta})\right\}^{\otimes 2} \widehat{\boldsymbol{Q}}_{\boldsymbol{\eta}}^{-1}(t,\hat{\theta}),$$

where

$$\begin{aligned}\widehat{\boldsymbol{W}}^{*}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta) &= \widehat{\boldsymbol{\zeta}}_{\boldsymbol{\eta},i}(t,\theta)+\widehat{\boldsymbol{\kappa}}_{\boldsymbol{\eta},i}(t,\theta)\\ &\quad -\widehat{\boldsymbol{Q}}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta)\left\{\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}(\theta)\right\}^{-1} n^{-\frac{1}{2}}\sum_{i=1}^{n}\widehat{\boldsymbol{W}}^{*}_{\boldsymbol{\gamma},i}(\tau,\boldsymbol{\eta}(\cdot),\boldsymbol{\gamma},\theta)\\ &\quad +\widehat{\boldsymbol{\kappa}}^{*}_{\boldsymbol{\eta},i}(t,\theta)-\widehat{\boldsymbol{\zeta}}^{*}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta),\end{aligned}$$

where $\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}(\theta)$ *is defined in Theorem 2.1.* $\widehat{\boldsymbol{Q}}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta)$, $\widehat{\boldsymbol{Q}}_{\boldsymbol{\eta}}(t,\theta)$, $\widehat{\boldsymbol{\zeta}}_{\boldsymbol{\eta},i}(t,\theta)$, *and* $\widehat{\boldsymbol{\kappa}}_{\boldsymbol{\eta},i}(t,\theta)$ *can be obtained as described in Theorem 2.2.* $\widehat{\boldsymbol{W}}^{*}_{\boldsymbol{\gamma},i}(\tau,\boldsymbol{\eta}(\cdot),\boldsymbol{\gamma},\theta)$ *is defined in Theorem 3.2.* $\widehat{\boldsymbol{\kappa}}^{*}_{\boldsymbol{\eta},i}(t,\theta)$ *and* $\widehat{\boldsymbol{\zeta}}^{*}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta)$ *are the estimators of* $\boldsymbol{\kappa}^{*}_{\boldsymbol{\eta},i}(t,\theta)$ *and* $\boldsymbol{\zeta}^{*}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta)$, *respectively, which are defined in (4.84) in section 4.2.*

3.3 Simulations

In this chapter, a simulation study has been conducted to evaluate the finite sample properties of the augmented inverse probability weighted methods. Let X_i be a Bernoulli random variable with $P(X_i = 1) = 0.6$ and Z_i be a Bernoulli random variable with $P(Z_i = 1|X_i) = 0.4X_i + 0.2$ for a subject i, $i = 1, 2, \ldots, n$. The covariate X_i can be missing and the covariate Z_i is always observed. Let $\epsilon_i = k$ be the types of failure and let $k = 1$ be the event of interest among two competing risks $k \in \{1, 2\}$. From model (1.22), we consider the following semi-parametric additive model with identity link function $h(x) = x$ for the cumulative incidence function with cause 1 :

$$F_{1i}(t; X_i, Z_i) = \eta_0(t) + \eta_1(t)X_i + \gamma Z_i t, \tag{3.30}$$

where $\eta_0(t) = 0.01t$, $\eta_1(t) = 0.03t$ and $\gamma = 0.1$ with $0 \le t \le \tau$ and $\tau = 3$.

We consider the auxiliary covariate A_i, which may give information on the missing covariate X_i. The correlation coefficient ρ can be obtained from the relationship $A_i = \alpha_1 X_i + \alpha_2$ with parameters α_1 and α_2. We consider the correlation coefficients $\rho = 0.5, 0.8$ and 0.9 given by the choice of $(\alpha_1,\ \alpha_2) = (0.5, 0.3), (0.8, 0.12)$ and $(0.92, 0.05)$. Given $\eta_0(t), \eta_1(t), \gamma, X_i, Z_i$, the conditional probability of failure for cause 1 is

$$F_{1i}(\tau) = \eta_0(\tau) + x_i^{\mathsf{T}}\eta_1(\tau) + \gamma^{\mathsf{T}} z_i \tau \quad \textit{where} \quad 0 < t \leq 3$$

for a subject i. The types of failure $\epsilon_i = k$ have been determined by generating a Bernoulli random variable with the probability $F_{1i}(\tau) = P(\epsilon_i = 1)$. The failure time T_i is generated by the conditional probability for cause 1:

$$\tilde{F}_{1i}(t) = P(T_1 \leq t | \epsilon_i = 1) = \frac{F_{1i}(t)}{P(\epsilon_i = 1)} = \frac{F_{1i}(t; x, z, \eta, \gamma)}{F_{1i}(\tau)} = \frac{F_{1i}(t)}{F_{1i}(\tau)},$$

for a subject i and $\tau = 3$.

Let C^* follow a uniform distribution on [0,3]. The censoring time C_i is generated by $C_i = \min(C_i^*, \tau)$. Let $\tilde{T}_i = \min(T_i, C_i)$ be the observed failure time, generating approximately 50% censoring of the failure time before $\tau = 3$. Let $\tilde{\epsilon}_i = \epsilon_i \Delta_i$, where $\Delta_i = I(T_i \leq C_i)$.

Let $\mathcal{V}_i = \{\tilde{T}_i, \Delta_i, \tilde{\epsilon}_i, Z_i, A_i\}$ be the phase one data for a subject i. Let X_i be the phase-two covariate, which can be missing. We consider three simulation scenarios I, II and III in terms of whether the missing probabilities depend on the outcome variables $\tilde{\epsilon}_i$ and how the phase-two covariate X_i is missing. The first two scenarios are called phase-two sampling designs. The first scenario I is a classical case cohort sampling design, where the phase-two covariate X_i is sampled for all cases $\tilde{\epsilon}_i = 1$ and the information of the covariate X_i will be missing for the non-cases $\tilde{\epsilon}_i = 0$ or

2. The second scenario II is a generalized case-cohort sampling design, which allows the phase-two covariate X_i to be missing for both cases and non-cases. In the third scenario III, the missing probability does not depend on $\tilde{\epsilon}_i$ and phase-two covariate X_i is a simple random sample from the phase-one covariates.

Let m_0 be the average of the total missing probabilities. We consider $m_0 = 0.3$ and 0.6 for each sampling scenario. Let m_1 and m_2 be the average missing probabilities for the cases and the non-cases, respectively.

First, we consider $m_0 = 0.3$ and $\rho = 0.5, 0.8, 0.9$ for each scenario. For scenario I, missing probability $\vartheta_{1i} = P(\xi_i = 0|\mathcal{V}_i, \tilde{\epsilon}_i = 1) = 0$ for the cases. For non-cases $\epsilon_i = 0$ or 2, we assume that the missing probability $\vartheta_{2i} = P(\xi_i = 0|\mathcal{V}_i, \tilde{\epsilon}_i \neq 1)$ follows the logistic regression model

$$\text{logit}(\varphi(\mathcal{V}_i, \theta_2)) = \theta_{20} + \theta_{21} A_i + \theta_{22} \tilde{T}_i + \theta_{23} Z_i + \theta_{24} I(\tilde{\epsilon}_i = 2) \qquad (3.31)$$

based on phase-one covariates. We have approximately $m_2 = 0.36$ by choosing $\theta_2 = (-1.5, 0.3, 0.4, 0.3, 0.5)$. We have the linear model with the observed non-case covariates A_i, Z_i and $log(\tilde{T}_i)$ to estimate $E\{X_i|V_i, \tilde{\epsilon}_i \neq 1\}$ and $E\{X_i{}^T X_i|V_i, \tilde{\epsilon}_i \neq 1\}$ for those with missing X_i:

$$E\{X_i|V_i, \tilde{\epsilon}_i \neq 1\} = \nu_{20} A_i + \nu_{21} Z_i + \nu_{22} log(\tilde{T}_i) + \nu_{23} I(\tilde{\epsilon}_i = 2), \qquad (3.32)$$

$$E\{X_i{}^T X_i|V_i, \tilde{\epsilon}_i \neq 1\} = \phi_{20} A_i + \phi_{21} Z_i + \phi_{22} log(\tilde{T}_i) + \phi_{23} I(\tilde{\epsilon}_i = 2), \qquad (3.33)$$

where estimators of coefficients $\hat{\nu}$ and $\hat{\phi}$ can be obtained by a fitting linear model based on the observed response variable X_i and the predictors A_i, Z_i and $log(\tilde{T}_i)$ that are observed non-cases.

For II, the missing probability $\vartheta_{ki} = P(\xi_i = 0|\mathcal{V}_i, \tilde{\epsilon}_i = k)$ for both cases and non-cases can be obtained by the following logistic regression model

$$\text{logit}(\varphi(\mathcal{V}_i, \theta_k)) = \theta_{k0} + \theta_{k1}A_i + \theta_{k2}\tilde{T}_i + \theta_{k3}Z_i + \theta_{k4}I(\tilde{\epsilon}_i = 1) + \theta_{k5}I(\tilde{\epsilon}_i = 2), \quad (3.34)$$

which gives $m_1 = 0.2$ and $m_2 = 0.3$ when $\theta_k = (-1.2, 0.1, 0.1, 0.1, -0.5, 0.5)$ for $k = 1, 2$. We use the following linear models to estimate $E\{X_i|V_i, \tilde{\epsilon}_i = k\}$ and $E\{{X_i}^T X_i|V_i, \tilde{\epsilon}_i = k\}$:

$$E\{X_i|V_i, \tilde{\epsilon}_i = k\} = \nu_{k0}A_i + \nu_{k1}Z_i + \nu_{k2}log(\tilde{T}_i) + \nu_{k3}I(\tilde{\epsilon}_i = 1) + \nu_{k4}I(\tilde{\epsilon}_i = 2), \quad (3.35)$$

$$E\{{X_i}^T X_i|V_i, \tilde{\epsilon}_i = k\} = \phi_{k0}A_i + \phi_{k1}Z_i + \phi_{k2}log(\tilde{T}_i) + \phi_{k3}I(\tilde{\epsilon}_i = 1) + \phi_{k4}I(\tilde{\epsilon}_i = 2) \quad (3.36)$$

for those missing X_i based on the observations that are cases and non-cases and with observed value of X_i.

For III, we use the following logistic model for the missing probability $\vartheta_i = P(\xi_i = 0|\mathcal{V}_i)$:

$$\text{logit}(\varphi(\mathcal{V}_i, \theta)) = \theta_{10} + \theta_{11}A_i + \theta_{12}Z_i. \quad (3.37)$$

We have $\vartheta_i = 0.3$ with $\theta = (-0.5, -0.6, 0.2)$ in (3.37). Therefore, we have $m_0 = 0.3$. To estimate conditional expectations, we use linear models $E\{X_i|V_i\} = \nu_{10} + \nu_{11}A_i + \nu_{12}Z_i + \nu_{13}log(\tilde{T}_i)$ and $E\{{X_i}^T X_i|V_i\} = \phi_{10} + \phi_{11}A_i + \phi_{12}Z_i + \phi_{13}log(\tilde{T}_i)$ for those with missing X_i.

Similarly, we consider $m_0 = 0.6$ and $\rho = 0.5, 0.8, 0.9$ for each scenario. For I, we have $m_1 = 0$ and $m_2 = 0.65$ by choosing $\theta_2 = (-1.5, 0.6, 0.6, 0.8, 2.5)$ in (3.31). Similar to (3.32) and (3.33), conditional expectations for those missing X_i can be estimated by $E\{X_i|V_i, \tilde{\epsilon}_i \neq 1\}$ and $E\{X_i{}^T X_i|V_i, \tilde{\epsilon}_i \neq 1\}$. For scenario II, we have about $m_1 = 0.45$ for the cases and $m_2 = 0.60$ for the non-cases by choosing $\theta_k = (-0.5, 0.3, 0.3, 0.4, -0.5, 1.0)$ in (3.34). Conditional expectations $E\{X_i|V_i, \tilde{\epsilon}_i = k\}$ and $E\{X_i{}^T X_i|V_i, \tilde{\epsilon}_i = k\}$ can be estimated by (3.35) and (3.36), respectively. For scenario III, $\vartheta_i = 0.6$ can be obtained by choosing $\theta = (-0.1, 0.5, 0.5)$ in (3.37). Similarly, $E\{X_i|V_i\}$ and $E\{X_i{}^T X_i|V_i\}$ can be estimated by using linear models with predictors A_i, Z_i and $log(\tilde{T_i})$.

We denote the full estimators as Full when all the values of the phase-two covariate X_i are fully observed, inverse probability weighted estimators as IPW obtained by the estimating procedure in chapter 2.1.2, and complete-case estimators as CC obtained by removing subjects, who have missing covariate X_i. The simulation results for the proposed AIPW estimators of γ and $\eta(t)$, where $t \in [0, 3]$, are summarized by the bias (Bias), the empirical standard error (SSE), the average of the estimated standard error (ESE), the empirical coverage probability (CP) of 95% confidence interval and the relative efficiency (REE), which is defined by SSE of the Full estimator divided by SSE of the each of estimators. We take sample size $n = 600, 700, 900$ and consider the total missing probabilities $m_0 = 0.3$ and 0.6 with different average missing probabilities (m_1, m_2) for the cases and the non-cases for each sampling scenarios I, II and III. Each entry of the tables is estimated based on 1000 simulations runs.

Tables 5-7 consider the average of total missing probability $m_0 = 0.3$ and the correlation coefficient $\rho = 0.5,\ 0.8,\ 0.9$ under three scenarios I, II and III. Table 5 summarizes the Bias, SSE, ESE, and CP for the AIPW estimator $\hat{\gamma}$. This table shows the AIPW estimator $\hat{\gamma}$ with $\rho = 0.5, 0.8, 0.9$ performs well in each scenario. The biases are very small for each sample size. The empirical standard errors decrease as the sample size increases and the averages of the estimated standard errors are very close to the empirical standard errors. The coverage probabilities are close to the 0.95 nominal level.

Table 6 compares the Bias, SSE, ESE, and CP for the AIPW estimator $\hat{\gamma}$ to those for IPW and CC estimators $\hat{\gamma}$. We use the Full estimator as a gold standard. The biases of AIPW, IPW estimators are very small as if all the covariates of X are fully observed. The CC estimator has much larger biases than the AIPW and IPW estimators have in scenarios II and III. Those two scenarios have biased sampling from the study population. Thus, applying standard methods for randomly sampled data results in biased estimation as seen in the CC estimator. However, the CC estimator has smaller biases in sampling scenario III because missingness of phase-two covariate X_i does not depend on outcome variable $\tilde{\epsilon}_i$ and the phase-two covariate X_i is a simple random sample from the phase-one covariates. The ESE for each estimator agrees to the SSE for the corresponding estimator, having a tendency to decrease as the sample size increases. The empirical coverage probabilities of the AIPW and IPW estimators are close to 0.95 nominal level for all scenarios. The empirical coverage probability of the CC estimator is close to 0.95 nominal level only in scenario III.

Table 7 shows the comparison of the relative efficiencies of AIPW, IPW and CC estimators for γ. The efficiency of AIPW estimator is better than that of IPW estimator, which is in turn better than that of the CC estimator in both scenarios II and III. However, the REE of the IPW estimator is similar to that of the CC estimator in scenario III.

Tables 8-10 show the similar results with $m_0 = 0.6$ and the correlation coefficient $\rho = 0.5,\ 0.8,\ 0.9$ for all scenarios I, II and III. Similar arguments can be applied to Tables 8-10.

The efficiency of the AIPW estimator gains the most over the IPW estimator when the average of the total missing probabilities is 0.6 than 0.3. For example, $m_0 = 0.3$ with $\rho = 0.5$ in table 7 can be compare to $m_0 = 0.6$ with $\rho = 0.5$ in table 10. This shows the AIPW estimator can more efficiently utilize information on other fully observable covariates for individuals with missing covariates.

Let AIPW-50, AIPW-80 and AIPW-90 estimators be the AIPW estimators corresponding to the correlation coefficients $\rho = 0.5,\ 0.8,\ 0.9$, respectively. Figures 15, 17 and 19 compare the Full, IPW, CC, AIPW-50, AIPW-80 and AIPW-90 estimators for the baseline cumulative coefficient $\eta_0(t)$ with $m_0 = 0.3$ and 0.6. Figures 16, 18 and 20 compare the Full, IPW, CC, AIPW-50, AIPW-80 and AIPW-90 estimators for the cumulative coefficient $\eta_1(t)$ with $m_0 = 0.3$ and 0.6. We take $n = 700$ with 50% censoring for scenarios I, II and III.

Figures 15-16 compare those estimators for $m_0 = 0.3$ and 0.6 under scenario I with $m_1 = 0$ for cases and $m_2 = 0.3$ for non-cases. In each Figure, (a) and (b) plot the biases of each of the estimators, respectively, (c) and (d) plot the empirical standard errors, (e) and (f) plot the average of the estimated standard errors of each of the estimators, respectively, and (g) and (h) plot the coverage probabilities of the those estimators, respectively, for $\eta_0(t)$ and $\eta_1(t)$. The biases of AIPW-50, AIPW-80 and AIPW-90 estimators are very small comparable to the bias of the Full estimator for both $\eta_1(t)$ and $\eta_0(t)$. The bias of the IPW estimator is relatively smaller than that of the CC estimator for both $\eta_1(t)$ and $\eta_0(t)$, which have much larger biases than the AIPW and IPW estimators. The averages of estimated standard errors of $\hat{\eta}_1(t)$ and $\hat{\eta}_0(t)$ have good agreements with the empirical standard errors. The SSE of the AIPW-50 estimator is slightly smaller than the SSE of the IPW estimator when the correlation between the auxiliary variable A_i and the phase-two covariate X_i is low with $\rho = 0.5$. However, the empirical standard errors of the AIPW-80 and AIPW-90 estimators are much smaller than that of the IPW estimator. This is because the auxiliary variable A_i carries more information on the phase-two covariate X_i with the correlation coefficients $\rho = 0.8$ and 0.9. This phenomenon is more obvious when $m_0 = 0.6$ than 0.3. The coverage probabilities for the AIPW-50, AIPW-80 and AIPW-90 estimators are close to the 0.95 nominal level.

Figures 17 and 18 show the performances of those estimators under scenario II and Figures 19 and 20 show the performances of those estimators under III. These results can be interpreted in a similar way to the scenario I. However, in scenario III in Figures 19 and 20, (a) and (b) show that all estimators have very small biases. (c) and (d) show the SSE of AIPW-50, IPW and CC estimators are similar, but SSE of AIPW-80 and AIPW-90 are much smaller than that of IPW. (g) and (h) show the coverage probabilities of all the estimators are close to the 0.95 nominal level.

3.4 Application

The RV144 vaccine efficacy trial randomized 16,394 HIV negative volunteers to the vaccine ($n = 8198$) and placebo ($n = 8196$) groups. We apply the proposed estimating procedures for AIPW method to the vaccine group, which included 5035 men and 3163 women. Subjects enrolled in the RV144 trial were vaccinated at weeks 0,4,12 and 24. They were monitored for 42 months for the occurrence of the primary end point of HIV infection after immune responses were measured at week 26, which turned out that 43 out of the 8198 vaccine recipients acquired HIV infection. Vaccine recipients were distributed in the Low, Medium, and High baseline behavioral risk scores, defined as in (Rerks-Ngarm et al., 2009) with 3863 Low, 2370 Medium, and 1965 High.

Three HIV gp120 sequences were included in the vaccine construct; 92TH023 in the ALVAC canarypox vector prime component; and A244 and MN in the AIDSVAX protein boost component. The 92TH023 and A244 are subtype E HIVs whereas MN is subtype B. However, the analysis focuses on the 92TH023 and A244 insert sequences. This is because the subtype E vaccine-insert sequences are genetically much closer to the infecting (and regional circulating) sequences than MN, meaning that the subtype E HIVs are more likely to stimulate protective immune responses. The observed failure time $\tilde{T}_i$ is the time to HIV infection diagnosis, which is the minimum of failure time or right-censoring time.

Because vaccine recipients with higher levels of antibodies binding to the V1V2 portion of the HIV envelope protein had a significantly lower rate of HIV infection (Haynes et al., 2012; Yates et al., 2014; Zolla-Pazner et al., 2014), the V1V2 subregion of gp120 may have been involved in the partial vaccine efficacy administered by the vaccine regimen. This region contains epitopes recognized by antibodies induced by the vaccine. Therefore, we study the genetic distance of an infecting HIV V1V2 sequence to the corresponding V1V2 sequence in the vaccine construct (using a multiple sequence alignment); these are called marks. The way of measuring the genetic distances (marks) is described in Nickle et al. (2007).

Let V be the genetic distances (marks). For the analysis, two marks are considered, based on the 92TH023 and A244 vaccine construct sequences. Let V_{1_i} be the genetic distance mark 92TH023V1V2 and let V_{2_i} be the genetic distance mark A244V1V2 for a subject i. These distances $V = (V_{1_i}, V_{2_i})$ were re-scaled to take values between 0 and 1. We use marks V_{1_i} and V_{2_i} to form two causes of failure each.

Let M_1 be the median of the observed mark V_{1_i} and M_2 be the median of the observed mark V_{2_i} for each subject i. Let ϵ_{1_i} be the cause of failure for the mark V_{1_i}, which is generated by using M_1. We define $\epsilon_{1_i} = 1$ for an uncensored subject i if V_{1_i} is less than M_1; otherwise $\epsilon_{1_i} = 2$. Let ϵ_{2_i} be the cause of failure for the mark V_{2_i}, which is generated by using M_2. Similarly, we define $\epsilon_{2_i} = 1$ for an uncensored subject i if V_{2_i} is less than M_2; otherwise $\epsilon_{2_i} = 2$. If subjects are censored, then $\epsilon_{j_i} = 0$ for $j = 1, 2$.

Following the analysis in Yang et al. (2016), we study IgG and IgG3 biomarkers as correlates of 92TH023V1V2 and A244V1V2 mark-specific HIV infection for the cumulative incidence model based on competing risks data under two-phase sampling. In particular, paired to the 92TH023V1V2 mark variable, we study the two biomarkers Week 26 IgG and IgG3 binding antibodies to 92TH023V1V2, namely IgG-92TH023V1V2 and IgG3-92TH023V1V2; and, paired to the A244V1V2 mark variable, we study Week 26 IgG and IgG3 binding antibodies to A244V1V2, namely IgG-A244V1V2 and IgG3-A244V1V2. Therefore, we consider four different immune responses IgG-92TH023V1V2, IgG3-92TH023V1V2, IgG-A244V1V2 and IgG3-A244V1

V2 for each analysis.

The immune response biomarkers were measured for 34 of the 43 HIV infected vaccine recipients with HIV V1V2 sequence data and 212 of the 8155 uninfected vaccine recipients at the Week 26 visit post entry. These observed biomarkers were each standardized to have mean 0 and variance 1 for each analysis.

Let R_i be the immune responses R_{11_i}, R_{12_i}, R_{21_i}, and R_{22_i}, respectively, for each analysis. Let δ_i be the infection status, whose value is 1 if a subject is infected with HIV; and 0 if a subject i is right censored over a follow-up period of 42 months. We consider two causes of failure $k = 1, 2$. Let $\epsilon_{1_i} = k$ be the causes of failure for immune responses R_{11_i} and R_{12_i}, respectively. Let $\epsilon_{2_i} = k$ be the causes of failure for immune responses R_{21_i} and R_{22_i}, respectively. Let B_{1_i} and B_{2_i} be the dummy variables for baseline behavioral risk score groups B_i (High=1, Low=2, Medium=3), where $B_{1_i} = 1$ if a subject i is in the low risk score group; 0 otherwise, $B_{2_i} = 1$ if a subject i is in the medium risk score group; 0 otherwise and $B_{1_i} = B_{2_i} = 0$ if a subject i is in the high risk group.

The immune responses R_i can be missing for both case and non-case subjects, and hence are phase two covariates. The baseline behavioral risk scores B_i are measure for all subjects, and hence are phase one covariates.

We consider the following semiparametric additive model for the cumulative incidence function by using the identity link function $h(x) = x$ in (1.22):

$$F_{ki}(t; X_i, Z_i) = \eta_0(t) + \eta_1(t)R_i + \gamma_1 B_{1_i} t + \gamma_2 B_{2_i} t \tag{3.38}$$

Let ξ_i be the missing indicator of the immune response data, whose value is $\xi_i = 1$

if each of the four immune response data R_i is measured as a phase-two covariate; otherwise $\xi_i = 0$. Let $\vartheta_i = P(\xi_i = 1|\mathcal{V}_i, \delta_i)$ be the selection probability for a subject i. To predict the probability of observing the immune response R_i, we consider the following logistic model

$$\text{logit}(\vartheta_i) = \theta_0 + \theta_1\delta_i. \tag{3.39}$$

The estimated selection probabilities $\hat{\vartheta}_i$ can be obtained by $\theta = (-3.6235, 4.9526)$ with standard errors (0.06959, 0.38127) in the model (3.39). The weights are given by $\psi(\hat{\theta}_i) = \xi_i/\hat{\vartheta}_i$.

To implement the AIPW method, we use the following linear models:

$$\begin{aligned} E\{R_i|\mathcal{V}_i, \delta_i\} &= \nu_{10} + \nu_{11}B_{1_i} + \nu_{12}B_{2_i} + \nu_{13}\log(\tilde{T}_i) + \nu_{14}\delta_i + \nu_{15}\delta_i * \log(\tilde{T}_i), \\ E\{R_i^{\otimes 2}|\mathcal{V}_i, \delta_i\} &= \varsigma_{10} + \varsigma_{11}B_{1_i} + \varsigma_{12}B_{2_i} + \varsigma_{13}\log(\tilde{T}_i) + \varsigma_{14}\delta_i + \varsigma_{15}\delta_i * \log(\tilde{T}_i). \end{aligned} \tag{3.40}$$

We analyze the semiparametric additive model (2.30) with the following four different settings: (S1). The model (3.38) is analyzed with the immune response $R_i = R_{11_i}$ for $\epsilon_{1_i} = 1$. The first moment $E\{R_i|\mathcal{V}_i, \delta_i\}$ and the second moment $E\{R_i^{\otimes 2}|\mathcal{V}_i, \delta_i\}$ of the missing immune response R_i can be estimated by plugging $\hat{\nu}_1 = (0.2538, -0.2678, 0.0118, -0.0963, -0.2691, 0.0050)$ and $\hat{\varsigma}_1 = (-0.5754, 1.3439, -0.2953, 0.8834, 1.1671, -0.7601)$ into the linear models (3.40). The estimates of B_{1i} and B_{2i} are $\hat{\gamma} = (-0.00156, -0.00157)$ with a standard error of $(0.000895, 0.000978)$, yielding p-value=$(0.08147, 0.10751)$ for testing $\gamma = 0$; Similarly, for $\epsilon_{1i} = 2$, the effects of B_{1_i} and B_{2_i} are $\hat{\gamma} = (-0.00056, -0.00010)$ with a standard error of $(0.00042, 0.00068)$, yielding p-value=$(0.18394, 0.87838)$ for testing $\gamma = 0$;

(S2). The model (3.38) is analyzed with the immune response $R_i = R_{12_i}$ for $\epsilon_{1_i} = 1$. The estimates of $E\{R_i|\mathcal{V}_i, \delta_i\}$ and $E\{R_i^{\otimes 2}|\mathcal{V}_i, \delta_i\}$ are obtained by plugging $\hat{\nu}_1 = (0.3166, -0.3996, -0.1080, -0.0979, -0.1440, 0.1683)$ and $\hat{\varsigma}_1 = (0.5267, 0.6413, 0.1473, 0.1257, 0.3857, -0.5138)$ into the linear models (3.40). This analysis gives the effects of B_{1_i} and B_{2i}, $\hat{\gamma} = (-0.001488, -0.001582)$ with standard error of $(0.000886, 0.000982)$, yielding p-value=$(0.093214, 0.107275)$ for testing $\gamma = 0$; Similarly, for $\epsilon_{1_i} = 2$, the effects of B_{1_i} and B_{2_i} are $\hat{\gamma} = (-0.00058, -0.00012)$ with standard error of $(0.00041, 0.00068)$, yielding p-value=$(0.15749, 0.86411)$ for testing $\gamma = 0$;

(S3). The model (3.38) with the immune response $R_i = R_{21_i}$ is analyzed for $\epsilon_{2_i} = 1$. The conditional expectations can be obtained by plugging $\hat{\nu}_1 = (0.2110, -0.2858, -0.0593, -0.0290, -0.2944, -0.0026)$ and $\hat{\varsigma}_1 = (-0.4441, 1.1693, -0.2345, 0.7951, 1.2254, -0.5539)$ into the linear models (3.40). In this analysis, the estimates of B_{1_i} and B_{2_i} are $\hat{\gamma} = (-0.00084, -0.00081)$ with standard error of $(0.00057, 0.00063)$, yielding p-value=$(0.14261, 0.19699)$ for testing $\gamma = 0$; Similarly, for $\epsilon_{2_i} = 2$, the estimates of B_{1_i} and B_{2_i} are $\hat{\gamma} = (-0.00124, -0.00081)$ with standard error of $(0.00072, 0.00082)$, yielding p-value=$(0.08619, 0.32526)$ for testing $\gamma = 0$;

(S4). The model (3.38) with immune response $R_i = R_{22_i}$ is analyzed for $\epsilon_{2_i} = 1$. The conditional expectations can be estimated by plugging $\hat{\nu}_1 = (0.3217, -0.4299, -0.1495, -0.0622, -0.2350, 0.0670)$ and $\hat{\varsigma}_1 = (0.3486, 0.5998, -0.0997, 0.3130, 0.9022, -0.6895)$ into the models (3.40). This analysis gives the estimates $\hat{\gamma} = (-0.000864, -0.000852)$ with standard error of $(0.000577, 0.000651)$, yielding p-value=$(0.13433, 0.19080)$ for testing $\gamma = 0$; Similarly, for cause $\epsilon_{2_i} = 2$, The effects of B_{1_i} and B_{2_i} are $\hat{\gamma} = (-0.00121, -0.00080)$ with standard error of $(0.00071, 0.00083)$, yielding p-value=$(0.08944, 0.33091)$ for testing $\gamma = 0$.

Figure 21 to 24 compares the AIPW estimates of the baseline cumulative coefficients $\eta_0(t)$ and the cumulative coefficients $\eta_1(t)$ with 95% pointwise confidence intervals for the four different immune responses of R_i for $\epsilon_{j_i} = 1$ and $\epsilon_{j_i} = 2$, respectively, where $j = 1, 2$. The analyses with R_{11_i} (IgG-92TH023V1V2) and R_{12_i} (IgG-92TH023V1V2) for $\epsilon_{1_i} = 1$ have larger AIPW estimates of baseline cumulative coefficients $\eta_0(t)$ than those for $\epsilon_{1_i} = 2$. The analyses with R_{21_i} (IgG-A244V1V2) and R_{22_i} (IgG3-A244V1V2) for $\epsilon_{2_i} = 1$ have smaller AIPW estimates of the baseline cumulative coefficients $\eta_0(t)$ than those for $\epsilon_{2_i} = 2$. For the estimates of the cumulative coefficients $\eta_1(t)$, while the effects of the immune responses IgG-92TH023V1V2, IgG-A244V1V2 and IgG3-A244V1V2 have negative effects on the cumulative incidence function with $\epsilon_{j_i} = 1$, $j = 1, 2$ in Figure 21, 23, and 24, the effects of immune responses IgG3-92TH023V1V2 are close to zero over study time with $\epsilon_{1_i} = 1$ in Figure 22. However, the effects of IgG-92TH023V1V2 are negative on the cumulative incidence function in the upper range of time, where $t \leq 1.8$. On the other hand, each of these four immune responses R_i with $\epsilon_{1_i} = 2$ or $\epsilon_{2_i} = 2$ has no effect on the cumulative incidence function over the study time. By comparing Figure 21 to Figure 23 and comparing Figure 22 to Figure 24, IgG and IgG3 binding antibodies responding to A244V1V2 have significantly negative effects on the cumulative incidence function than those antibodies responding to 92TH023V1V2, i.e A244 would be more relevant for protection of HIV with V1V2 sequences.

The relationship between the behavioral risk scores and cumulative incidence of HIV infection is summarized in Figures 25-28. For the four different immune responses

R_i, these four figures show the cumulative incidence functions, $F_k(t)$, are estimated at the first, second and third quartiles Q_1, Q_2 and Q_3 of the observed immune response R_i at each level of the behavioral risk score groups (Low, Medium and High). In each figure, for $\epsilon_{1_i} = 1$ and $\epsilon_{2_i} = 1$, the high risk score group tends to have a higher probability of getting infected by HIV with V1V2 sequences than the low and medium risk score groups. However, this tendency is less noticeable for the immune responses IgG-92TH023V1V2 and IgG3-92TH023V1V2 with $\epsilon_{1_i} = 2$ in Figures 25-26.

Figures 25-28 also describe the relationship between the immune responses and cumulative incidence of HIV infection. The estimated cumulative incidence functions for IgG-92TH023V1V2 with $\epsilon_{1_i} = 1$, IgG-A244V1V2 and IgG3-A244V1V2 with $\epsilon_{2_i} = 1$, respectively, show that subjects in the third quartile group (Q_3) have lower cumulative incidence of HIV infection than those in the second quartile group (Q_2), which have in turn lower cumulative incidence of HIV infection than those in the first quartile group (Q_1) at each behavioral risk group (a), (b) and (c) in Figures 25, 27, and 28. This tendency is clearly noticeable for immune responses IgG-A244V1V2 and IgG3-A244V1V2 with $\epsilon_{2_i} = 1$, meaning that one with a higher immune response may be strongly against HIV infection with the V1V2 sequences closer to A244.

Figures 25-28 show the relationship between the genetic distance (mark variable) and the cumulative incidence function $F_k(t)$. Note that the competing marks $\epsilon_{j_i} = 1$ and $\epsilon_{j_i} = 2$, where $\epsilon_{j_i} = 1$ is less than the median mark and $\epsilon_{j_i} = 2$ is greater than the median mark, where $j = 1, 2$. We expect there to be a lower probability of getting infected by HIV with V1V2 sequences closer to 92TH023 or A244 ($\epsilon_{j_i} = 1$, $j = 1, 2$) and expect a higher probability of getting infected by HIV with V1V2 sequences are far away from 92TH023 or A244 ($\epsilon_{j_i} = 2,\ j = 1, 2$). However, for immune responses IgG-92TH023V1V2 and IgG3-92TH023V1V2, it is not always true that the marks less than the median mark lead to lower probabilities of HIV infection since these analyses depend on the prevalence of each circulating HIV strain with genetic distance (specific mark) in the population. For example, (c) in Figure 25 shows that the cumulative incidence of HIV infection from the mark less than the median mark ($\epsilon_{1_i} = 1$) is larger than the cumulative incidence of HIV infection from the mark farther than the median mark $\epsilon_{1_i} = 2$. This is because the HIV infection from the shorter marks may be more prevalent and exposed to more people than HIV infection from the farther marks.

These analyses imply that the IgG3 antibody to 92TH023V1V2 does not have effect on the cumulative incidence of HIV infection in Figure 8 and other IgG subclasses besides type 3 induced by 92TH023 would have negative effects on the cumulative incidence function. These analyses also imply that A244 was more important than 92TH023 for induction of protective IgG3 antibodies. These analyses support the hypothesis that IgG and IgG3 binding antibodies protect less against exposures to HIV V1V2 sequences with farther distance away from the A244V1V2 vaccine sequences and strongly against HIV infecting sequences with small V1V2 distance. Therefore, vaccine recipients exposed to HIV with V1V2 sequences close to A244 (less marks than the median mark) may be more likely to be protected by these antibodies than vaccine recipients exposed to HIV with V1V2 sequences that have marks farther than the median mark.

Table 5: Bias, empirical standard error (SSE), average of the estimated standard error (ESE), and empirical coverage probability (CP) of 95% confidence intervals for the AIPW estimator of γ under model (3.30) with $\rho = 0.5,\ 0.8,\ 0.9$ and with the average of the total missing probabilities $m_0 = 0.3$ and about 50% censoring based on 1000 simulations for sampling scenarios, I, II and III, where m_1 and m_2 are the averages of the missing probabilities for the cases and the non-cases, respectively.

					γ			
Sampling	ρ	m_0	(m_1, m_2)	n	Bias	SSE	ESE	CP
I	0.5	0.3	(0, 0.36)	500	0.0022	0.0216	0.0220	0.956
				700	0.0025	0.0186	0.0186	0.946
				900	0.0018	0.0159	0.0163	0.961
II			(0.20, 0.30)	500	0.0019	0.0217	0.0221	0.954
				700	0.0024	0.0190	0.0187	0.943
				900	0.0018	0.0160	0.0164	0.955
III				500	0.0019	0.0220	0.0224	0.949
				700	0.0024	0.0192	0.0189	0.941
				900	0.0019	0.0163	0.0166	0.956
I	0.8	0.3	(0, 0.36)	500	0.0021	0.0215	0.0220	0.957
				700	0.0024	0.0187	0.0186	0.946
				900	0.0018	0.0159	0.0163	0.956
II			(0.20, 0.30)	500	0.0019	0.0216	0.0220	0.951
				700	0.0024	0.0187	0.0186	0.944
				900	0.0018	0.0160	0.0164	0.956
III				500	0.0019	0.0219	0.0222	0.950
				700	0.0024	0.0188	0.0187	0.942
				900	0.0018	0.0162	0.0165	0.953
I	0.9	0.3	(0, 0.30)	500	0.0021	0.0214	0.0220	0.955
				700	0.0024	0.0986	0.0186	0.947
				900	0.0018	0.0159	0.0163	0.956
II			(0.20, 0.30)	500	0.0020	0.0215	0.0220	0.951
				700	0.0024	0.0187	0.0186	0.950
				900	0.0018	0.0159	0.0163	0.955
III				500	0.0020	0.0216	0.0220	0.952
				700	0.0024	0.0187	0.0186	0.946
				900	0.0019	0.0161	0.0163	0.951

Table 6: Comparison of bias (Bias), empirical standard error (SSE), average of the estimated standard error (ESE) and empirical coverage probability (CP) of 95% confidence intervals for Full, AIPW, IPW and CC estimators of γ under (3.30) with $\rho = 0.5,\ 0.8,\ 0.9$ and with the average of the total missing probabilities $m_0 = 0.3$ and about 50% censoring based on 1000 simulations for sampling scenarios, I, II and III, where m_1 and m_2 are the averages of the missing probabilities for the cases and the non-cases, respectively.

					Bias(γ)				SSE(γ)				ESE(γ)				CP(γ)			
Sampling	ρ	m_0	(m_1, m_2)	n	Full	AIPW	IPW	CC	Full	AIPW	IPW	CC	Full	AIPW	IPW	CC	Full	AIPW	IPW	CC
I	0.5	0.3	(0,0.36)	500	0.0020	0.0022	0.0076	0.0575	0.0214	0.0216	0.0226	0.0323	0.0217	0.0220	0.0244	0.0325	0.954	0.956	0.956	0.567
				700	0.0024	0.0025	0.0080	0.0579	0.0186	0.0186	0.0196	0.0274	0.0183	0.0186	0.0206	0.0275	0.947	0.946	0.946	0.429
				900	0.0018	0.0018	0.0075	0.0570	0.0159	0.0159	0.0170	0.0238	0.0161	0.0163	0.0181	0.0242	0.954	0.961	0.955	0.340
II			(0.20,0.30)	500		0.0019	0.0057	0.0199		0.0217	0.0245	0.0278		0.0221	0.0258	0.0287		0.954	0.964	0.901
				700		0.0024	0.0068	0.0212		0.0190	0.0212	0.0244		0.0187	0.0218	0.0243		0.943	0.942	0.875
				900		0.0018	0.0058	0.0201		0.0160	0.0187	0.0215		0.0164	0.0192	0.0214		0.955	0.953	0.844
III				500		0.0019	0.0017	0.0017		0.0220	0.0255	0.0254		0.0224	0.0263	0.0263		0.949	0.953	0.956
				700		0.0024	0.0031	0.0030		0.0192	0.0224	0.0224		0.0189	0.0223	0.0223		0.941	0.937	0.938
				900		0.0019	0.0017	0.0017		0.0163	0.0194	0.0194		0.0166	0.0196	0.0196		0.956	0.955	0.956
I	0.8	0.3	(0,0.36)	500		0.0021	0.0076	0.0579		0.0215	0.0226	0.0324		0.0220	0.0244	0.0326		0.957	0.959	0.570
				700		0.0024	0.0079	0.0579		0.0187	0.0197	0.0276		0.0186	0.0206	0.0275		0.946	0.942	0.435
				900		0.0018	0.0074	0.0571		0.0159	0.0169	0.0238		0.0163	0.0181	0.0242		0.956	0.953	0.328
II			(0.20,0.30)	500		0.0019	0.0056	0.0199		0.0216	0.0245	0.0278		0.0220	0.0258	0.0287		0.951	0.963	0.902
				700		0.0024	0.0068	0.0213		0.0187	0.0211	0.0243		0.0186	0.0218	0.0243		0.944	0.945	0.873
				900		0.0018	0.0058	0.0202		0.0160	0.0188	0.0216		0.0164	0.0192	0.0214		0.956	0.953	0.842
III				500		0.0019	0.0018	0.0018		0.0219	0.0254	0.0253		0.0222	0.0262	0.0262		0.950	0.956	0.955
				700		0.0024	0.0031	0.0031		0.0188	0.0223	0.0223		0.0187	0.0222	0.0222		0.942	0.939	0.943
				900		0.0018	0.0018	0.0018		0.0162	0.0195	0.0195		0.0165	0.0195	0.0195		0.953	0.953	0.953
I	0.9	0.3	(0,0.36)	500		0.0021	0.0075	0.0578		0.0214	0.0224	0.0324		0.0220	0.0244	0.0326		0.955	0.963	0.570
				700		0.0024	0.0079	0.0580		0.0186	0.0197	0.0276		0.0186	0.0206	0.0276		0.947	0.941	0.439
				900		0.0018	0.0074	0.0572		0.0159	0.0169	0.0238		0.0163	0.0181	0.0242		0.956	0.954	0.334
II			(0.20,0.30)	500		0.0020	0.0055	0.0199		0.0215	0.0245	0.0278		0.0220	0.0258	0.0287		0.951	0.964	0.903
				700		0.0024	0.0067	0.0213		0.0187	0.0211	0.0243		0.0186	0.0219	0.0243		0.950	0.945	0.870
				900		0.0018	0.0058	0.0202		0.0159	0.0188	0.0216		0.0163	0.0192	0.0214		0.955	0.953	0.836
III				500		0.0020	0.0018	0.0018		0.0216	0.0254	0.0254		0.0220	0.0262	0.0262		0.952	0.954	0.954
				700		0.0024	0.0032	0.0031		0.0187	0.0223	0.0222		0.0186	0.0222	0.0222		0.946	0.943	0.943
				900		0.0019	0.0018	0.0017		0.0161	0.0195	0.0194		0.0163	0.0195	0.0195		0.951	0.955	0.957

Table 7: The Relative efficiencies (REE) of AIWP, IPW and CC estimators compared to the Full estimator for γ under model (3.30) with $\rho = 0.5,\ 0.8,\ 0.9$ and with the average of the total missing probabilities $m_0 = 0.3$ and about 50% censoring based on 1000 simulations for sampling scenarios, I, II and III, where m_1 and m_2 are the averages of the missing probabilities for the cases and the non-cases, respectively.

					REE(γ)		
Sampling	ρ	m_0	(m_1, m_2)	n	AIPW	IPW	CC
I	0.5	0.3	(0,0.36)	500	0.9920	0.9462	0.6627
				700	0.9966	0.9481	0.6793
				900	0.9985	0.9372	0.6678
II			(0.20,0.30)	500	0.9869	0.8729	0.7709
				700	0.9803	0.8764	0.7626
				900	0.9931	0.8502	0.7403
III				500	0.9718	0.8400	0.8420
				700	0.9666	0.8280	0.8293
				900	0.9791	0.8205	0.8215
I	0.8	0.3	(0,0.36)	500	0.9945	0.9486	0.6606
				700	0.9959	0.9439	0.6741
				900	1	0.9411	0.6681
II			(0.20,0.30)	500	0.9913	0.8720	0.7695
				700	0.9935	0.8790	0.7645
				900	0.9975	0.8479	0.7385
III				500	0.9769	0.8423	0.8447
				700	0.9874	0.8334	0.8349
				900	0.9860	0.8166	0.8177
I	0.9	0.3	(0,0.36)	500	0.9988	0.9541	0.6597
				700	0.9982	0.9448	0.6742
				900	0.9991	0.9414	0.6691
II			(0.20,0.30)	500	0.9941	0.8720	0.7702
				700	0.9947	0.8787	0.7645
				900	0.9990	0.8483	0.7388
III				500	0.9908	0.8411	0.8432
				700	0.9922	0.8342	0.8353
				900	0.9905	0.8178	0.8191

Table 8: Bias, empirical standard error (SSE), average of the estimated standard error (ESE) and empirical coverage probability (CP) of 95% confidence intervals for the AIPW estimator of γ under model (3.30) with $\rho = 0.5,\ 0.8,\ 0.9$ and with the average of the total missing probabilities $m_0 = 0.6$ and about 50% censoring based on 1000 simulations for sampling scenarios, I, II and III, where m_1 and m_2 are the averages of the missing probabilities for the cases and the non-cases, respectively.

					γ			
Sampling	ρ	m_0	(m_1, m_2)	n	Bias	SSE	ESE	CP
I	0.5	0.6	(0, 0.65)	500	0.0030	0.0226	0.0262	0.974
				700	0.0029	0.0193	0.0218	0.963
				900	0.0022	0.0167	0.0189	0.969
II			(0.45, 0.60)	500	0.0019	0.0228	0.0237	0.966
				700	0.0018	0.0195	0.0199	0.957
				900	0.0019	0.0166	0.0175	0.962
III				500	0.0023	0.0242	0.0242	0.954
				700	0.0030	0.0196	0.0204	0.956
				900	0.0017	0.0179	0.0180	0.943
I	0.8	0.6	(0, 0.65)	500	0.0025	0.0219	0.0264	0.980
				700	0.0026	0.0191	0.0219	0.967
				900	0.0020	0.0163	0.0190	0.976
II			(0.45, 0.60)	500	0.0021	0.0225	0.0232	0.958
				700	0.0026	0.0187	0.0197	0.961
				900	0.0017	0.0166	0.0173	0.952
III				500	0.0026	0.0231	0.0236	0.955
				700	0.0027	0.0188	0.0199	0.957
				900	0.0017	0.0172	0.0175	0.947
I	0.9	0.6	(0, 0.65)	500	0.0022	0.0216	0.0264	0.979
				700	0.0025	0.0187	0.0219	0.967
				900	0.0019	0.0161	0.0190	0.981
II			(0.45, 0.60)	500	0.0023	0.0222	0.0231	0.957
				700	0.0026	0.0184	0.0195	0.964
				900	0.0019	0.0164	0.0171	0.952
III				500	0.0026	0.0226	0.0232	0.954
				700	0.0027	0.0184	0.0196	0.959
				900	0.0018	0.0167	0.0172	0.951

Table 9: Comparison of bias (Bias), empirical standard error (SSE), average of the estimated standard error (ESE) and empirical coverage probability (CP) of 95% confidence intervals for Full, AIPW, IPW and CC estimators of γ under (3.30) with $\rho = 0.5,\ 0.8,\ 0.9$ and with the average of the total missing probabilities $m_0 = 0.6$ and about 50% censoring based on 1000 simulations for sampling scenarios, I, II and III, where m_1 and m_2 are the averages of the missing probabilities for the cases and the non-cases, respectively.

					Bias(γ)				SSE(γ)				ESE(γ)				CP(γ)			
Sampling	ρ	m_0	(m_1, m_2)	n	Full	AIPW	IPW	CC	Full	AIPW	IPW	CC	Full	AIPW	IPW	CC	Full	AIPW	IPW	CC
I	0.5	0.6	(0,0.65)	500	0.0020	0.0030	0.0395	0.2025	0.0214	0.0226	0.0395	0.0611	0.0217	0.0262	0.0433	0.0637	0.954	0.974	0.899	0.099
				700	0.0024	0.0029	0.0383	0.2041	0.0186	0.0193	0.0333	0.0525	0.0183	0.0218	0.0368	0.0539	0.947	0.963	0.871	0.036
				900	0.0018	0.0022	0.0370	0.2044	0.0159	0.0167	0.0292	0.0469	0.0161	0.0189	0.0327	0.0479	0.954	0.969	0.836	0.015
II			(0.45,0.60)	500		0.0019	0.0242	0.0662		0.0228	0.0395	0.0502		0.0237	0.0423	0.0505		0.966	0.961	0.744
				700		0.0018	0.0235	0.0660		0.0195	0.0340	0.0421		0.0199	0.0359	0.0430		0.957	0.939	0.670
				900		0.0019	0.0243	0.0678		0.0166	0.0291	0.0371		0.0175	0.0317	0.0382		0.962	0.925	0.576
III				500		0.0023	0.0036	0.0034		0.0242	0.0394	0.0386		0.0242	0.0377	0.0374		0.954	0.937	0.939
				700		0.0030	0.0022	0.0019		0.0196	0.0311	0.0206		0.0204	0.0319	0.0316		0.956	0.955	0.951
				900		0.0017	0.0013	0.0013		0.0179	0.0276	0.0274		0.0180	0.0281	0.0279		0.943	0.955	0.957
I	0.8	0.6	(0,0.65)	500		0.0025	0.0387	0.2030		0.0219	0.0395	0.0617		0.0264	0.0434	0.0639		0.980	0.903	0.104
				700		0.0026	0.0379	0.2039		0.0191	0.0336	0.0528		0.0219	0.0369	0.0540		0.967	0.865	0.037
				900		0.0020	0.0368	0.2048		0.0163	0.0293	0.0469		0.0190	0.0327	0.0479		0.976	0.839	0.014
II			(0.45,0.60)	500		0.0021	0.0252	0.0685		0.0225	0.0405	0.0515		0.0232	0.0421	0.0503		0.958	0.956	0.713
				700		0.0026	0.0275	0.0714		0.0187	0.0332	0.0418		0.0197	0.0364	0.0434		0.961	0.931	0.635
				900		0.0017	0.0255	0.0685		0.0166	0.0294	0.0370		0.0173	0.0317	0.0379		0.952	0.931	0.550
III				500		0.0026	0.0033	0.0033		0.0231	0.0387	0.0381		0.0236	0.0379	0.0376		0.955	0.944	0.945
				700		0.0027	0.0020	0.0017		0.0188	0.0310	0.0305		0.0199	0.0321	0.0318		0.957	0.962	0.961
				900		0.0017	0.0005	0.0006		0.0172	0.0272	0.0271		0.0175	0.0282	0.0281		0.947	0.951	0.953
I	0.9	0.6	(0,0.65)	500		0.0022	0.0385	0.2035		0.0216	0.0396	0.0616		0.0264	0.0434	0.0640		0.979	0.895	0.111
				700		0.0025	0.0375	0.2039		0.0187	0.0334	0.0528		0.0219	0.0369	0.0541		0.967	0.864	0.032
				900		0.0019	0.0369	0.2050		0.0161	0.0294	0.0470		0.0190	0.0328	0.0479		0.981	0.846	0.014
II			(0.45,0.60)	500		0.0023	0.0252	0.0683		0.0222	0.0406	0.0513		0.0231	0.0423	0.0504		0.957	0.957	0.723
				700		0.0026	0.0273	0.0711		0.0184	0.0331	0.0417		0.0195	0.0365	0.0434		0.964	0.932	0.638
				900		0.0019	0.0257	0.0685		0.0164	0.0294	0.0368		0.0171	0.0319	0.0380		0.952	0.929	0.555
III				500		0.0026	0.0037	0.0037		0.0226	0.0387	0.0381		0.0232	0.0379	0.0377		0.954	0.945	0.948
				700		0.0027	0.0022	0.0018		0.0184	0.0312	0.0306		0.0196	0.0322	0.0320		0.959	0.958	0.960
				900		0.0018	0.0006	0.0006		0.0167	0.0272	0.0270		0.0172	0.0283	0.0281		0.951	0.953	0.958

Table 10: The Relative efficiencies (REE) of AIPW, IPW and CC estimators compared to the Full estimator for γ under model (3.30) with $\rho = 0.5,\ 0.8,\ 0.9$ and with the average of the total missing probabilities $m_0 = 0.6$ and about 50% censoring based on 1000 simulations for sampling scenarios, I, II and III, where m_1 and m_2 are the averages of the missing probabilities for the cases and the non-cases, respectively.

					REE(γ)		
Sampling	ρ	m_0	$(\vartheta_1, \vartheta_2)$	n	AIPW	IPW	CC
I	0.5	0.6	(0,0.65)	500	0.9478	0.5414	0.3500
				700	0.9633	0.5574	0.3539
				900	0.9535	0.5445	0.3398
II			(0.45,0.60)	500	0.9396	0.5419	0.4264
				700	0.9505	0.5467	0.4409
				900	0.9599	0.5477	0.4297
III				500	0.8985	0.5518	0.5621
				700	0.9204	0.5815	0.5907
				900	0.9127	0.5921	0.5961
I	0.8	0.6	(0,0.65)	500	0.9763	0.5416	0.3471
				700	0.9732	0.5536	0.3522
				900	0.9760	0.5430	0.3395
II			(0.45,0.60)	500	0.9652	0.5365	0.4216
				700	0.9658	0.5437	0.4329
				900	0.9855	0.5542	0.4407
III				500	0.9414	0.5620	0.5703
				700	0.9598	0.5836	0.5930
				900	0.9506	0.5993	0.6022
I	0.9	0.6%	(0,0.65)	500	0.9851	0.7627	0.4626
				700	0.9875	0.7682	0.4636
				900	0.9943	0.7813	0.4606
II			(0.45,0.60)	500	0.9780	0.5348	0.4236
				700	0.9830	0.5459	0.4334
				900	0.9977	0.5554	0.4429
III				500	0.9603	0.5620	0.5706
				700	0.9810	0.5798	0.5901
				900	0.9772	0.6008	0.6405

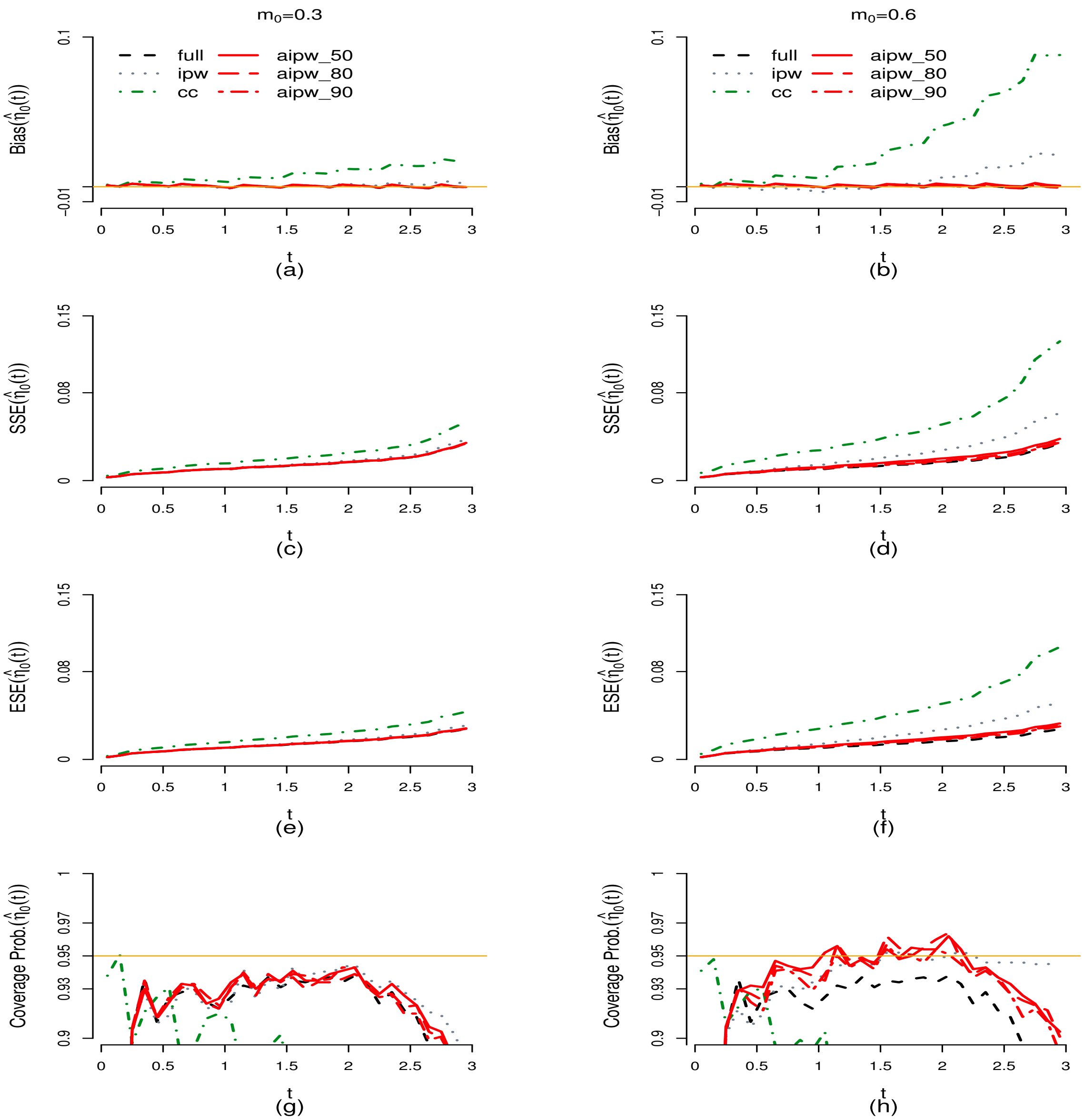


Figure 15: Comparison of Full, IPW, CC, AIPW-50, AIPW-80, AIPW-90 estimators for the baseline cumulative coefficient $\eta_0(t)$ for the average of the total missing probabilities $m_0 = 0.3$ and $m_0 = 0.6$, respectively, under (3.30) with sampling scenario I. For $m_0 = 0.3$, we have $m_1 = 0$ and $m_2 = 0.36$. For $m_0 = 0.6$, we have $m_1 = 0$ and $m_2 = 0.65$. These results are based on 1000 simulations with $n = 700$ and 50% censoring. (a), (b): The plots of the biases of the estimates of $\eta_0(t)$. (c), (d):The plots of the empirical standard errors of the estimates of $\eta_0(t)$. (e), (f): The plots of the average of the estimated standard errors of the estimates of $\eta_0(t)$. (g)(h): The plots of the coverage probabilities of the estimators of $\eta_0(t)$.

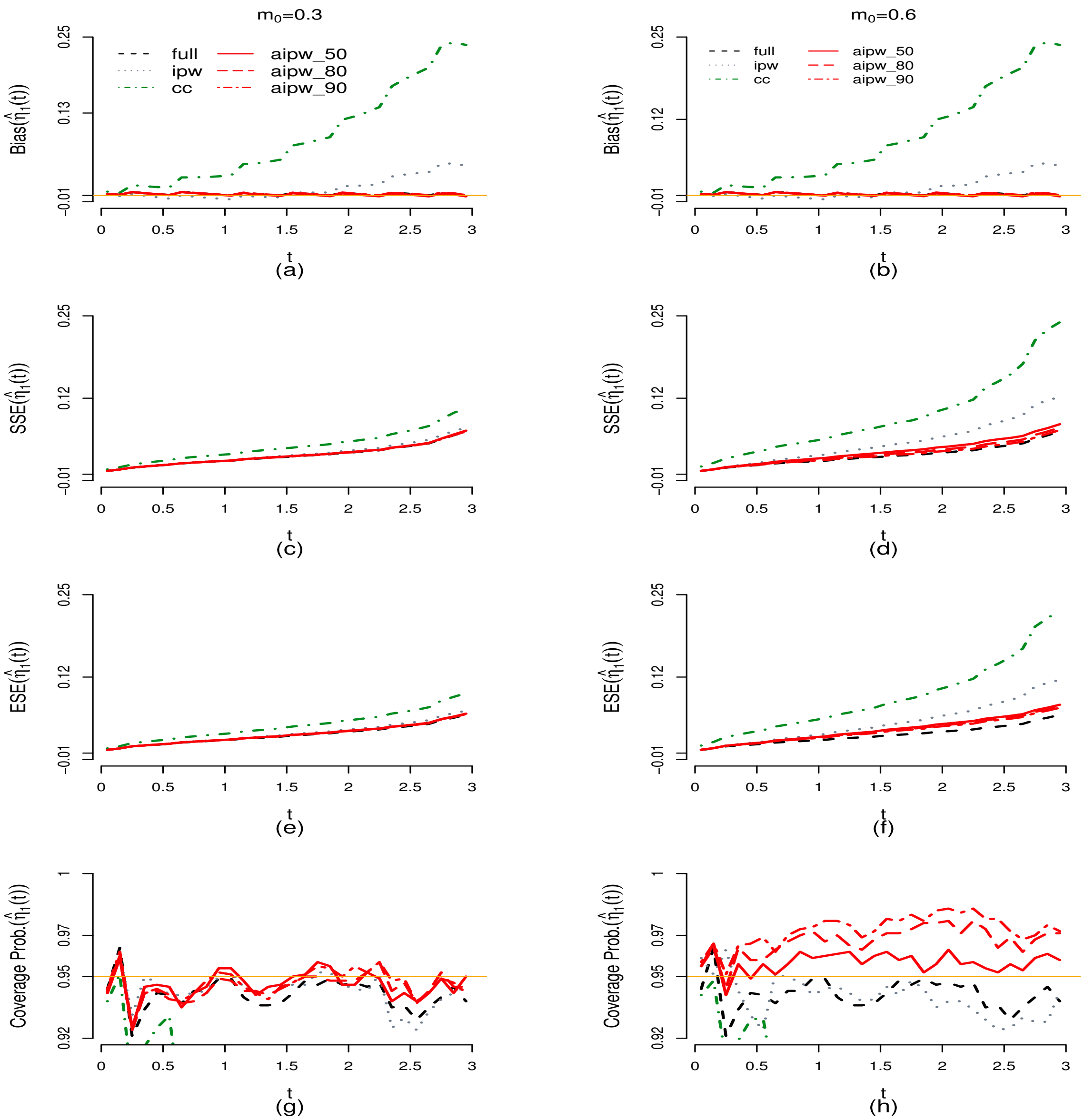


Figure 16: Comparison of Full, IPW, CC, AIPW-50, AIPW-80, AIPW-90 estimators for the cumulative coefficient $\eta_1(t)$ for the average of the total missing probabilities $m_0 = 0.3$ and $m_0 = 0.6$, respectively, under (3.30) with sampling scenario I. For $m_0 = 0.3$, we have $m_1 = 0$ and $m_2 = 0.36$. For $m_0 = 0.6$, we have $m_1 = 0$ and $m_2 = 0.65$. These results are based on 1000 simulations with $n = 700$ and 50% censoring. (a), (b): The plots of the biases of the estimates of $\eta_1(t)$. (c), (d):The plots of the empirical standard errors of the estimates of $\eta_1(t)$. (e), (f): The plots of the average of the estimated standard errors of the estimates of $\eta_1(t)$. (g), (h): The plots of the coverage probabilities of the estimators of $\eta_1(t)$.

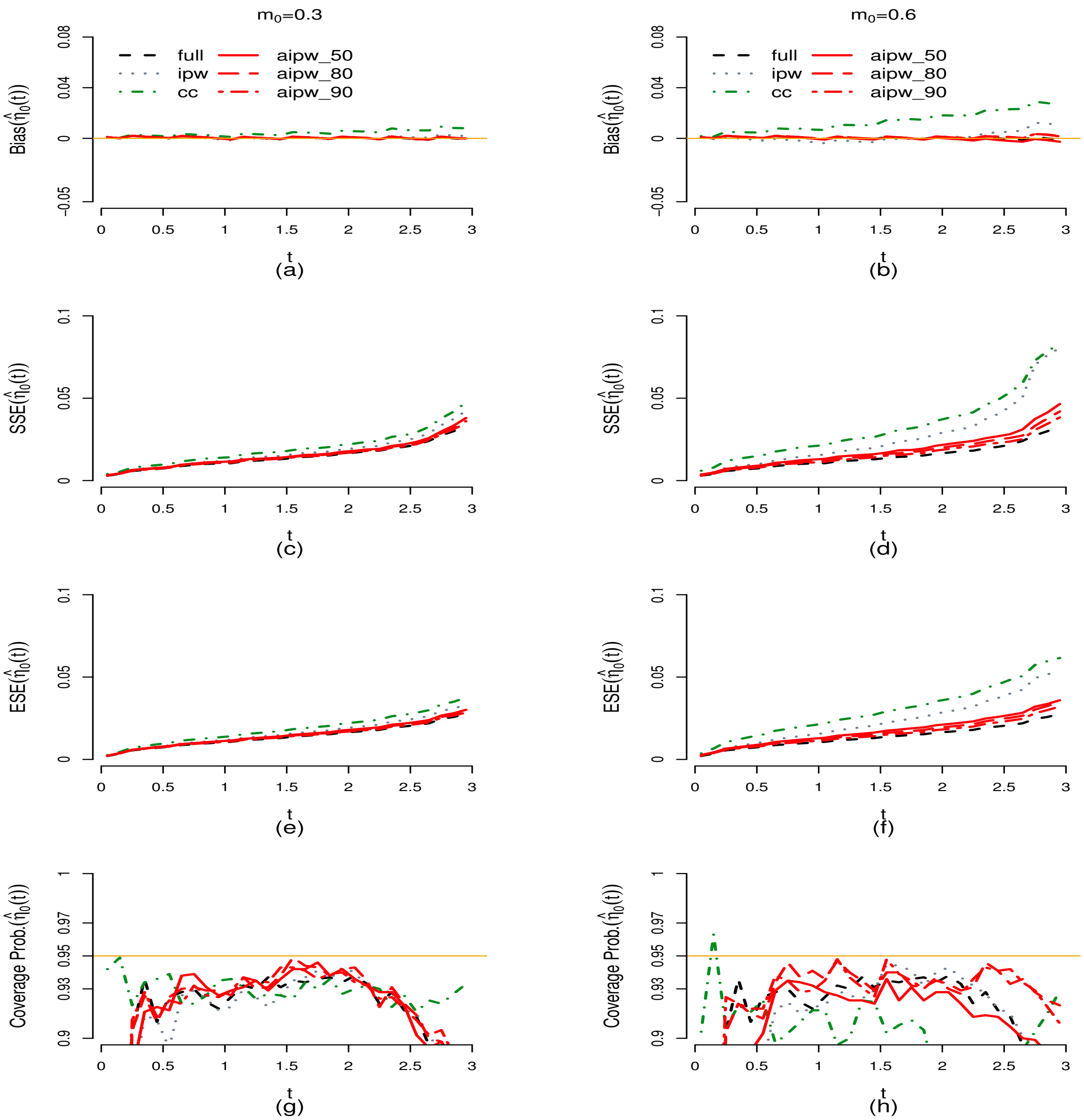


Figure 17: Comparison of Full, IPW, CC, AIPW-50, AIPW-80, AIPW-90 estimators for the baseline cumulative coefficient $\eta_0(t)$ for the average of the total missing probabilities $m_0 = 0.3$ and $m_0 = 0.6$, respectively, under (3.30) with sampling scenario II. For $m_0 = 0.3$, we have $m_1 = 0.2$ and $m_2 = 0.3$. For $m_0 = 0.6$, we have $m_1 = 0.45$ and $m_2 = 0.60$. These results are based on 1000 simulations with $n = 700$ and 50% censoring. (a), (b): The plots of the biases of the estimates of $\eta_0(t)$. (c), (d): The plots of the empirical standard errors of the estimates of $\eta_0(t)$. (e), (f): The plots of the average of the estimated standard errors of the estimates of $\eta_0(t)$. (g), (h): The plots of the coverage probabilities of the estimators of $\eta_0(t)$.

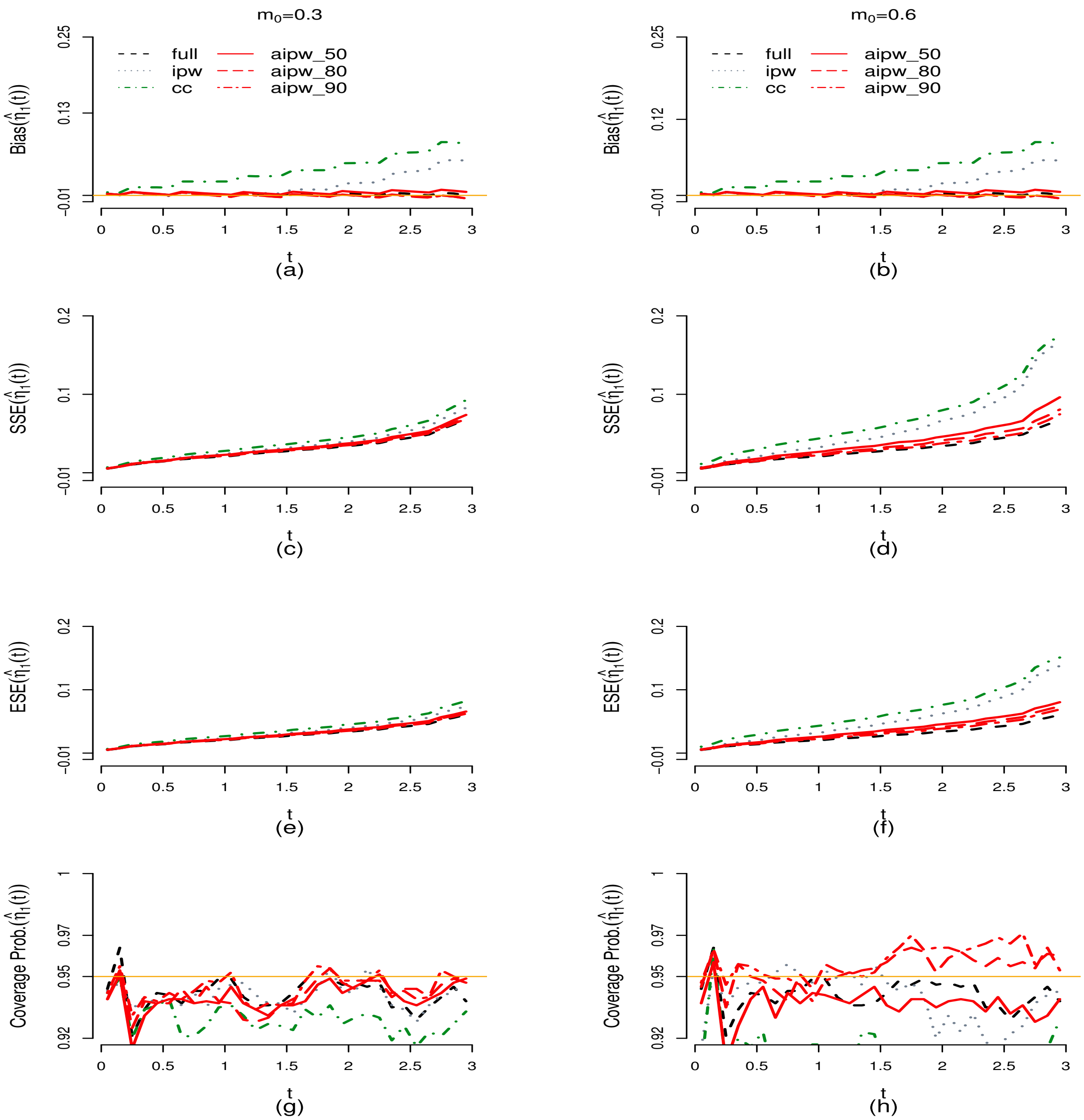


Figure 18: Comparison of Full, IPW, CC, AIPW-50, AIPW-80, AIPW-90 estimators for the cumulative coefficient $\eta_1(t)$ for the average of the total missing probabilities $m_0 = 0.3$ and $m_0 = 0.6$, respectively, under (3.30) with sampling scenario II. For $m_0 = 0.3$, we have $m_1 = 0.2$ and $m_2 = 0.3$. For $m_0 = 0.6$, we have $m_1 = 0.45$ and $m_2 = 0.60$. These results are based on 1000 simulations with $n = 700$ and 50% censoring. (a), (b): The plots of the biases of the estimates of $\eta_1(t)$. (c), (d):The plots of the empirical standard errors of the estimates of $\eta_1(t)$. (e),(f): The plots of the average of the estimated standard errors of the estimates of $\eta_1(t)$. (g),(h): The plots of the coverage probabilities of the estimators of $\eta_1(t)$.

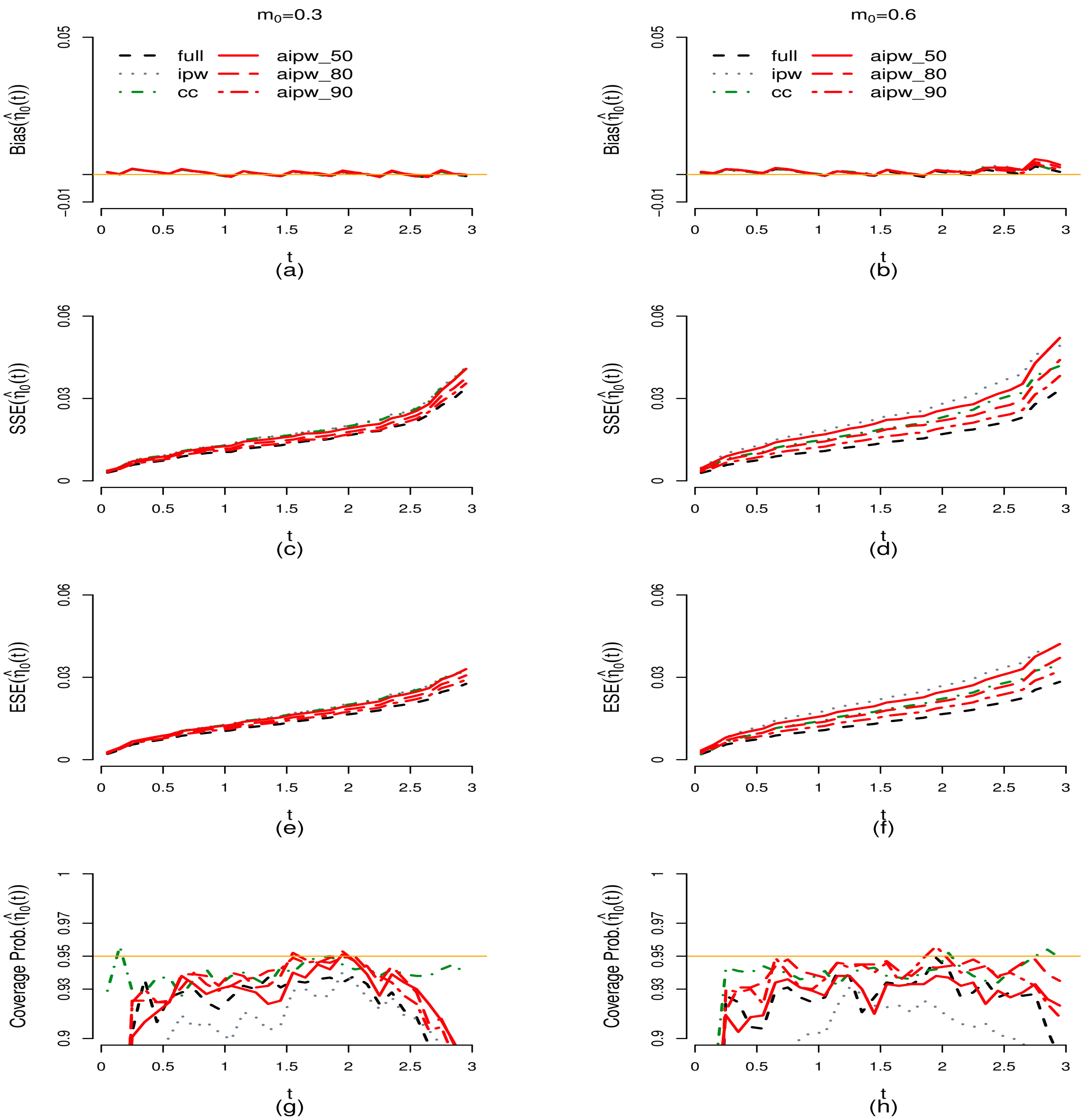


Figure 19: Comparison of Full, IPW, CC, AIPW-50, AIPW-80, AIPW-90 estimators for the baseline cumulative coefficient $\eta_0(t)$ for the average of the total missing probabilities $m_0 = 0.3$ and $m_0 = 0.6$, respectively, under (3.30) with sampling scenario III. These results are based on 1000 simulations with $n = 700$ and 50% censoring. (a), (b): The plots of the biases of the estimates of $\eta_0(t)$.(c), (d):The plots of the empirical standard errors of the estimates of $\eta_0(t)$. (e), (f): The plots of the average of the estimated standard errors of the estimates of $\eta_0(t)$. (g), (h): The plots of the coverage probabilities of the estimators of $\eta_0(t)$.

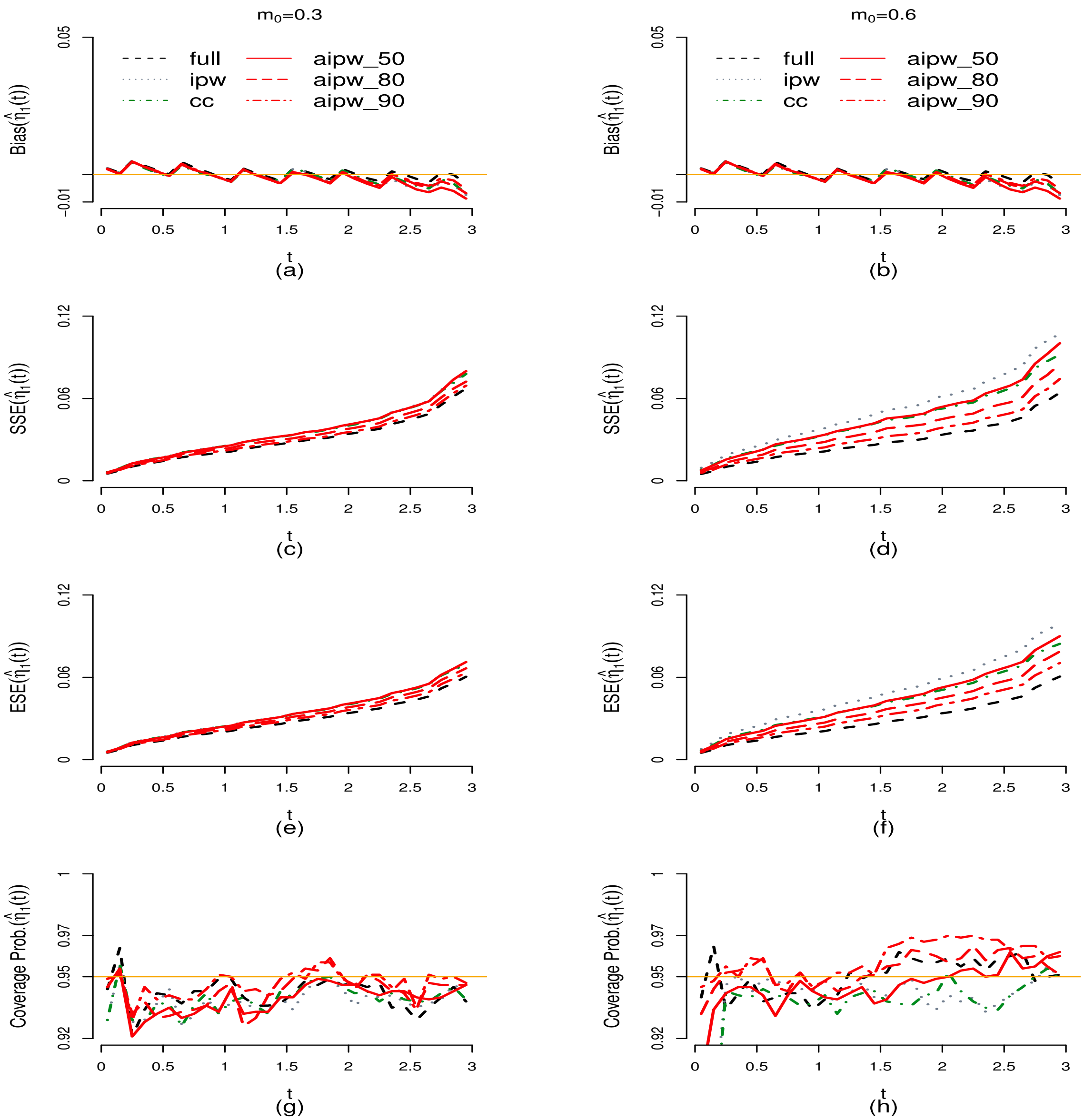
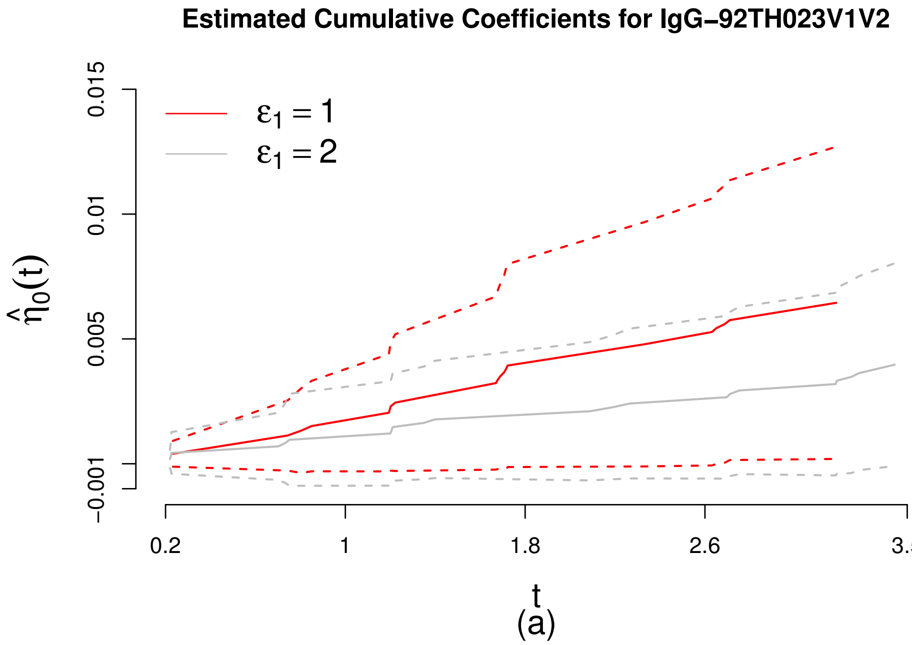


Figure 20: Comparison of Full, IPW, CC, AIPW-50, AIPW-80, AIPW-90 estimators for the cumulative coefficient $\eta_1(t)$ for the average of the total missing probabilities $m_0 = 0.3$ and $m_0 = 0.6$, respectively, under (3.30) with sampling scenario III. These results are based on 1000 simulations with $n = 700$ and 50% censoring. (a), (b): The plots of the biases of the estimates of $\eta_1(t)$.(c), (d):The plots of the empirical standard errors of the estimates of $\eta_1(t)$. (e), (f): The plots of the average of the estimated standard errors of the estimates of $\eta_1(t)$. (g), (h): The plots of the coverage probabilities of the estimators of $\eta_1(t)$.



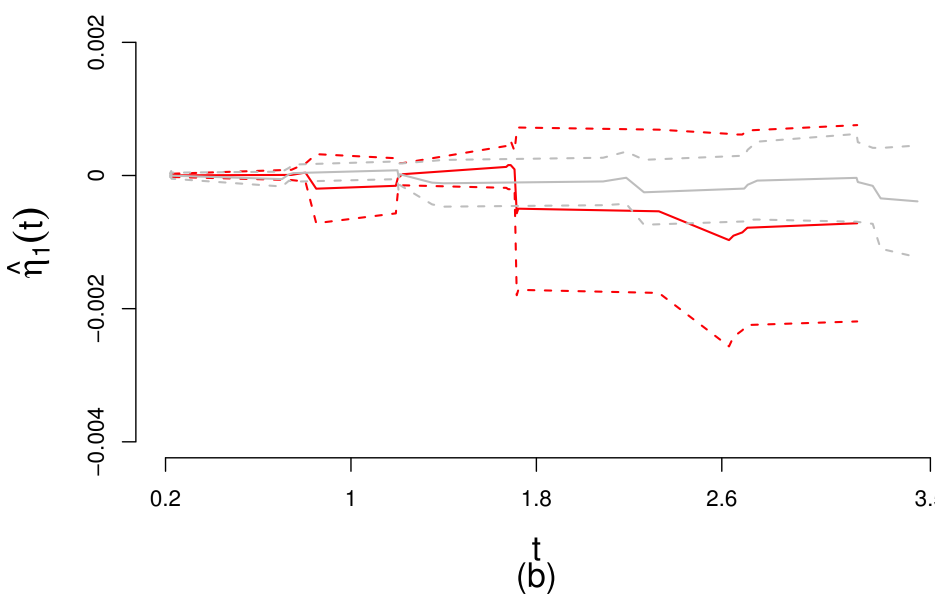


Figure 21: (a) and (b) show the comparison of the AIPW estimates of baseline cumulative coefficients $\eta_0(t)$ and cumulative coefficients $\eta_1(t)$ with 95% pointwise confidence intervals for the immune response R_i (IgG-92TH023V1V2) in model (3.38) for $\epsilon_{1_i} = 1$ and $\epsilon_{1_i} = 2$, respectively.

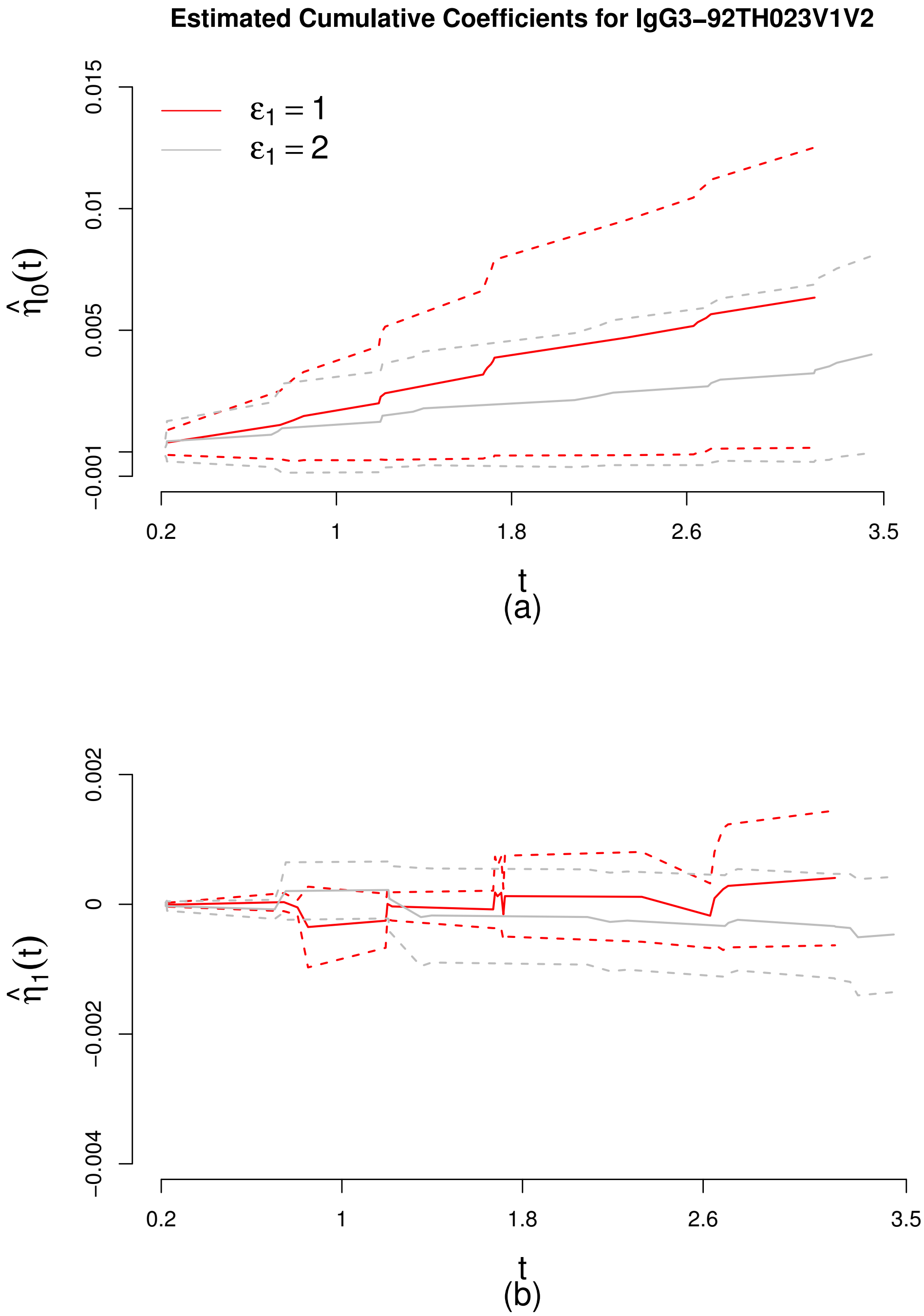


Figure 22: (a) and (b) show the comparison of the AIPW estimates of baseline cumulative coefficients $\eta_0(t)$ and cumulative coefficients $\eta_1(t)$ with 95% pointwise confidence intervals for the immune response R_i (IgG3-92TH023V1V2) in model (3.38) for $\epsilon_{1_i} = 1$ and $\epsilon_{1_i} = 2$, respectively.

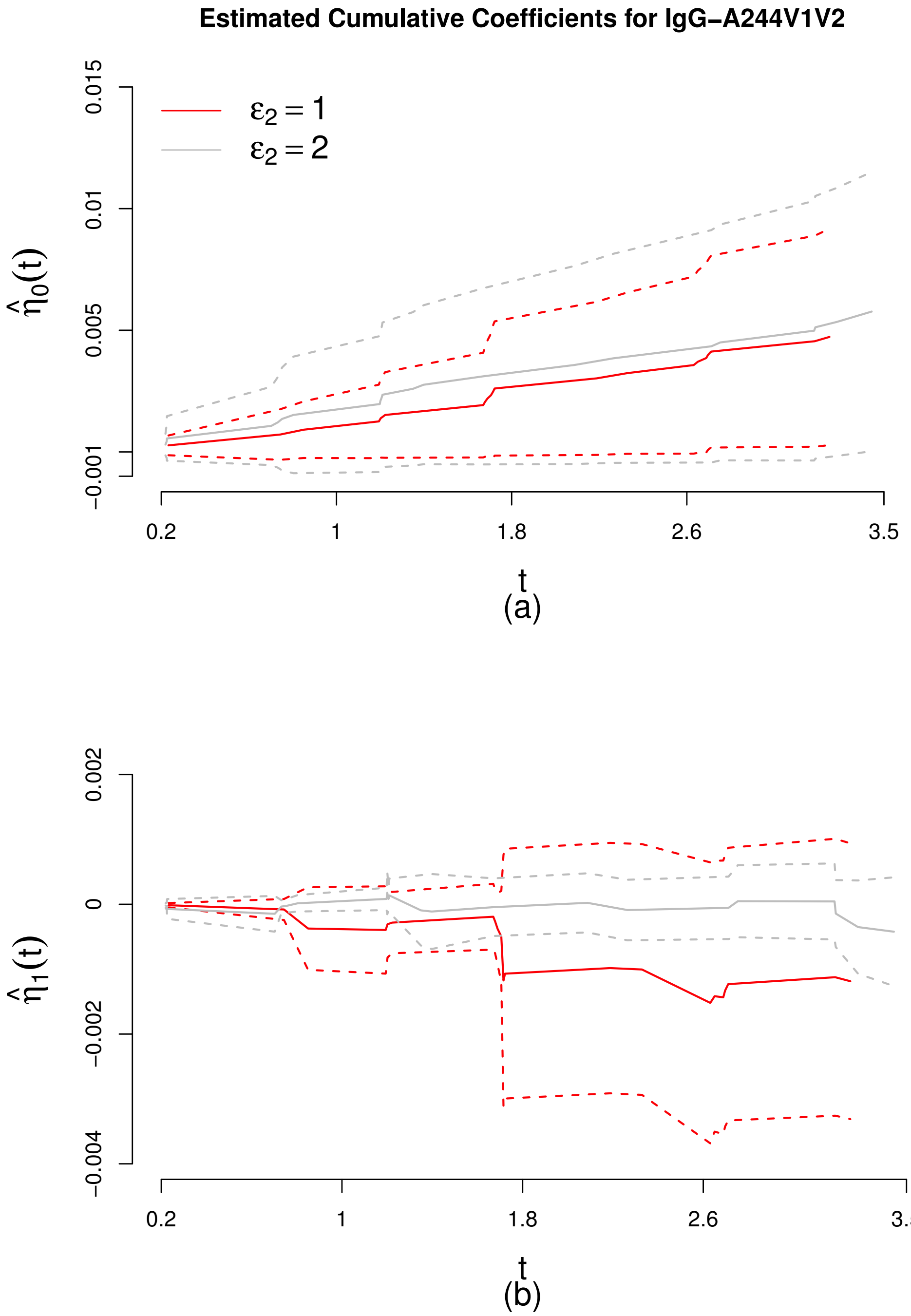


Figure 23: (a) and (b) show the comparison of the AIPW estimates of baseline cumulative coefficients $\eta_0(t)$ and cumulative coefficients $\eta_1(t)$ with 95% pointwise confidence intervals for the immune response R_i (IgG-A244V1V2) in model (3.38) for $\epsilon_{2_i} = 1$ and $\epsilon_{2_i} = 2$, respectively.

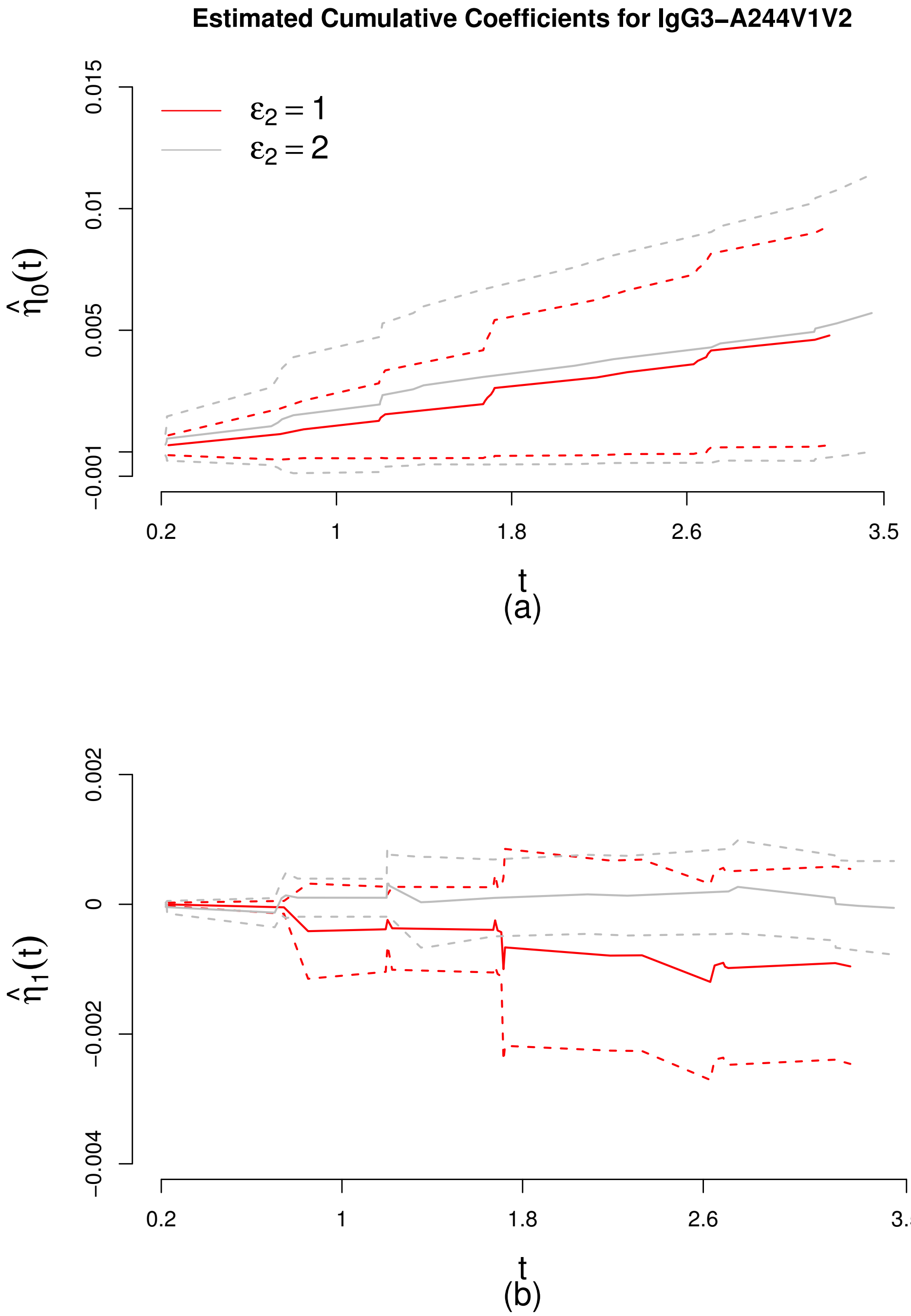


Figure 24: (a) and (b) show the comparison of the AIPW estimates of baseline cumulative coefficients $\eta_0(t)$ and cumulative coefficients $\eta_1(t)$ with 95% pointwise confidence intervals for the immune response R_i (IgG3-A244V1V2) in model (3.38) for $\epsilon_{2_i} = 1$ and $\epsilon_{2_i} = 2$, respectively.

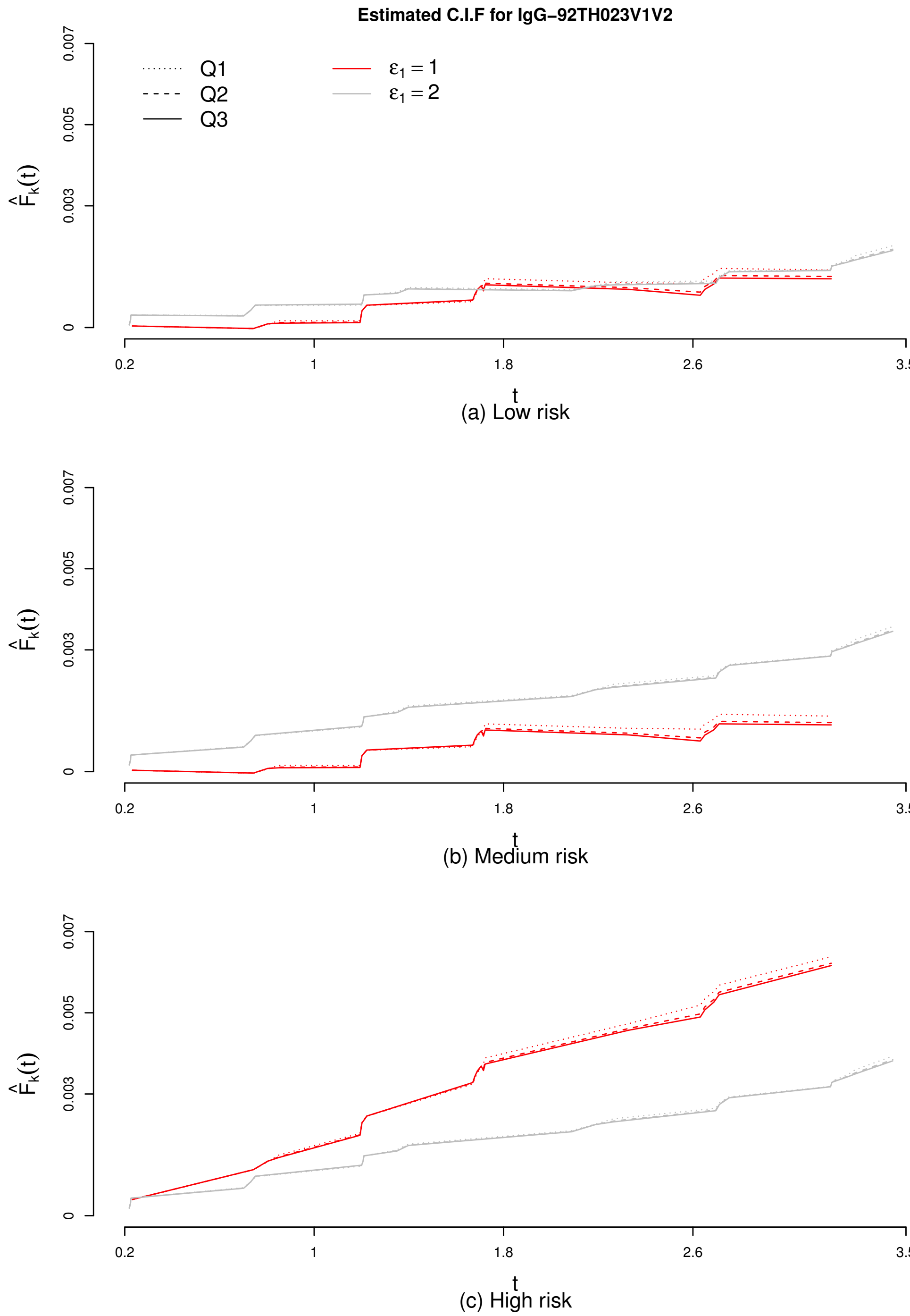


Figure 25: $Q_1 = 0.09027, Q_2 = 0.31310$ and $Q_3 = 0.39230$ are quartiles of the predicted immune response R_i (IgG-92TH023V1V2) using AIPW method. (a), (b) and (c) show the predicted cumulative incidence function $\hat{F}_k$ for $\epsilon_{1_i} = 1$ (red) and $\epsilon_{1_i} = 2$ (grey), respectively, at each level of behavioral risk score groups (low, medium and high) based on the model (3.38).

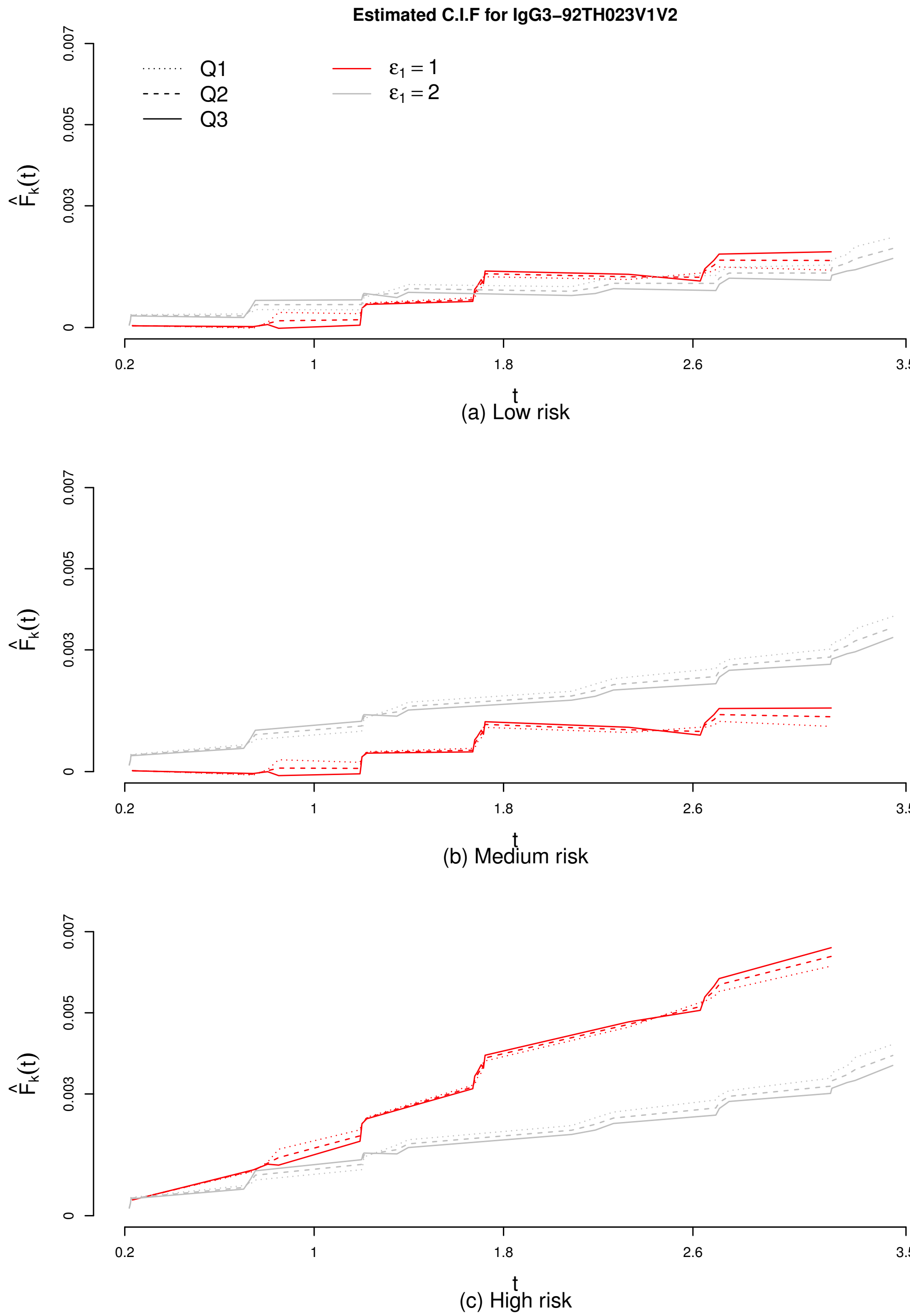


Figure 26: $Q_1 = -0.4677, Q_2 = 0.1196$ and $Q_3 = 0.6484$ are quartiles of the predicted immune response R_i (IgG3-92TH023V1V2) using AIPW method. (a), (b) and (c) show the predicted cumulative incidence function $\hat{F}_k$ for $\epsilon_{1_i} = 1$ (red) and $\epsilon_{1_i} = 2$ (grey), respectively, at each level of behavioral risk score groups (low, medium and high) based on the model (3.38).

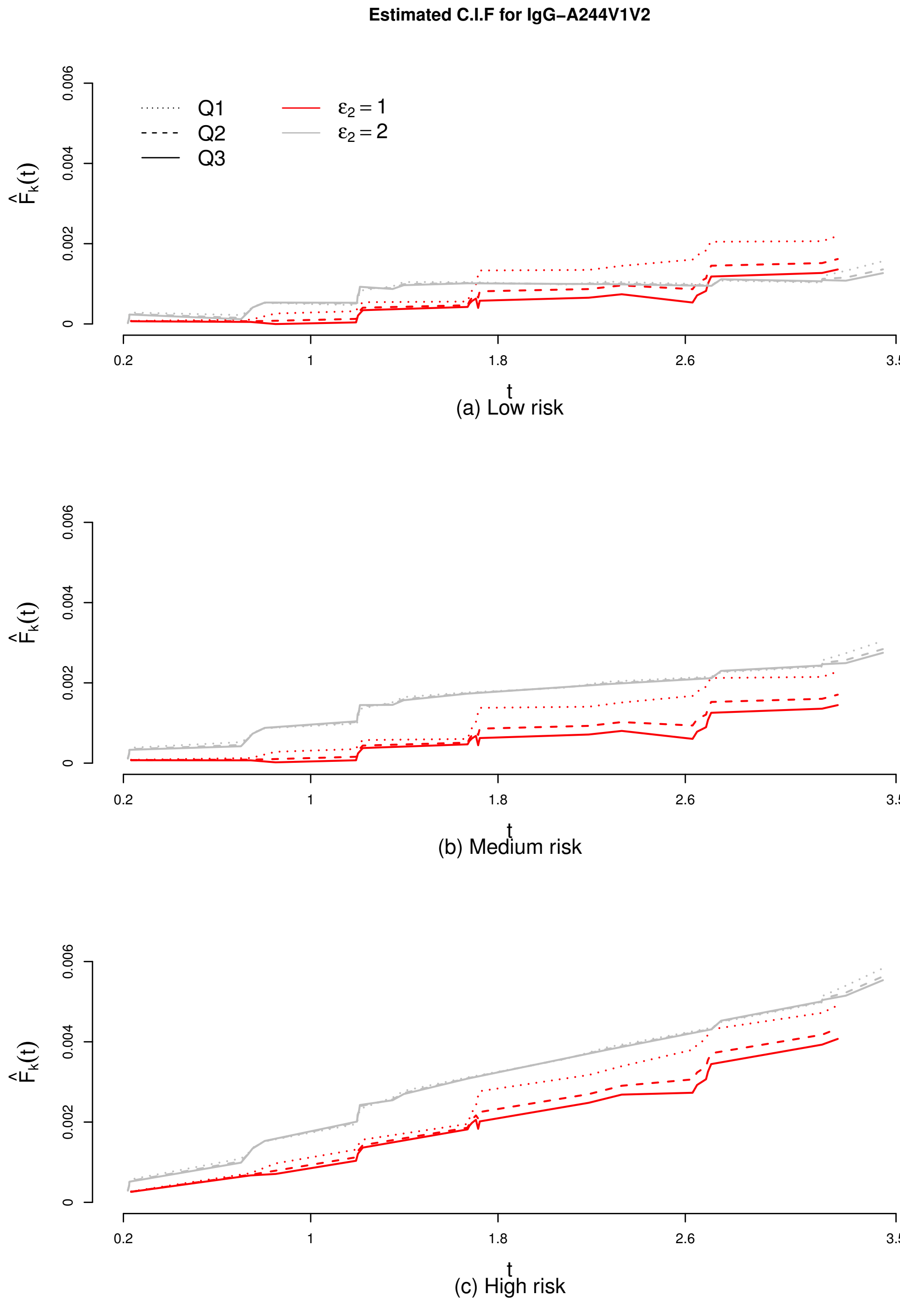


Figure 27: $Q_1 = -0.1530, Q_2 = 0.3321$ and $Q_3 = 0.5514$ are quartiles of the predicted immune response R_i (IgG-A244V1V2) using AIPW method. (a), (b) and (c) show the predicted cumulative incidence function $\hat{F}_k$ for $\epsilon_{1_i} = 1$ (red) and $\epsilon_{1_i} = 2$ (grey), respectively, at each level of behavioral risk score groups (low, medium and high) based on the model (3.38).

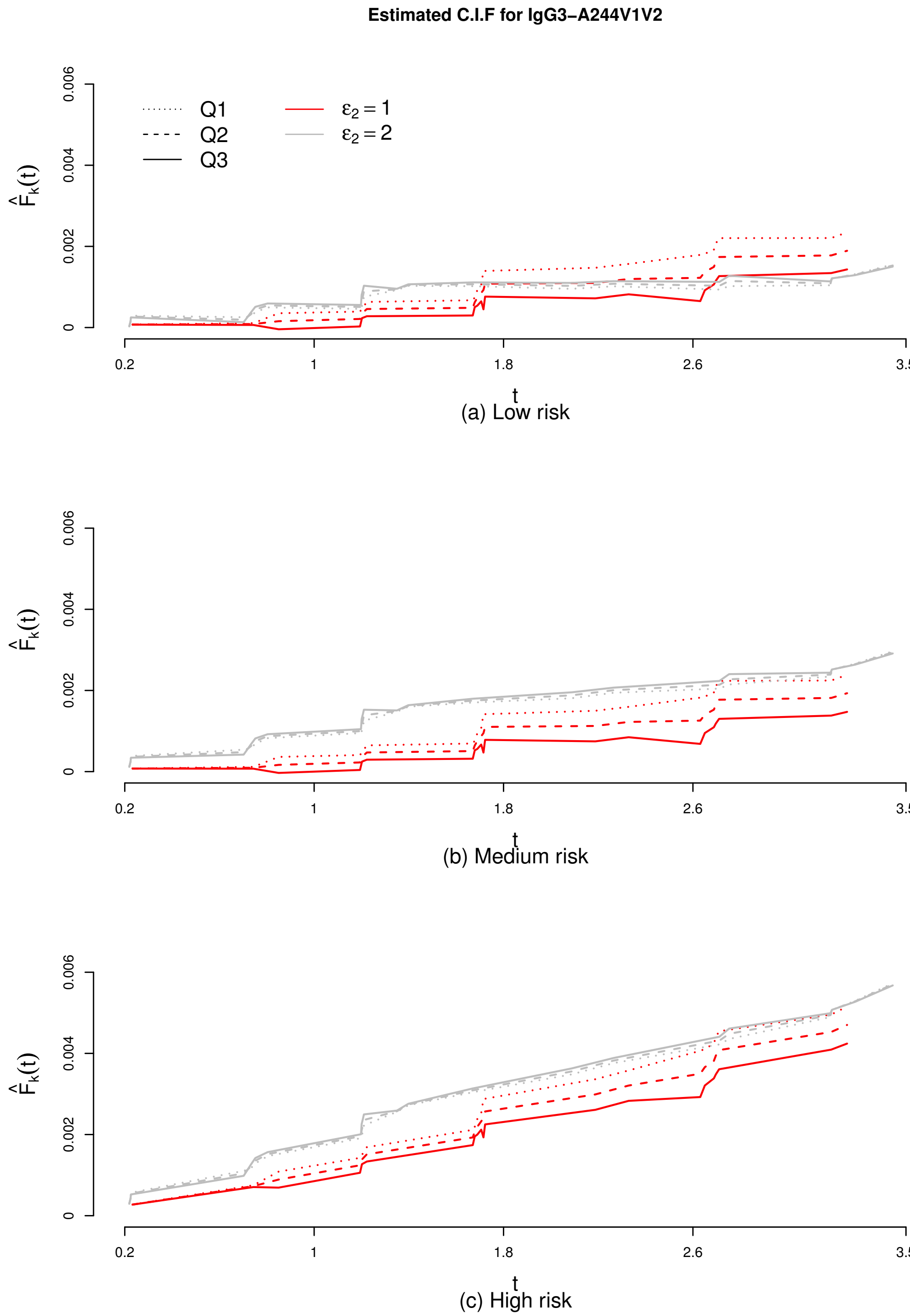


Figure 28: $Q_1 = -0.38510, Q_2 = 0.08807$ and $Q_3 = 0.56800$ are quartiles of the predicted immune response R_i (IgG3-A244V1V2) using AIPW method. (a), (b) and (c) show the predicted cumulative incidence function $\hat{F}_k$ for $\epsilon_{1_i} = 1$ (red) and $\epsilon_{1_i} = 2$ (grey), respectively, at each level of behavioral risk score groups (low, medium and high) based on the model (3.38).

CHAPTER 4: PROOFS OF THE THEOREMS

4.1 Proofs of the Theorems in Chapter 2

Condition I.

I.1. The regression function $\eta(t)$ is right-continuous with left-handed limits on $[0,\tau]$.

I.2. The link function $h(\cdot)$ is three times continuously differentiable and invertible, and $\partial h(\cdot)/\partial x$ is bounded away from zero. The function $w_i(t)$ is a possibly random weight and uniformly convergent in $t \in [0,\tau]$.

I.3. $\varphi(\mathcal{V}_i,\theta) = I(\tilde{\epsilon}_i = 1)\varphi_1(\mathcal{V}_i,\theta) + I(\tilde{\epsilon}_i \neq 1)\varphi_2(\mathcal{V}_i,\theta)$ is twice differentiable with respect to φ and $\varphi'(\mathcal{V}_i,\theta) = d\varphi(\mathcal{V}_i,\theta)/d\theta$ is uniformly bounded and bounded away from zero, i.e., $\varphi(\mathcal{V}_i,\theta_j) \geq \epsilon > 0$.

I.4. The estimator $\hat{\theta}$ satisfies $n^{\frac{1}{2}}(\hat{\theta}-\theta_0) = n^{-\frac{1}{2}}\left[J\left(\theta_0\right)\right]^{-1}\sum_{i=1}^{n} U(\mathcal{V}_i,\theta_0) + o_p(1)$, where $J\left(\theta_0\right)$ is the positive definite fisher information matrix and $U(\mathcal{V}_i,\theta_0)$, $i = 1,\ldots,n$, are identically independent distributed mean zero random variables.

I.5. The estimator $\widehat{G}(t)$ is asymptotically linear with influence function $\mathtt{IC}_G$ such that

$$n^{\frac{1}{2}}(\hat{G}-G)(t,x,z) = n^{-\frac{1}{2}}\sum_{i=1}^{n}\mathtt{IC}_G(t,x,z;Y_i) + o_p(1),$$

uniformly in (t,x,z).

I.6. Assume that $G(t)$ is continuous. If $\tau \in (0,\infty]$ is such that $Y(\tau) \xrightarrow{p} \infty$ as

$n \to \infty$, then $\sup_{0 \le t \le \tau} |\widehat{G}(t) - G(t)| \xrightarrow{p} 0$ as $n \xrightarrow{p} \infty$, where $Y(t) = I\{T \ge t\}$ at risk process.

Proof of Proposition 1

Let

$$\log \mathrm{L}_1(\theta_1) = \sum_{i=1}^{n} I(\tilde{\epsilon}_i = 1)\left[\xi_i\{\theta_1^\mathsf{T}\mathcal{V}_i\} - \log\{1 + \exp(\theta_1^\mathsf{T}\mathcal{V}_i)\}\right],$$
$$\log \mathrm{L}_2(\theta_2) = \sum_{i=1}^{n} I(\tilde{\epsilon}_i \ne 1)\left[\xi_i\{\theta_2^\mathsf{T}\mathcal{V}_i\} - \log\{1 + \exp(\theta_2^\mathsf{T}\mathcal{V}_i)\}\right].$$

From the likelihood defined in (2.1), we have the following log-likelihood function

$$\log \mathrm{L}(\theta) = \log \mathrm{L}_1(\theta_1) + \log \mathrm{L}_2(\theta_2).$$

By taking derivative with respect to $\theta = (\theta_1^\mathsf{T}, \theta_2^\mathsf{T})^\mathsf{T}$, the score functions are

$$U(\mathcal{V}_i, \theta) = \begin{pmatrix} U_1(\mathcal{V}_i, \theta_1) \\ U_2(\mathcal{V}_i, \theta_2) \end{pmatrix} = \begin{pmatrix} \frac{\partial \log \mathrm{L}(\theta)}{\partial \theta_1} \\ \frac{\partial \log \mathrm{L}(\theta)}{\partial \theta_2} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} I(\tilde{\epsilon}_i = 1)\left[\xi_i \mathcal{V}_i - \frac{\mathcal{V}_i \exp^{\theta_1^\mathsf{T}\mathcal{V}_i}}{1+\exp^{\theta_1^\mathsf{T}\mathcal{V}_i}}\right]^\mathsf{T} \\ \sum_{i=1}^{n} I(\tilde{\epsilon}_i \ne 1)\left[\xi_i \mathcal{V}_i - \frac{\mathcal{V}_i \exp^{\theta_2^\mathsf{T}\mathcal{V}_i}}{1+\exp^{\theta_2^\mathsf{T}\mathcal{V}_i}}\right]^\mathsf{T} \end{pmatrix},$$

which has zero at $\hat{\theta}$. The second-order partial derivatives of $\log \mathrm{L}(\theta)$ is

$$H(\theta) = \begin{pmatrix} -\sum_{i=1}^{n} I(\tilde{\epsilon}_i = 1)\frac{\exp^{\theta_1^\mathsf{T}\mathcal{V}_i} \mathcal{V}_i\mathcal{V}_i^\mathsf{T}}{(1+\exp^{\theta_1^\mathsf{T}\mathcal{V}_i})^2} & 0 \\ 0 & -\sum_{i=1}^{n} I(\tilde{\epsilon}_i \ne 1)\frac{\exp^{\theta_2^\mathsf{T}\mathcal{V}_i} \mathcal{V}_i\mathcal{V}_i^\mathsf{T}}{(1+\exp^{\theta_2^\mathsf{T}\mathcal{V}_i})^2} \end{pmatrix}, \qquad (4.1)$$

which is negative on $H(\theta)$. Moerover, the Jacobian of $H(\theta)$ is obviously positive. Thus, the log likelihood function $\log \mathrm{L}(\theta)$ has a local maximum at $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$. Therefore, the selection probability S_i can be estimated by its parametric model $\hat{S}_i = \varphi(\mathcal{V}_i, \hat{\theta})$.

Since $\hat{\theta}$ is maximum likelihood estimator of $\log \mathrm{L}(\theta)$, the $\hat{\theta}$ is consistent estimator of the true value θ_0. Moreover, by using Taylor series expansion,

$$U(\mathcal{V}_i, \hat{\theta}) = U(\mathcal{V}_i, \theta_0) + \frac{\partial}{\partial \theta} U(\mathcal{V}_i, \theta_0)(\hat{\theta} - \theta_0) + o_p(n^{-\frac{1}{2}}), \tag{4.2}$$

where $U(\mathcal{V}_i, \hat{\theta}) = 0$. By the standard arguments of asymptotic normality for M-estimators, it can be shown that we have the following asymptotic linear expression such that

$$\sqrt{n}(\hat{\theta} - \theta_0) = n^{-\frac{1}{2}} \left[J\left(\mathcal{V}_i, \theta_0\right)\right]^{-1} \sum_{i=1}^{n} U(\mathcal{V}_i, \theta_0) + o_p(1) \tag{4.3}$$

and its limiting distribution

$$\sqrt{n}(\hat{\theta} - \theta_0) \overset{d}{\to} N\left(0, J^{-1}(\mathcal{V}_i, \theta_0)\right), \tag{4.4}$$

where the fisher information matrix

$$J(\mathcal{V}_i, \theta_0) = \begin{pmatrix} E\left[I(\tilde{\epsilon}_i = 1) \frac{\exp^{\theta_{01}^{\mathsf{T}} \mathcal{V}_i} \mathcal{V}_i \mathcal{V}_i^{\mathsf{T}}}{(1+\exp^{\theta_{01}^{\mathsf{T}} \mathcal{V}_i})^2}\right] & 0 \\ 0 & E\left[I(\tilde{\epsilon}_i \neq 1) \frac{\exp^{\theta_{02}^{\mathsf{T}} \mathcal{V}_i} \mathcal{V}_i \mathcal{V}_i^{\mathsf{T}}}{(1+\exp^{\theta_{02}^{\mathsf{T}} \mathcal{V}_i})^2}\right] \end{pmatrix} \tag{4.5}$$

and

$$U(\mathcal{V}_i, \theta_0) = \begin{pmatrix} \sum_{i=1}^{n} I(\tilde{\epsilon}_i = 1) \left[\xi_i \mathcal{V}_i - \frac{\mathcal{V}_i \exp^{\theta_1^{\mathsf{T}} \mathcal{V}_i}}{1+\exp^{\theta_1^{\mathsf{T}} \mathcal{V}_i}}\right]^{\mathsf{T}} \\ \sum_{i=1}^{n} I(\tilde{\epsilon}_i \neq 1) \left[\xi_i \mathcal{V}_i - \frac{\mathcal{V}_i \exp^{\theta_2^{\mathsf{T}} \mathcal{V}_i}}{1+\exp^{\theta_2^{\mathsf{T}} \mathcal{V}_i}}\right]^{\mathsf{T}} \end{pmatrix}.$$

□

Proof of Proposition 2

(1) By the corollary 3.2.1 of (Fleming and Harrington, 2013), we have $G(t) > 0$

such that, for any $t \in [0, \tau]$,

$$
\begin{aligned}
n^{\frac{1}{2}}(\widehat{G}(t) - G(t)) &= n^{\frac{1}{2}} \left\{ -G(t) \int_0^\tau \frac{\widehat{G}(s-)}{G(s)} I(Y(s) > 0) \frac{dM(s)}{Y(s)} \right\} + o_p(1), \\
&= n^{\frac{1}{2}} \left\{ -G(t) \int_0^\tau I(Y(s) > 0) \frac{dM(s)}{Y(s)} \right\} \\
&\quad + n^{\frac{1}{2}} \left\{ G(t) \int_0^\tau \left(1 - \frac{\widehat{G}(s-)}{G(s)} \right) I(Y(s) > 0) \frac{dM(s)}{Y(s)} \right\} \\
&\quad + o_p(1). \qquad (4.6)
\end{aligned}
$$

By Lenglart's inequality, the second term of (4.6) is

$$
\begin{aligned}
& P\left(\sup_{0 \le t \le \tau} \left\{ n^{\frac{1}{2}} G(t) \int_0^\tau \left(1 - \frac{\widehat{G}(s-)}{G(s)} \right) I(Y(s) > 0) \frac{dM(s)}{Y(s)} \right\}^2 \ge \epsilon \right) \\
\le\ & \frac{\eta}{\epsilon} + P\left\{ nG^2(t) \int_0^\tau \left(1 - \frac{\widehat{G}(s-)}{G(s)} \right)^2 \frac{I(Y(s) > 0)}{Y^2(s)} Y(s) d\Lambda(s) \ge \eta \right\} \\
\le\ & \frac{\eta}{\epsilon} + P\left\{ G^2(t) \frac{nI(Y(\tau) > 0)\Lambda(\tau)}{Y(\tau)} \ge \eta \right\} + P(Y(\tau) = 0) \\
\le\ & \frac{\eta}{\epsilon} \qquad (4.7)
\end{aligned}
$$

since $Y(\tau) \xrightarrow{p} \infty$ as $n \xrightarrow{p} \infty$.

Therefore, we have

$$
\begin{aligned}
n^{\frac{1}{2}}(\widehat{G}(t) - G(t)) &= n^{\frac{1}{2}} \left\{ -G(t) \int_0^\tau I(Y(s) > 0) \frac{dM(s)}{Y(s)} \right\} + o_p(1) \\
&= n^{-\frac{1}{2}} \left\{ -G(t) \int_0^\tau I(Y(s) > 0) \frac{dM(s)}{y(s)} \right\} + o_p(1), \qquad (4.8)
\end{aligned}
$$

where $n^{-1}Y(s) = n^{-1} \sum_{i=1}^n Y_i(s) \xrightarrow{p} y(s)$ with $s \in [0, \tau]$.

(2) It is easy to derive when the censoring time follows the Cox model with hazard function $\lambda(t) = \lambda_0(t) \exp(\beta_0 X_i + \beta_1 Z_i)$ where baseline $\lambda_0(t)$ and possibly time dependent covariates X_i and Z_i. □

Proof of Theorem 2.1

Let $\boldsymbol{D}_{\boldsymbol{\eta}}(t)$, $\boldsymbol{D}_{\boldsymbol{\gamma}}(t)$, $\boldsymbol{R}(t)$, $\boldsymbol{F}_1(t)$, $\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})$, $\boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta})$, $\boldsymbol{H}(t,\hat{\theta})$, $\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\hat{\theta})$ are all evaluated at the true value $\{\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0\}$ of $\{\boldsymbol{\eta}(t),\boldsymbol{\gamma}\}$, where $\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})$, $\boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta})$ $\boldsymbol{H}(t,\hat{\theta})$, $\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\hat{\theta})$ are corresponding terms for (2.12). By (2.10), we have

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0) \;=\; \{n^{-1}\,\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})\}^{-1}n^{-\frac{1}{2}}\boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta})+o_p(1). \tag{4.9}$$

Let $\boldsymbol{A}(\theta)=\partial\boldsymbol{\Psi}(\theta)/\partial\theta$. The Taylor expansion of $\boldsymbol{\Psi}(\hat{\theta})$ around the true value θ_0 is

$$\boldsymbol{\Psi}(\hat{\theta})=\boldsymbol{\Psi}(\theta_0)+\boldsymbol{A}(\theta_0)(\hat{\theta}-\theta_0)+o_p(n^{-\frac{1}{2}}). \tag{4.10}$$

Let $\boldsymbol{K}(t,\hat{\theta}) = \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\eta}}(t)\left[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\hat{\theta})\right]^{-1}$ and evaluate at the true values $\{\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0\}$. By plugging $\boldsymbol{H}(t,\hat{\theta})$ in the formula for $\boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta})$ in (2.12), we have

$$\begin{aligned}
&n^{-\frac{1}{2}}\boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta})\\
&=n^{-\frac{1}{2}}\int_0^\tau \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\,\{I\\
&\qquad\qquad -\boldsymbol{D}_{\boldsymbol{\eta}}(t)\left[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\hat{\theta})\right]^{-1}\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\Big\}\,\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t)\}\;dt\\
&=n^{-\frac{1}{2}}\int_0^\tau \Big\{\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma}}(t)-\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\eta}}(t)\\
&\qquad\qquad \times\left[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\hat{\theta})\right]^{-1}\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta}}(t)\Big\}\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\,\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t)\}\,dt\\
&=n^{-\frac{1}{2}}\int_0^\tau \Big\{\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma}}(t)-\boldsymbol{K}(t,\hat{\theta})\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta}}(t)\Big\}\,\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\,\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t)\}\;dt.
\end{aligned} \tag{4.11}$$

By plugging (4.10) into (4.11) and decomposing $\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t)\}$ into $\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)}-\frac{\Delta_i N_i(t)}{G(T_i)}+\frac{\Delta_i N_i(t)}{G(T_i)}-\boldsymbol{F}_{1i}(t)\right\}$, then (4.11) can be split into the following four terms

$$
\begin{aligned}
& n^{-\frac{1}{2}} \boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta}) \\
= \; & n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t) - \boldsymbol{K}(t,\hat{\theta}) \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) \right] w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t) \right\} dt \\
& + n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t) - \boldsymbol{K}(t,\hat{\theta}) \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) \right] w_i(t) \\
& \qquad \times \boldsymbol{A}_i(\theta_0)(\hat{\theta} - \theta_0) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t) \right\} dt \\
& + n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t) - \boldsymbol{K}(t,\hat{\theta}) \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) \right] w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{ \frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - \frac{\Delta_i N_i(t)}{G(T_i)} \right\} dt \\
& + n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t) - \boldsymbol{K}(t,\hat{\theta}) \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) \right] w_i(t) \\
& \qquad \times \boldsymbol{A}_i(\theta_0)(\hat{\theta} - \theta_0) \left\{ \frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - \frac{\Delta_i N_i(t)}{G(T_i)} \right\} dt \\
& + o_p(1). \qquad (4.12)
\end{aligned}
$$

The fourth term of (4.12) is shown to be equal to $o_p(1)$ in Appendix A. That is,

$$
\begin{aligned}
& n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t) - \boldsymbol{K}(t,\hat{\theta}) \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) \right] w_i(t) \\
& \qquad \times \boldsymbol{A}_i(\theta_0)(\hat{\theta} - \theta_0) \left\{ \frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - \frac{\Delta_i N_i(t)}{G(T_i)} \right\} dt \\
& = o_p(1). \qquad (4.13)
\end{aligned}
$$

Denote the first term of (4.12) by

$$
\widetilde{\boldsymbol{B}}_{\boldsymbol{\gamma}}(\hat{\theta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t) - \boldsymbol{K}(t,\hat{\theta}) \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) \right] w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t) \right\} dt.
$$

It is shown in the Appendix A that

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^\tau \left[\boldsymbol{K}(t,\hat{\theta}) - \boldsymbol{k}(t,\theta_0)\right] \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\} dt = o_p(1). \tag{4.14}$$

Let

$$\boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t,\theta) = \left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t) - \boldsymbol{k}(t,\theta)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right] w_i(t) \boldsymbol{\psi}_i(\theta) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\}. \tag{4.15}$$

It follows by (4.14) that

$$\begin{aligned} \widetilde{\boldsymbol{B}}_{\boldsymbol{\gamma}}(\hat{\theta}) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^\tau \left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t) - \boldsymbol{k}(t,\theta_0)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right] w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\} dt \\ &- n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^\tau \left[\boldsymbol{K}(t,\hat{\theta}) - \boldsymbol{k}(t,\theta_0)\right] \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\} dt \\ &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^\tau \boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t,\theta_0)\, dt + o_p(1). \end{aligned} \tag{4.16}$$

Denote the second term of (4.12) by

$$\begin{aligned} \boldsymbol{D}_{\boldsymbol{\gamma}}(\hat{\theta}) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^\tau \left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t) - \boldsymbol{K}(t,\hat{\theta})\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right] w_i(t) \\ &\quad \times \boldsymbol{A}_i(\theta_0) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\} dt\, (\hat{\theta} - \theta_0). \end{aligned}$$

By similar argument in (4.14), we have

$$\begin{aligned} \boldsymbol{D}_{\boldsymbol{\gamma}}(\hat{\theta}) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^\tau \left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t) - \boldsymbol{k}(t,\theta_0)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right] w_i(t) \\ &\quad \times \boldsymbol{A}_i(\theta_0) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\} dt\, (\hat{\theta} - \theta_0) \\ &+ o_p(1). \end{aligned}$$

By the law of large numbers,

$$n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right]w_i(t)\boldsymbol{A}_i(\theta_0)\left\{\frac{\Delta_i N_i(t)}{G(T_i)}-\boldsymbol{F}_{1i}(t)\right\}dt\xrightarrow{p}g(\tau,\theta_0). \tag{4.17}$$

It follows by (2.15) in proposition 1 and (4.17) that

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{\gamma}}(\hat{\theta}) &= n^{\frac{1}{2}}(\hat{\theta}-\theta_0)\left(g(\tau,\theta_0)+o_p(1)\right)+o_p(1). \\ &= g(\tau,\theta_0)n^{\frac{1}{2}}(\hat{\theta}-\theta_0)+o_p(1). \\ &= g(\tau,\theta_0)\left\{n^{-\frac{1}{2}}J^{-1}(\mathcal{V}_i,\theta_0)\sum_{i=1}^{n}U(\mathcal{V}_i,\theta_0)\right\}+o_p(1). \end{aligned} \tag{4.18}$$

Now consider the third term of (4.12). Let

$$\begin{aligned}\Delta_{\boldsymbol{\gamma}}(\hat{\theta}) &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^{\tau}\left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t)-\boldsymbol{K}(t,\hat{\theta})\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right]w_i(t) \\ &\qquad\times\boldsymbol{\psi}_i(\theta_0)\Delta_i N_i(t)\left\{\frac{G(T_i)-\widehat{G}(T_i)}{\widehat{G}(T_i)G(T_i)}\right\}dt.\end{aligned}$$

With similar arguments in (4.14), we have

$$\begin{aligned}\Delta_{\boldsymbol{\gamma}}(\hat{\theta}) &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^{\tau}\left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right]w_i(t) \\ &\qquad\times\boldsymbol{\psi}_i(\theta_0)\Delta_i N_i(t)\left\{\frac{G(T_i)-\widehat{G}(T_i)}{\widehat{G}(T_i)G(T_i)}\right\}dt+o_p(1)\end{aligned} \tag{4.19}$$

Using (2.17) in proposition 2, we have

$$n^{\frac{1}{2}}\Delta_i N_i(t)\frac{\widehat{G}(T_i)-G(T_i)}{\widehat{G}(T_i)G(T_i)}=n^{-\frac{1}{2}}\Delta_i N_i(t)\frac{-\mathcal{I}(\widetilde{T}_i\le t)}{G(T_i)}\sum_{j=1}^{n}\int_0^{\tau}\mathcal{I}(s\le\widetilde{T}_i)\frac{dM_j{}^c(s)}{y(s)}+o_p(1). \tag{4.20}$$

By plugging (4.20) into (4.19), we have

$$\begin{aligned}\Delta_{\boldsymbol{\gamma}}(\hat{\theta}) = n^{-\frac{1}{2}} \sum_{j=1}^{n} \int_0^\tau \int_0^\tau n^{-1} \sum_{i=1}^{n} \Big\{ \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t) - \boldsymbol{k}(t,\theta_0) \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) \Big\} w_i(t) \\ \times \boldsymbol{\psi}_i(\theta_0) \frac{\Delta_i N_i(t)}{G(T_i)} \mathcal{I}(s \le \widetilde{T}_i \le t)\, dt \frac{dM_j^c(s)}{y(s)} \\ + o_p(1).\end{aligned}$$

By law of large numbers,

$$n^{-1} \sum_{j=1}^{n} \Big\{ \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},j}(t) - \boldsymbol{k}(t,\theta_0) \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},j}(t) \Big\} w_j(t) \boldsymbol{\psi}_j(\theta_0) \frac{\Delta_j N_j(t)}{G(T_j)} \mathcal{I}(s \le \widetilde{T}_j \le t) \xrightarrow{p} \boldsymbol{q}_{\boldsymbol{\gamma}}(s,t,\theta_0).$$

Let

$$\boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t,\theta) = \int_0^\tau \frac{\boldsymbol{q}_{\boldsymbol{\gamma}}(s,t,\theta)}{y(s)} dM_i^c(s). \tag{4.21}$$

Thus, we have

$$\Delta_{\boldsymbol{\gamma}}(\hat{\theta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^\tau \boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t,\theta_0) dt + o_p(1). \tag{4.22}$$

Let

$$\begin{aligned}\widetilde{\boldsymbol{B}}_{\boldsymbol{\gamma}}(\theta) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^\tau \boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t,\theta)\, dt, \\ \Delta_{\boldsymbol{\gamma}}(\theta) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^\tau \boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t,\theta)\, dt, \\ \boldsymbol{D}_{\boldsymbol{\gamma}}(\theta) &= g(\tau,\theta) \left\{ n^{-\frac{1}{2}} J^{-1}(\mathcal{V}_i,\theta) \sum_{i=1}^{n} U(\mathcal{V}_i,\theta) \right\}.\end{aligned} \tag{4.23}$$

It follows by (4.12), (4.13), (4.16), (4.18), (4.22) and (4.23) that

$$n^{-\frac{1}{2}} \boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta}) = \widetilde{\boldsymbol{B}}_{\boldsymbol{\gamma}}(\theta_0) + \Delta_{\boldsymbol{\gamma}}(\theta_0) + \boldsymbol{D}_{\boldsymbol{\gamma}}(\theta_0) + o_p(1). \tag{4.24}$$

By plugging the expression of $\boldsymbol{H}(t,\hat{\theta})$ into $\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})$ from (2.12), we have

$$n^{-1}\,\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta}) \;=\; \int_a^\tau \left[\boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t) - \boldsymbol{K}(t,\hat{\theta})\boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\right]\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\,dt. \tag{4.25}$$

Let

$$\begin{aligned}
n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta) &= n^{-1}\sum_{i=1}^{n}\int_a^\tau \left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t) - \boldsymbol{k}(t,\theta)\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right] w_i(t)\boldsymbol{\psi}_i(\theta)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt\\
\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta) &= E\left\{\int_a^\tau \left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t) - \boldsymbol{k}(t,\theta)\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right] w_i(t)\boldsymbol{\psi}_i(\theta)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt\right\}.
\end{aligned}$$

It is shown in the Appendix A that

$$\begin{aligned}
n^{-1}\,\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta}) &= n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0) + o_p(1)\\
&\xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0).
\end{aligned} \tag{4.26}$$

Let

$$\boldsymbol{W}_{\gamma,i}(\tau,\theta) = \int_0^\tau \boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t,\theta)\,dt + \int_0^\tau \boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t,\theta)\,dt + g(\tau,\theta)\left[J\left(\mathcal{V}_i,\theta\right)\right]^{-1}U(\mathcal{V}_i,\theta).$$

It follows by (4.9), (4.24), (4.26) that $\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ is asymptotically equivalent to the following identically independent distributed decomposition

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = \left\{\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0)\right\}^{-1} n^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{\boldsymbol{W}_{\gamma,i}(\tau,\theta_0)\right\} + o_p(1). \tag{4.27}$$

Since $\boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t,\theta_0)$ has mean zero by (2.2), $\boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t,\theta_0)$ is mean zero local square martingale, and the score function $U(\mathcal{V}_i,\theta_0)$ has mean zero, by the law of large numbers, $n^{-1}\sum_{i=1}^{n}\left\{\boldsymbol{W}_{\gamma,i}(\tau,\theta_0)\right\}$ has mean zero. By the central limit theorem,
$n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{W}_{\gamma,i}(\tau,\theta_0)$ converges in distribution to a mean zero normal random vector with covariance matrix $E\left\{\boldsymbol{W}_{\gamma,i}(\tau,\theta_0)\right\}^{\otimes 2}$.

By slutsky's theorem, we have

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\Sigma}_\gamma\right), \tag{4.28}$$

where $\boldsymbol{\Sigma}_\gamma = \boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0)^{-1}\, E\left\{\boldsymbol{W}_{\gamma,i}(\tau,\theta_0)\right\}^{\otimes 2} \boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0)^{-1}$.

Let $\widehat{\boldsymbol{F}}_{1i}(t)$, $\widehat{\boldsymbol{D}}_{\boldsymbol{\eta},i}(t)$ and $\widehat{\boldsymbol{D}}_{\boldsymbol{\gamma},i}(t)$ be the estimator of $\boldsymbol{F}_{1i}(t)$, $\boldsymbol{D}_{\boldsymbol{\eta},i}(t)$ and $\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)$ by plugging estimators $\widehat{\boldsymbol{\eta}}(t)$ and $\widehat{\boldsymbol{\gamma}}$ into $F_{1i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})$, $\boldsymbol{D}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})$, $\boldsymbol{D}_{\boldsymbol{\gamma},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma})$, respectively, and let $\boldsymbol{A}_i(\hat{\theta}) = \partial\psi_i(\hat{\theta})/\partial\theta$ where $\psi_i(\hat{\theta}) = \xi_i/\varphi(\mathcal{V}_i,\hat{\theta})$.

Let

$$\begin{aligned}
\widehat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\eta}}(t,\theta) &= n^{-1}\sum_{i=1}^{n} \widehat{\boldsymbol{D}}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \boldsymbol{\psi}_i(\theta) \widehat{\boldsymbol{D}}_{\boldsymbol{\eta},i}(t),\\
\widehat{\boldsymbol{K}}(t,\theta) &= n^{-1}\sum_{i=1}^{n} \widehat{\boldsymbol{D}}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t) w_i(t) \boldsymbol{\psi}_i(\theta) \widehat{\boldsymbol{D}}_{\boldsymbol{\eta},i}(t) \left[\widehat{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\eta}}(t,\theta)\right]^{-1},\\
\widehat{\boldsymbol{\zeta}}_{\boldsymbol{\gamma},i}(t,\theta) &= \left[\widehat{\boldsymbol{D}}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t) - \widehat{\boldsymbol{K}}(t,\theta)\widehat{\boldsymbol{D}}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right] w_i(t) \boldsymbol{\psi}_i(\theta) \left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - \widehat{\boldsymbol{F}}_{1i}(t)\right\},\\
\widehat{\boldsymbol{q}}_{\boldsymbol{\gamma}}(s,t,\theta) &= n^{-1}\sum_{j=1}^{n} \left\{\widehat{\boldsymbol{D}}_{\boldsymbol{\gamma},j}^{\mathsf{T}}(t) - \widehat{\boldsymbol{K}}(t,\theta)\widehat{\boldsymbol{D}}_{\boldsymbol{\eta},j}^{\mathsf{T}}(t)\right\} w_j(t) \boldsymbol{\psi}_j(\theta) \frac{\Delta_j N_j(t)}{\widehat{G}(T_j)} \mathcal{I}(s \le \widetilde{T}_j \le t),\\
\hat{y}(t) &= n^{-1}\sum_{i=1}^{n} \mathcal{I}(\widetilde{T}_i \ge t),\\
\widehat{M_j}^{c}(s) &= \mathcal{I}(\widetilde{T}_j \le s, \Delta_j = 0) - \int_0^s \mathcal{I}(\widetilde{T}_j \ge u)\ d(-\log \widehat{G}(u)),\\
\widehat{\boldsymbol{\kappa}}_{\boldsymbol{\gamma},i}(t,\theta) &= \int_0^\tau \frac{\widehat{\boldsymbol{q}}_{\boldsymbol{\gamma}}(s,t,\theta)}{\hat{y}(s)} d\widehat{M}_i^c(s),\\
\hat{g}(\tau,\theta) &= n^{-1}\sum_{i=1}^{n}\int_0^\tau \left[\widehat{\boldsymbol{D}}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t) - \widehat{\boldsymbol{K}}(t,\theta)\widehat{\boldsymbol{D}}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right] w_i(t)\\
&\qquad\qquad \times \boldsymbol{A}_i(\theta) \left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - \widehat{\boldsymbol{F}}_{1i}(t)\right\} dt.
\end{aligned} \tag{4.29}$$

The asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ can be consistently estimated by

$$\widehat{\boldsymbol{\Sigma}}_\gamma = \widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}^{-1}(\hat{\theta})\, n^{-1} \sum_{i=1}^{n} \left\{ \widehat{\boldsymbol{W}}_{\gamma,i}(\tau, \hat{\theta}) \right\}^{\otimes 2} \widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}^{-1}(\hat{\theta}),$$

where

$$\begin{aligned} \widehat{\boldsymbol{W}}_{\gamma,i}(\tau,\theta) &= \int_0^\tau \widehat{\boldsymbol{\zeta}}_{\boldsymbol{\gamma},i}(t,\theta)\, dt + \int_0^\tau \widehat{\boldsymbol{\kappa}}_{\boldsymbol{\gamma},i}(t,\theta)\, dt + \hat{g}(\tau,\theta) \left[\widehat{J}(\theta)\right]^{-1} U(\mathcal{V}_i, \theta), \\ \widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}(\theta) &= n^{-1} \sum_{i=1}^{n} \int_0^\tau \left[\widehat{\boldsymbol{D}}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t) - \widehat{\boldsymbol{K}}(t,\theta)\widehat{\boldsymbol{D}}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right] w_i(t) \boldsymbol{\psi}_i(\theta) \widehat{\boldsymbol{D}}_{\boldsymbol{\gamma},i}(t) dt. \end{aligned} \quad (4.30)$$

□

Proof of Theorem 2.2.

From (2.11), we have

$$\begin{aligned} \sqrt{n}(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t)) &= \left[n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\hat{\theta})\right]^{-1} n^{-\frac{1}{2}} \boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t) \boldsymbol{W}(t) \boldsymbol{\Psi}(\hat{\theta}) \{\boldsymbol{R}(t) - \boldsymbol{F}_1(t) \\ &\quad -\boldsymbol{D}_{\boldsymbol{\gamma}}(t) \left\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})\right\}^{-1} \boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta}) \Big\} + o_p(1). \end{aligned} \quad (4.31)$$

We consider the expression

$$n^{-\frac{1}{2}} \boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t) \boldsymbol{W}(t) \boldsymbol{\Psi}(\hat{\theta}) \left\{ \boldsymbol{R}(t) - \boldsymbol{F}_1(t) - \boldsymbol{D}_{\boldsymbol{\gamma}}(t) \left\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})\right\}^{-1} \boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta}) \right\}. \quad (4.32)$$

It can be decomposed into two terms

$$\begin{aligned} & n^{-\frac{1}{2}} \boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t) \boldsymbol{W}(t) \boldsymbol{\Psi}(\hat{\theta}) \left\{ \boldsymbol{R}(t) - \boldsymbol{F}_1(t) - \boldsymbol{D}_{\boldsymbol{\gamma}}(t) \left\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})\right\}^{-1} \boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta}) \right\} \\ &= n^{-\frac{1}{2}} \boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t) \boldsymbol{W}(t) \boldsymbol{\Psi}(\hat{\theta}) \{\boldsymbol{R}(t) - \boldsymbol{F}_1(t)\} \\ &\quad -n^{-\frac{1}{2}} \boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t) \boldsymbol{W}(t) \boldsymbol{\Psi}(\hat{\theta}) \boldsymbol{D}_{\boldsymbol{\gamma}}(t) \left\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})\right\}^{-1} \boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta}). \end{aligned} \quad (4.33)$$

The first term of (4.33) can be decomposed into four terms

$$
\begin{aligned}
&n^{-\frac{1}{2}}\boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\left\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t)\right\}\\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)w_i(t)\boldsymbol{\psi}_i(\theta_0)\left\{\frac{\Delta_i N_i(t)}{G(T_i)}-\boldsymbol{F}_{1i}(t)\right\}\\
&\quad +n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)w_i(t)\boldsymbol{\psi}_i(\theta_0)\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)}-\frac{\Delta_i N_i(t)}{G(T_i)}\right\}\\
&\quad +n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)w_i(t)(\boldsymbol{\psi}_i(\hat{\theta})-\boldsymbol{\psi}_i(\theta_0))\left\{\frac{\Delta_i N_i(t)}{G(T_i)}-\boldsymbol{F}_{1i}(t)\right\}\\
&\quad +n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)w_i(t)(\boldsymbol{\psi}_i(\hat{\theta})-\boldsymbol{\psi}_i(\theta_0))\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)}-\frac{\Delta_i N_i(t)}{G(T_i)}\right\}\\
&\quad +o_p(1). \qquad (4.34)
\end{aligned}
$$

It is shown in the Appendix A that the third and the fourth term are $o_p(1)$, that is,

$$
n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)w_i(t)(\boldsymbol{\psi}_i(\hat{\theta})-\boldsymbol{\psi}_i(\theta_0))\left\{\frac{\Delta_i N_i(t)}{G(T_i)}-\boldsymbol{F}_{1i}(t)\right\}=o_p(1), \qquad (4.35)
$$

$$
n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)w_i(t)(\boldsymbol{\psi}_i(\hat{\theta})-\boldsymbol{\psi}_i(\theta_0))\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)}-\frac{\Delta_i N_i(t)}{G(T_i)}\right\}=o_p(1). \qquad (4.36)
$$

It follows that

$$
\begin{aligned}
&n^{-\frac{1}{2}}\boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\left\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t)\right\}\\
&= n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{\zeta}_{\boldsymbol{\eta},i}(t,\theta_0)+n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{\kappa}_{\boldsymbol{\eta},i}(t,\theta_0)+o_p(1), \qquad (4.37)
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\zeta}_{\boldsymbol{\eta},i}(t,\theta) &= \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)w_i(t)\boldsymbol{\psi}_i(\theta)\left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\},\\
\boldsymbol{q}_{\boldsymbol{\eta}}(s,t,\theta) &= E\left\{\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},j}(t)w_j(t)\boldsymbol{\psi}_j(\theta)\frac{\Delta_j N_j(t)}{G(T_j)}\mathcal{I}(s \le \tilde{T}_j \le t)\right\},\\
\boldsymbol{\kappa}_{\boldsymbol{\eta},i}(t,\theta) &= \left\{\int_0^{\tau}\frac{\boldsymbol{q}_{\boldsymbol{\eta}}(s,t,\theta)}{y(s)}dM_i^c(s)\right\},\\
y(s) &= \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\mathcal{I}(\widetilde{T}_i \ge s), \quad \text{where } s\in[0,\tau].
\end{aligned}
$$

Now we consider the second term of (4.33). Note that

$$
n^{-1}\sum_{i=1}^{n}\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)w_i(t)\psi_i(\theta_0)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t) \xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta_0),
$$

where $\boldsymbol{Q}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta) = E\left\{\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)w_i(t)\psi_i(\theta)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\right\}$. It follows by (4.9) and (4.27) that

$$
\begin{aligned}
&-n^{-\frac{1}{2}}\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\left\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})\right\}^{-1}\boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta})\\
=\;& -n^{-1}\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\left\{n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})\right\}^{-1}n^{-\frac{1}{2}}\boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta})\\
=\;& \left\{\boldsymbol{Q}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta_0)+o_p(1)\right\}\left\{\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0)\right\}^{-1}\left\{n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{W}_{\gamma,i}(\tau,\theta_0)\right\}+o_p(1)\\
=\;& \boldsymbol{Q}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta_0)\left\{\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0)\right\}^{-1}\left\{n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{W}_{\gamma,i}(\tau,\theta_0)\right\}+o_p(1). \qquad (4.38)
\end{aligned}
$$

It follows by (4.33), (4.37) and (4.38) that

$$
\begin{aligned}
&n^{-\frac{1}{2}}\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\left\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t)-\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\left\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta})\right\}^{-1}\boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta})\right\}\\
=\;& n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{W}_{\boldsymbol{\eta},i}(t,\theta_0)+o_p(1), \qquad (4.39)
\end{aligned}
$$

where

$$\boldsymbol{W}_{\boldsymbol{\eta},i}(t,\theta) = \left\{\boldsymbol{\zeta}_{\boldsymbol{\eta},i}(t,\theta) + \boldsymbol{\kappa}_{\boldsymbol{\eta},i}(t,\theta) - \boldsymbol{Q}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta)\left\{\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta)\right\}^{-1}\boldsymbol{W}_{\gamma,i}(\tau,\theta)\right\}. \quad (4.40)$$

From (2.12), we have

$$\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\hat{\theta}) = \sum_{i=1}^{n} \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t)\psi_i(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\eta},i}(t).$$

It is shown in the Appendix A that

$$\begin{aligned} n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\hat{\theta}) &= n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0) + o_p(1) \\ &\xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{\eta}}(t,\theta_0), \end{aligned} \quad (4.41)$$

where $\boldsymbol{Q}_{\boldsymbol{\eta}}(t,\theta) = E\{\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t)\psi_i(\theta)\boldsymbol{D}_{\boldsymbol{\eta},i}(t)\}$.

By plugging (4.39) and (4.41) into (4.31), $\sqrt{n}(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t))$ is asymptotically equivalent to the following identically independent distributed decomposition

$$\sqrt{n}(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t))) = \left\{\boldsymbol{Q}_{\boldsymbol{\eta}}(t,\theta_0)\right\}^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} \boldsymbol{W}_{\boldsymbol{\eta},i}(t,\theta_0) + o_p(1). \quad (4.42)$$

By the functional central limit theorem for empirical process, $n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{W}_{\boldsymbol{\eta},i}(t,\theta_0)$ converges in distribution to a normal random vector with zero-mean and covariance matrix $E\left\{\boldsymbol{W}_{\boldsymbol{\eta},i}(t,\theta_0)\right\}^{\otimes 2}$. By the slutsky's theorem and an application of Theorem 19.5 of van der Vaart(1998), $\sqrt{n}(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t))$ converges weakly to a mean zero Gaussian process on $t \in [0,\tau]$ with the covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}} = \boldsymbol{Q}_{\boldsymbol{\eta}}^{-1}(t,\theta_0)\, E\left\{\boldsymbol{W}_{\boldsymbol{\eta},i}(t,\theta_0)\right\}^{\otimes 2} \boldsymbol{Q}_{\boldsymbol{\eta}}^{-1}(t,\theta_0). \quad (4.43)$$

Let

$$
\begin{aligned}
\widehat{\boldsymbol{\zeta}}_{\boldsymbol{\eta},i}(t,\theta) &= \widehat{\boldsymbol{D}}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)w_i(t)\boldsymbol{\psi}_i(\theta)\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - \widehat{\boldsymbol{F}}_{1i}(t)\right\},\\
\widehat{\boldsymbol{q}}_{\boldsymbol{\eta}}(s,t,\theta) &= n^{-1}\sum_{j=1}^{n}\widehat{\boldsymbol{D}}^{\mathsf{T}}_{\boldsymbol{\eta},j}(t)w_j(t)\boldsymbol{\psi}_j(\theta)\frac{\Delta_j N_j(t)}{\widehat{G}(T_j)}\mathcal{I}(s \le \widetilde{T}_j \le t),\\
\hat{y}(t) &= n^{-1}\sum_{i=1}^{n}\mathcal{I}(\widetilde{T}_i \ge t),\\
\widehat{M}_j^{\,c}(s) &= \mathcal{I}(\widetilde{T}_j \le s, \Delta_j = 0) - \int_0^s \mathcal{I}(\widetilde{T}_j \ge u)\ d(-\log \widehat{G}(u)),\\
\widehat{\boldsymbol{\kappa}}_{\boldsymbol{\eta},i}(t,\theta) &= \int_0^\tau \frac{\widehat{\boldsymbol{q}}_{\boldsymbol{\eta}}(s,t,\theta)}{\hat{y}(s)}d\widehat{M}_i^c(s),\\
\widehat{\boldsymbol{Q}}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta) &= n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{D}}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)w_i(t)\boldsymbol{\psi}_i(\theta)\widehat{\boldsymbol{D}}_{\boldsymbol{\gamma},i}(t). \qquad (4.44)
\end{aligned}
$$

The asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t))$ can be consistently estimated by

$$
\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}} \;=\; \widehat{\boldsymbol{Q}}_{\boldsymbol{\eta}}^{-1}(t,\hat{\theta})\, n^{-1}\sum_{i=1}^{n}\left\{\widehat{\boldsymbol{W}}_{\boldsymbol{\eta},i}(\tau,\hat{\theta})\right\}^{\otimes 2}\,\widehat{\boldsymbol{Q}}_{\boldsymbol{\eta}}^{-1}(t,\hat{\theta}),
$$

where

$$
\begin{aligned}
\widehat{\boldsymbol{W}}_{\boldsymbol{\eta},i}(t,\theta) &= \left\{\widehat{\boldsymbol{\zeta}}_{\boldsymbol{\eta},i}(t,\theta) + \widehat{\boldsymbol{\kappa}}_{\boldsymbol{\eta},i}(t,\theta) - \widehat{\boldsymbol{Q}}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta)\left\{\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}(\theta)\right\}^{-1}\widehat{\boldsymbol{W}}_{\gamma,i}(\tau,\theta)\right\},\\
\widehat{\boldsymbol{Q}}_{\boldsymbol{\eta}}(t,\theta) &= n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{D}}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)w_i(t)\boldsymbol{\psi}_i(\theta)\widehat{\boldsymbol{D}}_{\boldsymbol{\eta},i}(t),
\end{aligned}
$$

and where $\left\{\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}(\theta)\right\}^{-1}$ and $\widehat{\boldsymbol{W}}_{\gamma,i}(\tau,\theta)$ are defined in (4.30). □

4.2 Proofs of the Theorems in Chapter 3

Proof of Theorem 3.1

We have the following estimating equations from (3.13) and (3.14)

$$\widehat{\widehat{\boldsymbol{U}}}_{\boldsymbol{\eta}}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \hat{\theta}) = \boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\left\{\boldsymbol{R}(t) - \boldsymbol{F}_1(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma})\right\} + \hat{a}_{\eta}(t, \eta(t), \gamma, \hat{\theta}), \tag{4.45}$$

$$\widehat{\widehat{\boldsymbol{U}}}_{\boldsymbol{\gamma}}(\tau, \boldsymbol{\eta}(\cdot), \boldsymbol{\gamma}, \hat{\theta}) = \int_0^{\tau} \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\hat{\theta})\left\{\boldsymbol{R}(t) - \boldsymbol{F}_1(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma})\right\} dt + \hat{a}_{\gamma}(\tau, \eta(\cdot), \gamma, \hat{\theta}). \tag{4.46}$$

For model (3.5), $\boldsymbol{D}_{\boldsymbol{\eta}}(t)$ is the $n \times (p+1)$ matrix of $X = (X_1, \cdots, X_n)^{\mathsf{T}}$ with the ith row vector $\boldsymbol{D}_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}) = X_i^{\mathsf{T}} = (1, X_{i1}, \cdots, X_{ip})$, $\boldsymbol{D}_{\boldsymbol{\gamma}}(t)$ is the the $n \times q$ matrix of $\partial g(\gamma, Z, t)/\partial\gamma$ with the ith row vector $\boldsymbol{D}_{\gamma,i}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}) = \partial g(\gamma, Z_i, t)/\partial\gamma$, and where $\hat{a}_{\eta}(t, \eta(t), \gamma, \hat{\theta})$ and $\hat{a}_{\gamma}(\tau, \eta(\cdot), \gamma, \hat{\theta})$ are defined in (3.11) and (3.12) with $\widehat{V}_x = (\widehat{E}\{X_1|\mathcal{V}_1\}, \cdots, \widehat{E}\{X_n|\mathcal{V}_n\})^{\mathsf{T}}$ and $\widehat{V}_{xx}(\hat{\theta}) = \sum_{i=1}^n (1 - \psi_i(\hat{\theta}))w_i(t)\widehat{E}\left[X_i X_i^{\mathsf{T}}|\mathcal{V}_i\right]$.

Let

$$\widetilde{\boldsymbol{U}}_{\boldsymbol{\eta}}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \theta) = \boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta)\left\{\boldsymbol{R}(t) - \boldsymbol{F}_1(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma})\right\} + a_{\eta}(t, \eta(t), \gamma, \theta), \tag{4.47}$$

$$\widetilde{\boldsymbol{U}}_{\boldsymbol{\gamma}}(\tau, \boldsymbol{\eta}(\cdot), \boldsymbol{\gamma}, \theta) = \int_0^{\tau} \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta)\left\{\boldsymbol{R}(t) - \boldsymbol{F}_1(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma})\right\} dt + a_{\gamma}(\tau, \eta(\cdot), \gamma, \theta), \tag{4.48}$$

where

$$a_{\eta}(t, \eta(t), \gamma, \theta) = V_x^T W(t)(I - \boldsymbol{\Psi}(\theta))\left\{\boldsymbol{R}(t) - g(\gamma, Z, t)\right\} - V_{xx}(\theta)\eta(t),$$

$$a_{\gamma}(\tau, \eta(\cdot), \gamma, \theta) = \int_0^{\tau} \{\frac{\partial g(\gamma, Z, t)}{\partial \gamma}\}^{\mathsf{T}}\boldsymbol{W}(t)(I - \boldsymbol{\Psi}(\theta))\left\{\boldsymbol{R}(t) - V_x\eta(t) - g(\gamma, Z, t)\right\} dt,$$

where $V_x = (E\{X_1|\mathcal{V}_1\}, \cdots, E\{X_n|\mathcal{V}_n\})^{\mathsf{T}}$ and $V_{xx}(\theta) = \sum_{i=1}^n (1 - \psi_i(\theta))w_i(t)$

$\times E\left[X_i X_i^{\mathsf{T}} | \mathcal{V}_i\right].$

By adapting the theory of Robins, Rotnizky and Zhao (1994), if either $P(\xi_i = 1|\mathcal{V}_i)$ or $E\{X_i^{(2)}|\mathcal{V}_i\}$ and $E\{X_i^{(2)}(X_i^{(2)})^{\mathsf{T}}|\mathcal{V}_i\}$ is correctly specified, we have $\hat{\theta} \xrightarrow{p} \theta_0$, $\hat{a}_\eta(t, \eta(t), \gamma, \hat{\theta}) \xrightarrow{p} a_\eta(t, \eta(t), \gamma, \theta_0)$ and $\hat{a}_\gamma(\tau, \eta(\cdot), \gamma, \hat{\theta}) \xrightarrow{p} a_\gamma(\tau, \eta(\cdot), \gamma, \theta_0)$. It follows that the estimating equation (4.45) and (4.46) are asymptotically equivalent to (4.47) and (4.48). That is,

$$\begin{aligned} \widehat{\widehat{\widetilde{\boldsymbol{U}}}}_{\boldsymbol{\eta}}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \hat{\theta}) &= \widetilde{\boldsymbol{U}}_{\boldsymbol{\eta}}(t, \boldsymbol{\eta}(t), \boldsymbol{\gamma}, \theta_0) + o_p(n^{\frac{1}{2}}), \\ \widehat{\widehat{\widetilde{\boldsymbol{U}}}}_{\boldsymbol{\gamma}}(\tau, \boldsymbol{\eta}(\cdot), \boldsymbol{\gamma}, \hat{\theta}) &= \widetilde{\boldsymbol{U}}_{\boldsymbol{\gamma}}(\tau, \boldsymbol{\eta}(\cdot), \boldsymbol{\gamma}, \theta_0) + o_p(n^{\frac{1}{2}}). \end{aligned}$$

By using the Taylor expansion in (2.7) and replacing it into the estimation equations (4.47) and (4.48), we have

$$\begin{aligned} \widetilde{\boldsymbol{U}}_{\boldsymbol{\eta}}(t, \widehat{\boldsymbol{\eta}}(t), \widehat{\boldsymbol{\gamma}}, \theta_0) &= \boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\left[\boldsymbol{R}(t) - \boldsymbol{F}_1(t) - \boldsymbol{D}_{\boldsymbol{\eta}}(t)\left\{\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t)\right\}\right. \\ &\quad \left. - \boldsymbol{D}_{\boldsymbol{\gamma}}(t)\left\{\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\}\right] + \tilde{a}_\eta(t, \theta_0) + o_p(n^{\frac{1}{2}}) = 0, \qquad (4.49) \\ \widetilde{\boldsymbol{U}}_{\boldsymbol{\gamma}}(\tau, \widehat{\boldsymbol{\eta}}(\cdot), \widehat{\boldsymbol{\gamma}}, \hat{\theta}) &= \int_0^\tau \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\big[\boldsymbol{R}(t) - \boldsymbol{F}_1(t) - \boldsymbol{D}_{\boldsymbol{\eta}}(t)\{\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t)\} \\ &\quad \left. - \boldsymbol{D}_{\boldsymbol{\gamma}}(t)\left\{\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\right\}\right] dt + \tilde{a}_\gamma(\theta_0) + o_p(n^{\frac{1}{2}}) = 0. \qquad (4.50) \end{aligned}$$

where $\boldsymbol{D}_{\boldsymbol{\eta}}(t) = \boldsymbol{D}_{\boldsymbol{\eta}}(t, \boldsymbol{\eta}_0(t), \boldsymbol{\gamma}_0)$, $\boldsymbol{D}_{\boldsymbol{\gamma}}(t) = \boldsymbol{D}_{\boldsymbol{\gamma}}(t, \boldsymbol{\eta}_0(t), \boldsymbol{\gamma}_0)$, $\boldsymbol{F}_1(t) = \boldsymbol{F}_1(t, \boldsymbol{\eta}_0(t), \boldsymbol{\gamma}_0)$, $\tilde{a}_\eta(t, \theta_0) = a_\eta(t, \widehat{\boldsymbol{\eta}}(t), \widehat{\boldsymbol{\gamma}}, \theta_0)$ and $\tilde{a}_\gamma(\theta_0) = a_\gamma(\tau, \widehat{\boldsymbol{\eta}}(\cdot), \widehat{\boldsymbol{\gamma}}, \theta_0)$.

From (4.49), it can be solved for $\{\widehat{\boldsymbol{\eta}}(t)-\boldsymbol{\eta}_0(t)\}$. That is,

$$\boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\boldsymbol{D}_{\boldsymbol{\eta}}(t)\left\{\widehat{\boldsymbol{\eta}}(t)-\boldsymbol{\eta}_0(t)\right\} = \boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\left[\boldsymbol{R}(t)-\boldsymbol{F}_1(t)\right.$$

$$\left.-\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\left\{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\}\right] + \tilde{a}_\eta(t,\theta_0),$$

$$\widehat{\boldsymbol{\eta}}(t)-\boldsymbol{\eta}_0(t) = \left[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\right]^{-1}\boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\left[\boldsymbol{R}(t)-\boldsymbol{F}_1(t)\right.$$

$$\left.-\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\left\{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\}\right] + \left[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\right]^{-1}\tilde{a}_\eta(t,\theta_0) + o_p(n^{\frac{1}{2}}), \quad (4.51)$$

where $\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta) = \boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta)\boldsymbol{D}_{\boldsymbol{\eta}}(t)$.

By solving (4.50) for $\{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\}$,

$$\int_0^\tau \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\left\{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\}dt = \int_0^\tau \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\left[\boldsymbol{R}(t)-\boldsymbol{F}_1(t)\right]dt$$

$$-\int_0^\tau \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\boldsymbol{D}_{\boldsymbol{\eta}}(t)\left\{\widehat{\boldsymbol{\eta}}(t)-\boldsymbol{\eta}_0(t)\right\}dt + \tilde{a}_\gamma(\theta_0) + o_p(n^{-\frac{1}{2}}). \quad (4.52)$$

By substituting (4.51) into (4.52),

$$\begin{aligned}
&\int_0^\tau \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\left\{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\}dt \\
= &\int_0^\tau \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)[\boldsymbol{R}(t)-\boldsymbol{F}_1(t)]\,dt \\
&-\int_0^\tau \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\boldsymbol{D}_{\boldsymbol{\eta}}(t)\left[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\right]^{-1}\boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\left[\boldsymbol{R}(t)-\boldsymbol{F}_1(t)\right] \\
&+\int_0^\tau \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\boldsymbol{D}_{\boldsymbol{\eta}}(t)\left[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\right]^{-1}\boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\left\{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right\} \\
&-\int_0^\tau \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\boldsymbol{D}_{\boldsymbol{\eta}}(t)\left[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\right]^{-1}\tilde{a}_\eta(t,\theta_0)dt + \tilde{a}_\gamma(\theta_0) + o_p(n^{\frac{1}{2}}). \quad (4.53)
\end{aligned}$$

By combining like terms of $\{\widehat{\gamma}-\boldsymbol{\gamma}_0\}$ and $\boldsymbol{R}(t)-\boldsymbol{F}_1(t)$ in (4.53), we have

$$\begin{aligned}
&\int_0^\tau \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\,[I\\
&\qquad\qquad -\boldsymbol{D}_{\boldsymbol{\eta}}(t)\big[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\big]^{-1}\boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\Big]\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\,\{\widehat{\gamma}-\boldsymbol{\gamma}_0\}\,dt\\
= &\int_0^\tau \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\,[I\\
&\qquad\qquad -\boldsymbol{D}_{\boldsymbol{\eta}}(t)\,\big[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\big]^{-1}\boldsymbol{D}_{\boldsymbol{\eta}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\Big]\,[\boldsymbol{R}(t)-\boldsymbol{F}_1(t)]\\
&-\int_0^\tau \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\boldsymbol{D}_{\boldsymbol{\eta}}(t)\,\big[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\big]^{-1}\tilde{a}_\eta(t,\theta_0)dt+\tilde{a}_\gamma(\theta_0)\\
&+o_p(n^{\frac{1}{2}}).
\end{aligned} \tag{4.54}$$

Let

$$\begin{aligned}
\boldsymbol{A}_\eta(\theta) &= \int_0^\tau \boldsymbol{K}(t,\theta)\tilde{a}_\eta(t,\theta)\,dt\\
\boldsymbol{K}(t,\theta) &= \boldsymbol{D}_{\boldsymbol{\gamma}}^{\mathsf{T}}(t)\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta)\boldsymbol{D}_{\boldsymbol{\eta}}(t)\,\big[\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta)\big]^{-1}.
\end{aligned}$$

Using (2.12), (4.54) can be reduced to

$$\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0)\,\{\widehat{\gamma}-\boldsymbol{\gamma}_0\} = \boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0)-\boldsymbol{A}_\eta(\theta_0)+\tilde{a}_\gamma(\theta_0)+o_p(n^{\frac{1}{2}}).$$

Thus, we have

$$\widehat{\gamma}-\boldsymbol{\gamma}_0 = \big[\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0)\big]^{-1}\big\{\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0)-\boldsymbol{A}_\eta(\theta_0)+\tilde{a}_\gamma(\theta_0)\big\}+o_p(n^{-\frac{1}{2}}). \tag{4.55}$$

Again by substituting (4.55) into (4.51), we have

$$\begin{aligned}
\widehat{\boldsymbol{\eta}}(t)-\boldsymbol{\eta}_0(t) = &\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\}^{-1}\{\boldsymbol{D}_{\boldsymbol{\eta}}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\,\big\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t)-\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0)\}^{-1}\\
&\big\{\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0)+\tilde{a}_\gamma(\theta_0)-\boldsymbol{A}_\eta(\theta_0)\big\}\big\}+\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\}^{-1}\tilde{a}_\eta(t,\theta_0)+o_p(n^{-\frac{1}{2}}).
\end{aligned}$$

□

Proof of Theorem 3.2

From (3.24), we have

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0) \;=\; \left\{\frac{1}{n}\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0)\right\}^{-1} n^{-\frac{1}{2}}\left\{\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0)+\tilde{a}_{\gamma}(\theta_0)-\boldsymbol{A}_{\eta}(\theta_0)\right\}+o_p(1). \quad (4.56)$$

Consider $n^{-\frac{1}{2}}\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0)$ in (4.56). Using (2.12), with similar arguments to (4.11), we have

$$n^{-\frac{1}{2}}\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0) = n^{-\frac{1}{2}}\int_0^{\tau}\left\{\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma}}(t)-\boldsymbol{K}(t,\theta_0)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta}}(t)\right\}\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\left\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t)\right\}\,dt.$$

With similar argument to (4.12), it can be decomposed by

$$\begin{aligned} n^{-\frac{1}{2}}\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0) \;=\; & n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^{\tau}\left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t)-\boldsymbol{K}(t,\theta_0)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right]w_i(t) \\ & \times\boldsymbol{\psi}_i(\theta_0)\left\{\frac{\Delta_i N_i(t)}{G(T_i)}-\boldsymbol{F}_{1i}(t)\right\}\,dt \\ & +n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^{\tau}\left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t)-\boldsymbol{K}(t,\theta_0)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right]w_i(t) \\ & \times\boldsymbol{\psi}_i(\theta_0)\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)}-\frac{\Delta_i N_i(t)}{G(T_i)}\right\}\,dt. \end{aligned} \quad (4.57)$$

By using consistency of $\boldsymbol{K}(t,\theta_0)$ shown in (A.3) in Appendix 4.2, we have

$$n^{-\frac{1}{2}}\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0) = n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^{\tau}\boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t,\theta_0)\,dt + n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^{\tau}\boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t,\theta_0)\,dt + o_p(1), \quad (4.58)$$

where $\boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t,\theta)$ and $\boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t,\theta_0)$ are defined in (4.15) and (4.21), respectively.

Consider $\tilde{a}_{\gamma}(\theta_0)$ is defined in Theorem 3.1. The term $n^{-\frac{1}{2}}\tilde{a}_{\gamma}(\theta_0)$ can be decomposed into four parts

$$
\begin{aligned}
n^{-\frac{1}{2}}\tilde{a}_{\gamma}(\theta_0) = {} & n^{-\frac{1}{2}}\int_0^{\tau}\sum_{i=1}^{n}\{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0))\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - \frac{\Delta_i N_i(t)}{G(T_i)}\right\} dt \\
& + n^{-\frac{1}{2}}\int_0^{\tau}\sum_{i=1}^{n}\{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0))\left\{\frac{\Delta_i N_i(t)}{G(T_i)}\right. \\
& \qquad\qquad\qquad\qquad \left. -V_{x,i}\eta_0(t) - g(\gamma_0, Z_i, t)\right\} dt \\
& - n^{-\frac{1}{2}}\int_0^{\tau}\sum_{i=1}^{n}\{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0))V_{x,i}\left\{\widehat{\eta}(t) - \eta_0(t)\right\} dt \\
& - n^{-\frac{1}{2}}\int_0^{\tau}\sum_{i=1}^{n}\{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0)) \\
& \qquad\qquad\qquad\qquad \times\left\{g(\widehat{\gamma}, Z_i, t) - g(\gamma_0, Z_i, t)\right\} dt. \qquad (4.59)
\end{aligned}
$$

It is shown in the Appendix A that the third and fourth terms of the above equation are equal to $o_p(1)$, respectively. That is,

$$
\begin{aligned}
& - n^{-\frac{1}{2}}\int_0^{\tau}\sum_{i=1}^{n}\{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0)) \\
& \qquad\qquad\qquad \times V_{x,i}\left\{\widehat{\eta}(t) - \eta_0(t)\right\} dt = o_p(1), \qquad (4.60) \\
& - n^{-\frac{1}{2}}\int_0^{\tau}\sum_{i=1}^{n}\{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0)) \\
& \qquad\qquad\qquad \times\left\{g(\widehat{\gamma}, Z_i, t) - g(\gamma_0, Z_i, t)\right\} dt = o_p(1). \qquad (4.61)
\end{aligned}
$$

Consider the first term of (4.59). By using (4.20) and the law of large numbers,

$$
n^{-1}\sum_{i=1}^{n}\{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0))\frac{\Delta_i N_i(t)}{G(T_i)} I(s \le \tilde{T}_i \le t) \xrightarrow{p} \boldsymbol{q}^*_{\boldsymbol{\gamma}}(s, t, \boldsymbol{\gamma}_0, \theta_0). \qquad (4.62)
$$

It follows that

$$
\begin{aligned}
& n^{-\frac{1}{2}} \int_0^\tau \sum_{i=1}^n \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^\mathsf{T} w_i(t)(1-\psi_i(\theta_0)) \left\{ \frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - \frac{\Delta_i N_i(t)}{G(T_i)} \right\} dt \\
&= -n^{-\frac{1}{2}} \sum_{j=1}^n \int_0^\tau \int_0^\tau n^{-1} \sum_{i=1}^n \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^\mathsf{T} w_i(t)(1-\psi_i(\theta_0)) \\
&\qquad \times \frac{\Delta_i N_i(t)}{G(T_i)} \mathcal{I}(s \le \widetilde{T}_i \le t) \frac{d{M_j}^c(s)}{y(s)}\, dt + o_p(1). \\
&= -n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \left\{ \frac{\boldsymbol{q}^*_{\boldsymbol{\gamma}}(s,t,\boldsymbol{\gamma}_0,\theta_0)}{y(s)} + o_p(1) \right\} dM_i^c(s) + o_p(1) \\
&= -n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \boldsymbol{\kappa}^*_{\boldsymbol{\gamma},i}(t,\boldsymbol{\gamma}_0,\theta_0) dt + o_p(1), \qquad (4.63)
\end{aligned}
$$

where $\boldsymbol{\kappa}^*_{\boldsymbol{\gamma},i}(t,\boldsymbol{\gamma},\theta) = \int_0^\tau \frac{\boldsymbol{q}^*_{\boldsymbol{\gamma}}(s,t,\boldsymbol{\gamma},\theta)}{y(s)} dM_i^c(s)$.

Now consider the second term of (4.59).

$$
\begin{aligned}
& n^{-\frac{1}{2}} \int_0^\tau \sum_{i=1}^n \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^\mathsf{T} w_i(t)(1-\psi_i(\theta_0)) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} \right. \\
&\qquad \left. - V_{x,i}\eta_0(t) - g(\gamma_0, Z_i, t) \right\} dt \\
&= n^{-\frac{1}{2}} \int_0^\tau \sum_{i=1}^n \boldsymbol{\zeta}^*_{\boldsymbol{\gamma},i}(t, \boldsymbol{\eta}_0(t), \boldsymbol{\gamma}_0, \theta_0)\, dt, \qquad (4.64)
\end{aligned}
$$

where $\boldsymbol{\zeta}^*_{\boldsymbol{\gamma},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta) = \{\frac{\partial g(\gamma, Z_i, t)}{\partial \gamma}\}^\mathsf{T} w_i(t)(1-\psi_i(\theta)) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - V_{x,i}\eta(t) - g(\gamma, Z_i, t) \right\}$.

Therefore, it follows by (4.59), (4.60), (4.61), (4.63) and (4.64) that

$$
\begin{aligned}
n^{-\frac{1}{2}} \tilde{a}_\gamma(\theta_0) &= -n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \boldsymbol{\kappa}^*_{\boldsymbol{\gamma},i}(t,\boldsymbol{\gamma}_0,\theta_0) dt + n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \boldsymbol{\zeta}^*_{\boldsymbol{\gamma},i}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0,\theta_0) dt \\
&\quad + o_p(1). \qquad (4.65)
\end{aligned}
$$

Consider $-n^{-\frac{1}{2}} \boldsymbol{A}_\eta(\theta_0)$ in (4.56). From (3.26), we have

$$
-n^{-\frac{1}{2}} \boldsymbol{A}_\eta(\theta_0) = \int_0^\tau \boldsymbol{K}(t,\theta_0) \tilde{a}_\eta(t,\theta_0)\, dt. \qquad (4.66)
$$

To finish this, first we consider i.i.d expression for $\tilde{a}_\eta(t,\theta_0)$. It can be decomposed into four terms

$$\begin{aligned}
n^{-\frac{1}{2}}\tilde{a}_\eta(t,\theta_0) &= n^{-\frac{1}{2}}\sum_{i=1}^n V_{x,i}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0))\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - \frac{\Delta_i N_i(t)}{G(T_i)}\right\} \\
&+ n^{-\frac{1}{2}}\sum_{i=1}^n \Bigg[V_{x,i}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0))\left\{\frac{\Delta_i N_i(t)}{G(T_i)} - g(\gamma_0, Z_i, t)\right\} \\
&\qquad - V_{xx,i}(\theta_0)\eta_0(t)\Bigg] \\
&- n^{-\frac{1}{2}}\sum_{i=1}^n V_{x,i}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0))\left\{g(\widehat{\gamma}, Z_i, t) - g(\gamma_0, Z_i, t)\right\} \\
&- n^{-\frac{1}{2}}\sum_{i=1}^n V_{xx,i}(\theta_0)\left\{\widehat{\eta}(t) - \eta_0(t)\right\}. \qquad (4.67)
\end{aligned}$$

It is shown in the Appendix A that the third and fourth terms of (4.67) are shown to be equal to $o_p(1)$, respectively. That is,

$$- \quad n^{-\frac{1}{2}}\sum_{i=1}^n V_{x,i}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0))\left\{g(\widehat{\gamma}, Z_i, t) - g(\gamma_0, Z_i, t)\right\} = o_p(1), \qquad (4.68)$$

$$- \quad n^{-\frac{1}{2}}\sum_{i=1}^n V_{xx,i}(\theta_0)\left\{\widehat{\eta}(t) - \eta_0(t)\right\} = o_p(1). \qquad (4.69)$$

Consider the first term of (4.67). By using (4.20) and the law of large numbers,

$$n^{-1}\sum_{j=1}^n V_{x,j}^{\mathsf{T}} w_j(t)(1-\psi_j(\theta_0))\frac{\Delta_j N_j(t)}{G(T_j)}\mathcal{I}(s \le \widetilde{T}_j \le t) \xrightarrow{p} \boldsymbol{q}^*_{\boldsymbol{\eta}}(s,t,\theta_0). \qquad (4.70)$$

It follows that

$$\begin{aligned}
& n^{-\frac{1}{2}}\sum_{i=1}^n V_{x,i}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0))\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - \frac{\Delta_i N_i(t)}{G(T_i)}\right\} \\
&= -n^{-\frac{1}{2}}\sum_{i=1}^n \int_0^\tau n^{-1}\sum_{j=1}^n V_{x,j}^{\mathsf{T}} w_j(t)(1-\psi_j(\theta_0))\frac{\Delta_j N_j(t)}{G(T_j)}\mathcal{I}(s \le \widetilde{T}_j \le t)\frac{dM_i{}^c(s)}{y(s)} \\
&= -n^{-\frac{1}{2}}\sum_{i=1}^n \int_0^\tau \frac{\boldsymbol{q}^*_{\boldsymbol{\eta}}(s,t,\theta_0)}{y(s)} dM_i{}^c(s) + o_p(1) \\
&= -n^{-\frac{1}{2}}\sum_{i=1}^n \boldsymbol{\kappa}^*_{\boldsymbol{\eta},i}(t,\theta_0) + o_p(1), \qquad (4.71)
\end{aligned}$$

where $\boldsymbol{\kappa}^*_{\boldsymbol{\eta},i}(t,\theta) = \int_0^\tau \{\boldsymbol{q}^*_{\boldsymbol{\eta}}(s,t,\theta)/y(s)\} dM_i{}^c(s)$.

Consider the second term of (4.67). We have,

$$\begin{aligned} & n^{-\frac{1}{2}} \sum_{i=1}^n \left[V_{x,i}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0)) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - g(\gamma_0, Z_i, t) \right\} - V_{xx,i}(\theta_0)\eta_0(t) \right] \\ = \quad & n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{\zeta}^*_{\boldsymbol{\eta},i}(t, \boldsymbol{\eta}_0(t), \boldsymbol{\gamma}_0, \theta_0), \qquad (4.72) \end{aligned}$$

where $\boldsymbol{\zeta}^*_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta) = V_{x,i}^T w_i(t)(1-\psi_i(\theta)) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - g(\gamma, Z_i, t) \right\} - V_{xx,i}(\theta)\eta(t)$.

It follows by (4.67), (4.68), (4.69), (4.71) and (4.72) that

$$n^{-\frac{1}{2}} \tilde{a}_\eta(t,\theta_0) \quad = \quad -n^{-\frac{1}{2}} \sum_{i=1}^n \left\{ \boldsymbol{\kappa}^*_{\boldsymbol{\eta},i}(t,\theta_0) - \boldsymbol{\zeta}^*_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0,\theta_0) \right\} + o_p(1). \quad (4.73)$$

Note that $\boldsymbol{K}(t,\theta_0) \xrightarrow{p} \boldsymbol{k}(t,\theta_0)$. From (4.66) and (4.73), we have

$$\begin{aligned} -n^{-\frac{1}{2}} \boldsymbol{A}_\eta(\theta_0) &= n^{-\frac{1}{2}} \int_0^\tau \boldsymbol{k}(t,\theta_0)\tilde{a}_\eta(t,\theta_0)\, dt + n^{-\frac{1}{2}} \int_0^\tau \{\boldsymbol{K}(t,\theta_0) - \boldsymbol{k}(t,\theta_0)\} \tilde{a}_\eta(t,\theta_0)\, dt \\ &= n^{-\frac{1}{2}} \int_0^\tau \boldsymbol{k}(t,\theta_0)\tilde{a}_\eta(t,\theta_0)\, dt + o_p(1) \\ &= -n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\tau \boldsymbol{k}(t,\theta_0) \left\{ \boldsymbol{\kappa}^*_{\boldsymbol{\eta},i}(t,\theta_0) - \boldsymbol{\zeta}^*_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0,\theta_0) \right\}\, dt \\ &\quad + o_p(1). \qquad (4.74) \end{aligned}$$

It follows by (4.26), (4.56),(4.58), (4.65), (4.74) that

$$\begin{aligned} n^{\frac{1}{2}}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) &= \left\{ n^{-1} \boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0) \right\}^{-1} n^{-\frac{1}{2}} \left\{ \boldsymbol{B}_{\boldsymbol{\gamma}}(\hat{\theta}) + \tilde{a}_\gamma(\theta_0) - \boldsymbol{A}_\eta(\theta_0) \right\} \\ &= \left\{ \boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0) \right\}^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{W}^*_{\boldsymbol{\gamma},i}(\tau, \boldsymbol{\eta}_0(\cdot), \boldsymbol{\gamma}_0, \theta_0) + o_p(1), \quad (4.75) \end{aligned}$$

where $\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta)$ is defined in Theorem 2.1, and where

$$\begin{aligned}\boldsymbol{W}^*_{\boldsymbol{\gamma},i}(\tau,\boldsymbol{\eta}(\cdot),\boldsymbol{\gamma},\theta) &= \int_0^\tau \boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t,\theta)\,dt + \int_0^\tau \boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t,\theta)\,dt \\ &- \int_0^\tau \boldsymbol{\kappa}^*_{\boldsymbol{\gamma},i}(t,\boldsymbol{\gamma},\theta)dt + \int_0^\tau \boldsymbol{\zeta}^*_{\boldsymbol{\gamma},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta)dt \\ &- \int_0^\tau \boldsymbol{k}(t,\theta_0)\left\{\boldsymbol{\kappa}^*_{\boldsymbol{\eta},i}(t,\theta_0) - \boldsymbol{\zeta}^*_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0,\theta_0)\right\} dt. \quad (4.76)\end{aligned}$$

Since $\boldsymbol{\zeta}_{\boldsymbol{\gamma},i}(t,\theta_0)$ and $\boldsymbol{\kappa}_{\boldsymbol{\gamma},i}(t,\theta_0)$ has mean zero from Theorem 2.1, and since $\boldsymbol{\kappa}^*_{\boldsymbol{\gamma},i}(t,\boldsymbol{\gamma},\theta_0)$ has mean zero local square martingale and, by missing at random assumption, $\boldsymbol{\zeta}^*_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0,\theta_0)$ has mean zero, then, by the standard central limit theorem, $n^{-1}\ \sum_{i=1}^n \boldsymbol{W}^*_{\boldsymbol{\gamma},i}(\tau,\boldsymbol{\eta}_0(\cdot),\boldsymbol{\gamma}_0,\theta_0)$ has mean zero normal random vector with covariance matrix $E\{\boldsymbol{W}^*_{\boldsymbol{\gamma},i}(\tau,\boldsymbol{\eta}_0(\cdot),\boldsymbol{\gamma}_0,\theta_0)\}^{\otimes 2}$. By slutsky's theorem, we have

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{d} N(0,\Sigma^*_{\boldsymbol{\gamma}}), \quad (4.77)$$

where $\Sigma^*_{\boldsymbol{\gamma}} = {\boldsymbol{Q}_{\boldsymbol{\gamma}}}^{-1}(\theta_0)E\{\boldsymbol{W}^*_{\boldsymbol{\gamma},i}(\tau,\boldsymbol{\eta}_0(\cdot),\boldsymbol{\gamma}_0,\theta_0)\}^{\otimes 2}{\boldsymbol{Q}_{\boldsymbol{\gamma}}}^{-1}(\theta_0)$.

Let

$$\begin{aligned}\widehat{\boldsymbol{q}}^*_{\boldsymbol{\gamma}}(s,t,\boldsymbol{\gamma},\theta) &= n^{-1}\sum_{i=1}^n\{\frac{\partial g(\gamma,Z_i,t)}{\partial\gamma}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta))\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} I(s\le \tilde{T}_i \le t), \\ \hat{y}(t) &= n^{-1}\sum_{i=1}^n \mathcal{I}(\widetilde{T}_i \ge t), \\ \widehat{M}_i^c(t) &= \mathcal{I}(\widetilde{T}_j \le t, \Delta_j = 0) - \int_0^t \mathcal{I}(\widetilde{T}_j \ge s)\ d(-\log \widehat{G}(s)), \\ \widehat{\boldsymbol{\kappa}}^*_{\boldsymbol{\gamma},i}(t,\theta) &= \int_0^\tau \frac{\widehat{\boldsymbol{q}}^*_{\boldsymbol{\gamma}}(s,t,\boldsymbol{\gamma},\theta)}{\hat{y}(s)} d\widehat{M}_i^c(s), \\ \widehat{\boldsymbol{\zeta}}^*_{\boldsymbol{\gamma},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta)\,dt &= n^{-1}\sum_{i=1}^n \int_0^\tau \{\frac{\partial g(\gamma,Z_i,t)}{\partial\gamma}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta))\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)}\right. \\ &\qquad \left. -\widehat{V}_{x,i}\eta(t) - g(\gamma,Z_i,t)\right\}\ dt. \quad (4.78)\end{aligned}$$

The asymptotic covariance matrix of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ can be consistently estimated by

$$\left\{\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}^{-1}(\hat{\theta})\right\} \; n^{-1}\sum_{i=1}^{n}\left\{\widehat{\boldsymbol{W}}^{*}_{\boldsymbol{\gamma},i}(\tau,\widehat{\boldsymbol{\eta}}(\cdot),\widehat{\gamma},\hat{\theta})\right\}^{\otimes 2}\left\{\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}^{-1}(\hat{\theta})\right\},$$

where $\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}(\theta)$ is defined in (4.30), and where

$$\begin{aligned}\widehat{\boldsymbol{W}}^{*}_{\boldsymbol{\gamma},i}(\tau,\boldsymbol{\eta}(\cdot),\boldsymbol{\gamma},\theta) &= \int_0^\tau \widehat{\boldsymbol{\zeta}}_{\boldsymbol{\gamma},i}(t,\theta)dt + \int_0^\tau \widehat{\boldsymbol{\kappa}}_{\boldsymbol{\gamma},i}(t,\theta)dt \\ &- \int_0^\tau \widehat{\boldsymbol{\kappa}}^{*}_{\boldsymbol{\gamma},i}(t,\gamma,\theta)dt + \int_0^\tau \widehat{\boldsymbol{\zeta}}^{*}_{\boldsymbol{\gamma},i}(t,\boldsymbol{\eta}(t),\gamma,\theta)dt \\ &- \int_0^\tau \boldsymbol{K}(t,\theta)\left\{\boldsymbol{\kappa}^{*}_{\boldsymbol{\eta},i}(t,\theta) - \boldsymbol{\zeta}^{*}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\gamma,\theta)\right\}dt. \quad (4.79)\end{aligned}$$

□

Proof of Theorem 3.3

From (3.25),

$$\begin{aligned}& n^{\frac{1}{2}}(\widehat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}_0(t)) \\ =\; & \left\{n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\theta_0)\right\}^{-1} \Big[n^{-\frac{1}{2}}\{\boldsymbol{D}_{\boldsymbol{\eta}}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\left\{\boldsymbol{R}(t)-\boldsymbol{F}_1(t)\right\} \\ & -n^{-\frac{1}{2}}\{\boldsymbol{D}_{\boldsymbol{\eta}}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\left\{\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0)\}^{-1}\left\{\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0) + \tilde{a}_{\gamma}(\theta_0) - \boldsymbol{A}_{\eta}(\theta_0)\right\}\right\} \\ & +n^{-\frac{1}{2}}\tilde{a}_{\eta}(t,\theta_0)\Big] + o_p(1). \quad (4.80)\end{aligned}$$

Consider

$$-n^{-\frac{1}{2}}\{\boldsymbol{D}_{\boldsymbol{\eta}}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\left\{\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\{\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0)\}^{-1}\left\{\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0) + \tilde{a}_{\gamma}(\theta_0) - \boldsymbol{A}_{\eta}(\theta_0)\right\}\right\}. \quad (4.81)$$

Note that $n^{-1}\{\boldsymbol{D}_{\boldsymbol{\eta}}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\boldsymbol{D}_{\boldsymbol{\gamma}}(t) \xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta_0)$. It follows by (4.75) that

$$\begin{aligned} & -n^{-1}\{\boldsymbol{D}_{\boldsymbol{\eta}}(t)\}^{\mathsf{T}}\boldsymbol{W}(t)\boldsymbol{\Psi}(\theta_0)\left\{\boldsymbol{D}_{\boldsymbol{\gamma}}(t)\{n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0)\}^{-1}\right. \\ & \qquad \left.\times n^{-\frac{1}{2}}\left\{\boldsymbol{B}_{\boldsymbol{\gamma}}(\theta_0)+\tilde{a}_{\gamma}(\theta_0)-\boldsymbol{A}_{\eta}(\theta_0)\right\}\right\} \\ = & -\boldsymbol{Q}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta_0)\left\{\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0)\right\}^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n}\boldsymbol{W}^*_{\boldsymbol{\gamma},i}(\tau,\boldsymbol{\eta}_0(\cdot),\boldsymbol{\gamma}_0,\theta_0)+o_p(1). \end{aligned} \tag{4.82}$$

It follows by (4.37), (4.73), (4.82) that (4.80) is

$$\begin{aligned} n^{\frac{1}{2}}(\widehat{\boldsymbol{\eta}}(t)-\boldsymbol{\eta}_0(t)) &= \{n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t,\hat{\theta})\}^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{W}^*_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0,\theta_0) \\ &\quad +o_p(1), \end{aligned} \tag{4.83}$$

where

$$\begin{aligned} \boldsymbol{W}^*_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0,\theta_0) &= \boldsymbol{\zeta}_{\boldsymbol{\eta},i}(t,\theta_0)+\boldsymbol{\kappa}_{\boldsymbol{\eta},i}(t,\theta_0) \\ &\quad -\boldsymbol{Q}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta_0)\left\{\boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0)\right\}^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n}\boldsymbol{W}^*_{\boldsymbol{\gamma},i}(\tau,\boldsymbol{\eta}_0(\cdot),\boldsymbol{\gamma}_0,\theta_0) \\ &\quad -\boldsymbol{\kappa}^*_{\boldsymbol{\eta},i}(t,\theta_0)+\boldsymbol{\zeta}^*_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0,\theta_0). \end{aligned}$$

By using (4.41), we have the following i.i.d decomposition

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\eta}}(t)-\boldsymbol{\eta}_0(t)) = \{\boldsymbol{Q}_{\boldsymbol{\eta}}(t,\theta_0)\}^{-1}\ n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{W}^*_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0,\theta_0)+o_p(1).$$

Let

$$\begin{aligned}
\widehat{\boldsymbol{q}}^{*}_{\boldsymbol{\eta}}(s,t,\theta) &= n^{-1}\sum_{i=1}^{n}\widehat{V}_{x,i}^{T}w_i(t)(1-\psi_i(\theta))\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)}\mathcal{I}(s\le \widetilde{T}_i\le t),\\
\hat{y}(t) &= n^{-1}\sum_{i=1}^{n}\mathcal{I}(\widetilde{T}_i\ge t),\\
\widehat{M}_j^{c}(t) &= \mathcal{I}(\widetilde{T}_j\le t,\Delta_j=0)-\int_0^t \mathcal{I}(\widetilde{T}_j\ge s)\; d(-\log\widehat{G}(s)),\\
\widehat{\boldsymbol{\kappa}}^{*}_{\boldsymbol{\eta},i}(t,\theta) &= \int_0^{\tau}\frac{\widehat{\boldsymbol{q}}^{*}_{\boldsymbol{\eta}}(s,t,\theta)}{\hat{y}(s)}d\widehat{M}_j^{c}(s),\\
\widehat{\boldsymbol{\zeta}}^{*}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta) &= n^{-1}\sum_{i=1}^{n}\left\{\widehat{V}_{x,i}^{T}w_i(t)(1-\psi_i(\theta))\left\{\frac{\Delta_i N_i(t)}{\widehat{G}(T_i)}-g_i(\gamma,Z_i,t)\right\}\right.\\
&\qquad\left.-\widehat{V}_{xx,i}(\theta)\eta(t)\right\}. \qquad (4.84)
\end{aligned}$$

By using lemma 1 of Sun and Wu (2005), $n^{\frac{1}{2}}(\widehat{\boldsymbol{\eta}}(t)-\boldsymbol{\eta}_0(t))$ converges weakly to a mean-zero Gaussian process on $t\in[0,\tau]$ with the covariance matrix

$$\boldsymbol{\Sigma}^{*}_{\boldsymbol{\eta}}=\boldsymbol{Q}_{\boldsymbol{\eta}}(t,\theta_0)^{-1}E\{\boldsymbol{W}^{*}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}_0(t),\boldsymbol{\gamma}_0,\theta_0)\}^{\otimes 2}\boldsymbol{Q}_{\boldsymbol{\eta}}(t,\theta_0)^{-1},$$

which can be consistently estimated by

$$\widehat{\boldsymbol{\Sigma}}^{*}_{\boldsymbol{\eta}} = \widehat{\boldsymbol{Q}}_{\boldsymbol{\eta}}^{-1}(t,\hat{\theta})\,n^{-1}\sum_{i=1}^{n}\left\{\widehat{\boldsymbol{W}}^{*}_{\boldsymbol{\eta},i}(t,\widehat{\boldsymbol{\eta}}(t),\widehat{\gamma},\hat{\theta})\right\}^{\otimes 2}\,\widehat{\boldsymbol{Q}}_{\boldsymbol{\eta}}^{-1}(t,\hat{\theta}),$$

where

$$\begin{aligned}
\widehat{\boldsymbol{W}}^{*}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta) &= \widehat{\boldsymbol{\zeta}}_{\boldsymbol{\eta},i}(t,\theta)+\widehat{\boldsymbol{\kappa}}_{\boldsymbol{\eta},i}(t,\theta)\\
&\quad -\widehat{\boldsymbol{Q}}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta)\left\{\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}(\theta)\right\}^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}\widehat{\boldsymbol{W}}^{*}_{\boldsymbol{\gamma},i}(\tau,\boldsymbol{\eta}(\cdot),\boldsymbol{\gamma},\theta)\\
&\quad -\widehat{\boldsymbol{\kappa}}^{*}_{\boldsymbol{\eta},i}(t,\theta)+\widehat{\boldsymbol{\zeta}}^{*}_{\boldsymbol{\eta},i}(t,\boldsymbol{\eta}(t),\boldsymbol{\gamma},\theta),
\end{aligned}$$

and where $\widehat{\boldsymbol{Q}}_{\boldsymbol{\gamma}}(\theta)$ is defined in (4.30), $\widehat{\boldsymbol{Q}}_{\boldsymbol{\eta},\boldsymbol{\gamma}}(t,\theta)$, $\widehat{\boldsymbol{Q}}_{\boldsymbol{\eta}}(t,\theta)$,$\widehat{\boldsymbol{\zeta}}_{\boldsymbol{\eta},i}(t,\theta)$, and $\widehat{\boldsymbol{\kappa}}_{\boldsymbol{\eta},i}(t,\theta)$ are defined in (4.44) and $\widehat{\boldsymbol{W}}^{*}_{\boldsymbol{\gamma},i}(\tau,\boldsymbol{\eta}(\cdot),\boldsymbol{\gamma},\hat{\theta})$ is defined in (4.79). □

REFERENCES

Aly, E., Kochar, S., and McKeague, I. (1994). Some tests for com paring cumulative incidence functions and cause-specific hazard rates. *Journal of the American Statistical Association*, 89:994–999.

Aras, G. and Deshpandé, J. (1992). Statistical analysis of dependent competing risks. *Statistics and Decisions*, 10:323–336.

Bagai, I.and Deshpandé, J. and Kochar, S. (1989 a). A distribution-free test for the equality of failure rates due to two competing risks. *Comm. Statist. A-Theory Methods*, 18:107–120.

Bagai, I.and Deshpandé, J. and Kochar, S. (1989 b). Distribution-free tests for stochastic ordering among two independent risks. *Biometrika*, 76:775–778.

Barlow, W. (1994). Robust variance estimation for the case-cohort design. *Biometrics*, 50:1064–1072.

Benichou, J. and Gail, M. H. (1990). Estimates of absolute cause-specific risk in cohort studies. *Biometrics*, 46:813–826.

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., and Ritov, Y. (1993). *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press Baltimore.

Borgan, Ø., Langholz, B., Samuelsen, S., Goldstein, L., and Pogoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Analysis*, 6:39–58.

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

Chen, K. and Lo, S. (1999). Case-cohort and case-control analysis with coxs model. *Biometrika*, 86:755–764.

Cheng, S., Fine, J., and Wei, L. (1988). Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*, 54:219–228.

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

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

Fine, J. and Gray, R. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94:496–509.

Fleming, T. and Harrington, D. (2013). *Counting Processes and Survival Analysis*. John Wiley & Sons, Inc., New York.

Flinn, C. and Heckman, J. (1983). Are unemployment and out of the labor force behaviorally distinct labor force states? *Journal of Labor Economics*, I:28–42.

Gaynor, J., Feuer, E., Tan, C., Wu, D., Little, C., Straus, D., Clarkson, B., and Brennan, M. (1993). On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. *Journal of the Aemrican Statistical Association*, 88:400–409.

Gilbert, P., McKeague, I., and Sun, Y. (2004). Tests for comparing mark-specfic hazards and cumulative incidence functions. *Lifetime Data Analysis*, 10:5–28.

Gray, R. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, 16:1141–1154.

Haynes, B., Gilbert, P., McElrath, M., Zolla-Pazner, S., Tomaras, G., Alam, S., Evans, D., Montefiori, D., Karnasuta, C., Sutthent, R., Liao, H., DeVico, A., Lewis, G., Williams, C., Pinter, A., Fong, Y., Janes, H., deCamp, A., Huang, Y., Rao., M., Billings, E., Karasavvas, N., Robb, M., Ngauy, V., de Souza, M., Paris, R., Ferrari, G., Bailer, R., Soderberg, K., Andrews, C., Berman, P., Frahm, N., De Rosa, S., Alpert, M., Yates, N., Shen, X., Koup, R., Pitisuttithum, P., Kaewkungwal, J., Nitayaphan, S., Rerks-Ngarm, S., Michael, N., and Kim, J. (2012). Immune-correlates analysis of an hiv-1 vaccine efficacy trial. *New England Journal of Medicine*, 366:1275–1286.

Hinds, P. A. (1996). Reliability analysis and techniques for equipment management. *MSc Dissertation*.

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

Kalbfleisch, J. and Lawless, J. (1988). Likelihood analysis of multi-state models for disease incidence and mortality. *Statistics in Medicine*, 7:149–160.

Kang, S., Cai, J., and Chambless, L. (2013). Marginal additive hazards model for case-cohort studies with multiple disease outcomes:an application to the atherosclerosis risk in communities (aric) study. *Biostatistics*, 14:28–41.

Kong, L. and Cai, J. (2009). Case-cohort analysis with accelerated failure time model. *Biometrics*, 65:135–142.

Korn, E. L. and Dorey, F. J. (1992). Applications of crude incidence curves. *Statistics in Medicine*, 11:813–829.

Koul, H., Susarla, V., and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.*, 9(6):1276–1288.

Kulich, M. and Lin, D. (2000). Additive hazard regressions for case-cohort studies. *Biometrika*, 87:73–87.

Kulich, M. and Lin, D. (2004). Improving the efficiency of relative-risk estimation in case-cohort studies. *Journal of the American Statistical Association*, 99(467):832–844.

Lawless, J. (2003). *Statistical Models and Methods for Lifetime Data.* John Wiley & Sons, New York, 2003.

Lin, D. and Ying, Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association*, 88:1341–1349.

Lin, D. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika*, 81:61–71.

McKeague, I., Gilbert, P., and Kanki, P. (2001). Omnibus tests for comparison of competing risks with adjustment for covariate effects. *Biometrics*, 57:818–828.

McKeague, I. and Sasieni, P. (1994). A partly parametric additive risk model. *Biometrika*, 81:501–514.

Neuhaus, G. (1991). Some linear and nonlinear rank tests for competing risks models. *Comm. Statist. A-Theory Methods*, 20:667–701.

Nickle, D. C., L., H., Jensen, M. A., Gilbert, P. B., Mullins, J. I., and Kosakovsky Pond, S. L. (2007). Hiv-specific probabilistic models of protein evolution. *PLoS ONE*, 2(6):e503.

Pintilie, M. (2007). Analysing and interpreting competing risk data. *Statistics in Medicine*, 26:1360–1367.

Prentice, R. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, 73:1–11.

Prentice, R., Kalbfleisch, J., Peterson, A., Flournoy, N., Farewell, V., and Breslow, N. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 34:541–554.

Putter, H., Fiocco, M., and Geskus, R. (2007). Tutorial in biostatistics: Competing risks and multi-state models. *STATISTICS IN MEDICINE*, 26:2389–2430.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E., Benenson, M., Gurunathan, S., Tartaglia, J., McNeil, J., Francis, D., Stablein, D., Birx, D., Chunsuttiwat, S., Khamboonruang, C., Thongcharoen, P., Robb, M., Michael, N., Kunasol, P., and Kim, J. (2009). Vaccination with alvac and aidsvax to prevent hiv-1 infection in thailand. *New England Journal of Medicine*, 361:2209–2220.

Robins, J. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology-Methodological Issues, eds. N.Jewell, K. Dietz, and V.Farewell, Boston:Birkhauser*, pages 24–33.

Robins, J., Rotnitzky, A., and Zhao, L. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89:846–866.

Samuelsen, S., Ånested, H., and Skrondal, A. (2007). Stratified case-cohort analysis of general cohort sampling designs. *Scandinavian Journal of Statistics*, 34:103–119.

Scheike, T. and Zhang, M. (2002). An additive-multiplicative cox-aalen regression model. *Scandinavian Journal of Statistics*, 29(1):75–88.

Scheike, T. and Zhang, M. (2003). Extensions and applications of the cox-aalen survival model. *Biometrics*, 59:1036–1045.

Scheike, T. and Zhang, M. (2008). Flexible competing risks regression modelling and goodness-of-fit. *Lifetime Data Analysis*, 14:464–483.

Scheike, T., Zhang, M., and Gerds, T. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, 95(1):205–220.

Self, S. and Prentice, R. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *The Annals of Statistics*, 16:64–81.

Sen, P. (1979). Nonparametric tests for interchangeability under competing risks. *Contributions to Statistics*, J. Hajek Memorial Volume (Ed. Jureckeva,J.), Reidel, Dordrecht:224–228.

Shen, Y. and Cheng, S. (1999). Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics*, 55:1093–1100.

Sørensen, P. and P.K., A. (2000). Competing risks analysis of the case-cohort design. *Biometrika*, 87:49–59.

Sun, Y. (2001). Generalized nonparametric test procedures for comparing multiple cause-specific hazard rates. *Journal of Nonparametric Statistics*, 13:171–207.

Sun, Y., Qian, X., Shou, Q., and Gilbert, P. (2016). Analysis of two-phase sampling data with semiparametric additive hazards models. *Lifetime Data Analysis*, pages 1–23.

Sun, Y. and Tiwari, R. (1995). Comparing cause-specific hazard rates of a competing risks model with censored data. *IMS Lecture Notes-Monograph Series*, 27:255–270, Analysis of Censored data, edited by Koul, H. L. and Deshpandé, J.V.

Tsiatis, A. A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72:20–22.

Yang, G., Sun, Y., Qi, L., and Gilbert, P. (2016). Estimation of stratified mark-specific proportional hazards models under two-phase sampling with application to hiv vaccine efficacy trials. *Biostatistics*, 0(0):1–33.

Yates, N., Liao, H., Fong, Y., deCamp, A., Vandergrift, N., Williams, W., Alam, S., Ferrari, G., Yang, Z., Seaton, K., Berman, P., Alpert, M., Evans, D., OConnell, R., Francis, D., Sinangil, F., Lee, C., Nitayaphan, S., Rerks-Ngarm, S., Kaewkungwal, J., Pitisuttithum, P., Tartaglia, J., Pinter, A., Zolla-Pazner, S., Gilbert, P., Nabel, G., Michael, N., Kim, J., Montefiori, D., Haynes, B., and Tomaras, G. (2014.). Vaccine-induced env v1-v2 igg3 correlates with lower hiv-1 infection risk and declines soon after vaccination. *Science Translational Medicine*, 6:228–239.

Yip, P. and Lam, K. (1992). A class of nonparametric tests for the equality of failure rates in a competing risks model. *Comm. Statist. A-Theory Methods*, 21:2541–2556.

Zhang, M., Zhang, X., and Scheike, T. (2008). Modeling cumulative incidence function for competing risks data. *Expert Rev Clin Pharmacol*, 1(3):391–400.

Zolla-Pazner, S., deCamp, A., Gilbert, P., Williams, C., Yates, N., Williams, W., Howington, R., Fong, Y., Morris, D., Soderberg, K., Irene, C., Reichman, C., Pinter, A., Parks, R., Pitisuttithum, P., Kaewkungwal, J., Rerks-Ngarm, S., Nitayaphan, S., Andrews, C., OConnell, R., Yang, Z., Nabel, G., Kim, J., Michael, N., Montefiori, D., Liao, H., Haynes, B., and Tomaras, G. (2014). Vaccine-induced igg antibodies to v1v2 regions of multiple hiv-1 subtypes correlate with decreased risk of hiv-1 infection. *PLoS One*, 9:(2):e87572.

APPENDIX A: PROOFS IN CHAPTER 4

Lemmas

Lemma A.1. *Suppose* $X_n \xrightarrow{d} X$. *Then,* $X_n = Op(1)$.

Proof of Lemma A.1

Given $\epsilon > 0$, we choose sufficiently large k so that $P(|X| > k) < \epsilon$. By the assumption, $P(|X_n| > k) \to P(|X| > k)$. There exists some m such that for $n \geq m$, $P(|X_n| > k) < \epsilon$. We also choose sufficiently large k_1 so that $P(|X_i| > \epsilon) < \epsilon$, for $i = 1, ..., m-1$. Then, for $k_0 = \max(k, k_1)$, we have $P(|X_n| > k_0) < \epsilon$ for all n.

□

Proof of 4.13

Let the fourth term of (4.12) be

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{\gamma}}(\hat{\theta}) &= n^{-\frac{1}{n}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t) - \boldsymbol{K}(t,\hat{\theta})\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right] w_i(t) \\ &\qquad \times \boldsymbol{A}_i(\theta_0)\Delta_i N_i(t) \left\{\frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i)G(T_i)}\right\} dt(\hat{\theta} - \theta_0).\end{aligned}$$

By using (A.3), it can be divided into two parts

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{\gamma}}(\hat{\theta}) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t) - \boldsymbol{k}(t,\theta_0)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right] w_i(t) \\ &\qquad \times \boldsymbol{A}_i(\theta_0)\Delta_i N_i(t) \left\{\frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i)G(T_i)}\right\} dt(\hat{\theta} - \theta_0) \\ &\quad - n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{K}(t,\hat{\theta}) - \boldsymbol{k}(t,\theta_0)\right] \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \\ &\qquad \times \boldsymbol{A}_i(\theta_0)\Delta_i N_i(t) \left\{\frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i)G(T_i)}\right\} dt(\hat{\theta} - \theta_0). \qquad (A.1)\end{aligned}$$

In order to prove related $\boldsymbol{K}(t,\hat{\theta})$ term, we use the following properties. First, since $n^{\frac{1}{2}}(\hat{\theta}-\theta_0)=O_p(1)$, by using Delta method, we have

$$n^{\frac{1}{2}}(\boldsymbol{K}(t,\hat{\theta})-\boldsymbol{K}(t,\theta_0))=O_p(1). \tag{A.2}$$

Second, from definition of $\boldsymbol{K}(t,\theta_0)$,

$$\boldsymbol{K}(t,\theta_0)=\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t)w_i(t)\boldsymbol{\psi}_i(\theta_0)\boldsymbol{D}_{\boldsymbol{\eta},i}(t)\right]\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)w_i(t)\boldsymbol{\psi}_i(\theta_0)\boldsymbol{D}_{\boldsymbol{\eta},i}(t)\right]^{-1}.$$

By law of large numbers, we have

$$\boldsymbol{K}(t,\theta_0)\xrightarrow{p}\boldsymbol{k}(t,\theta_0). \tag{A.3}$$

It follows by uniform consistency of $\widehat{G}(t)$ in condition (I.6) and (A.1), the first term of (A.1) is, for $G(\tau)>0$,

$$\begin{aligned}
& n^{\frac{1}{2}}(\hat{\theta}-\theta_0)\ n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right]w_i(t) \\
&\qquad\qquad \times\boldsymbol{A}_i(\theta_0)\Delta_i N_i(t)\left\{\frac{G(T_i)-\widehat{G}(T_i)}{\widehat{G}(T_i)G(T_i)}\right\}dt \\
\leq\ & O_p(1)\times n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right]w_i(t) \\
&\qquad\qquad \times\boldsymbol{A}_i(\theta_0)\Delta_i N_i(t)\left\{\frac{1}{\widehat{G}(\tau)G(\tau)}\right\}\sup_{0\leq t\leq\tau}|G(t)-\widehat{G}(t)|\,dt \\
=\ & O_p(1)o_p(1) \\
=\ & o_p(1).
\end{aligned} \tag{A.4}$$

It follows by (A.2), (A.3) and (I.6) that the last term of (A.1) is equal to

$$
\begin{aligned}
& -n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{K}(t,\hat{\theta}) - \boldsymbol{k}(t,\theta_0) \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) \right] w_i(t) \\
& \qquad \times \boldsymbol{A}_i(\theta_0) \Delta_i N_i(t) \left\{ \frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i) G(T_i)} \right\} dt (\hat{\theta} - \theta_0) \\
= \; & -(\hat{\theta} - \theta_0) \times n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{K}(t,\hat{\theta}) - \boldsymbol{K}(t,\theta_0) \right] \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \\
& \qquad \times \boldsymbol{A}_i(\theta_0) \Delta_i N_i(t) \left\{ \frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i) G(T_i)} \right\} dt \\
& -n^{\frac{1}{2}} (\hat{\theta} - \theta_0) \times n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{K}(t,\theta_0) - \boldsymbol{k}(t,\theta_0) \right] \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \\
& \qquad \times \boldsymbol{A}_i(\theta_0) \Delta_i N_i(t) \left\{ \frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i) G(T_i)} \right\} dt \\
= \; & -o_p(1) \times O_p(1) \\
& \times n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \boldsymbol{A}_i(\theta_0) \Delta_i N_i(t) \left\{ \frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i) G(T_i)} \right\} dt \\
& -O_p(1) \times o_p(1) \\
& \times n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \boldsymbol{A}_i(\theta_0) \Delta_i N_i(t) \left\{ \frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i) G(T_i)} \right\} dt. \qquad \text{(A.5)}
\end{aligned}
$$

By the similar argument to (A.4) and Slutsky's theorem, the above equation reduce to $o_p(1)$. It follows by (A.1), (A.4) and (A.5) that the fourth term of (4.12) is $\boldsymbol{C}_{\boldsymbol{\gamma}}(\hat{\theta}) \xrightarrow{p} 0$ uniformly in $t \in [0, \tau]$. □

Proof of 4.14

From (4.14), we have

$$\begin{aligned} & n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{K}(t, \hat{\theta}) - \boldsymbol{k}(t, \theta_0) \right] \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t) \right\} dt \\ = & \; n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{K}(t, \hat{\theta}) - \boldsymbol{K}(t, \theta_0) \right] \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t) \right\} dt \\ & + n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \left[\boldsymbol{K}(t, \theta_0) - \boldsymbol{k}(t, \theta_0) \right] \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \\ & \qquad \times \boldsymbol{\psi}_i(\theta_0) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t) \right\} dt. \end{aligned} \tag{A.6}$$

By (2.2), we note that

$$n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \frac{\partial \boldsymbol{K}(t, \theta)}{\partial \theta} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t) \right\} dt$$

has mean zero. By the central limit theorem and Lemma (A.1), we have

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \frac{\partial \boldsymbol{K}(t, \theta)}{\partial \theta} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t) \right\} dt = O_p(1).$$

By the Taylor expansion of $\boldsymbol{K}(t, \hat{\theta})$ around the true value θ_0, the second term of

(A.6) is equal to

$$
\begin{aligned}
& n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_{0}^{\tau} \left[\boldsymbol{K}(t,\hat{\theta}) - \boldsymbol{K}(t,\theta_0)\right] \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\} dt \\
= \;& n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_{0}^{\tau} \left[\frac{\partial \boldsymbol{K}(t,\theta)}{\partial \theta}(\hat{\theta}-\theta_0) + o_p(n^{-\frac{1}{2}})\right] \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \\
& \qquad \times \boldsymbol{\psi}_i(\theta_0) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\} dt \\
= \;& (\hat{\theta}-\theta_0) \times n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_{0}^{\tau} \frac{\partial \boldsymbol{K}(t,\theta)}{\partial \theta} \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\} dt \\
& + o_p(1) \times n^{-1} \sum_{i=1}^{n} \int_{0}^{\tau} \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\} dt \\
= \;& o_p(1) \times O_p(1) + o_p(1) \\
= \;& o_p(1).
\end{aligned}
$$

With similar arguments, by the central limit theorem, we have

$$
n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_{0}^{\tau} \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\} dt = O_p(1).
$$

It follows by Slutsky's theorem that the second term of (A.6) is equal to

$$
\begin{aligned}
& n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_{0}^{\tau} \left[\boldsymbol{K}(t,\theta_0) - \boldsymbol{k}(t,\theta_0)\right] \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\} dt \\
& = o_p(1) \times n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_{0}^{\tau} \boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t) w_i(t) \boldsymbol{\psi}_i(\theta_0) \left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\} dt \\
& = o_p(1) \times O_P(1) \\
& = o_p(1),
\end{aligned}
$$

where $\boldsymbol{K}(t,\theta_0) - \boldsymbol{k}(t,\theta_0) = o_p(1)$ uniformly in $t \in [0,\tau]$. □

Proof of 4.26

From (4.25), it can be decomposed into two parts.

$$
\begin{aligned}
n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta}) &= n^{-1}\sum_{i=1}^{n}\int_{0}^{\tau}\left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t)-\boldsymbol{K}(t,\hat{\theta})\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right]w_i(t)\boldsymbol{\psi}_i(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt \\
&= n^{-1}\sum_{i=1}^{n}\int_{0}^{\tau}\left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right]w_i(t)\boldsymbol{\psi}_i(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt \qquad \text{(A.7)} \\
&- n^{-1}\sum_{i=1}^{n}\int_{0}^{\tau}\left[\boldsymbol{K}(t,\hat{\theta})-\boldsymbol{k}(t,\theta_0)\right]\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)w_i(t)\boldsymbol{\psi}_i(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt \qquad \text{(A.8)} \\
&= n^{-1}\sum_{i=1}^{n}\int_{0}^{\tau}\left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right]w_i(t)\boldsymbol{\psi}_i(\hat{\theta})\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt \\
&\quad +o_p(1). \qquad \text{(A.9)}
\end{aligned}
$$

By using (4.10), (A.9) can be decomposed as

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\int_{a}^{\tau}\left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right]w_i(t)\boldsymbol{\psi}_i(\theta_0)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt \\
& +n^{-1}\sum_{i=1}^{n}\int_{a}^{\tau}\left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right]w_i(t) \\
& \qquad\qquad \times\left(\boldsymbol{A}_i(\theta_0)(\hat{\theta}-\theta_0)+o_p(n^{-\frac{1}{2}})\right)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt+o_p(1) \\
= {} & n^{-1}\sum_{i=1}^{n}\int_{a}^{\tau}\left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right]w_i(t)\boldsymbol{\psi}_i(\theta_0)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt \\
& +(\hat{\theta}-\theta_0)\times n^{-1}\sum_{i=1}^{n}\int_{a}^{\tau}\left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right]w_i(t)\boldsymbol{A}_i(\theta_0)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt \\
& +o_p(n^{-\frac{1}{2}})\times n^{-1}\sum_{i=1}^{n}\int_{a}^{\tau}\left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right]w_i(t)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt+o_p(1) \\
= {} & n^{-1}\sum_{i=1}^{n}\int_{a}^{\tau}\left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right]w_i(t)\boldsymbol{\psi}_i(\theta_0)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt+o_p(1),
\end{aligned}
$$

followed by sums of the last three terms in second equality is equal to $o_p(1)$.

Let

$$
\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0) = n^{-1}\sum_{i=1}^{n}\int_{a}^{\tau}\left[\boldsymbol{D}_{\boldsymbol{\gamma},i}^{\mathsf{T}}(t)-\boldsymbol{k}(t,\theta_0)\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t)\right]w_i(t)\boldsymbol{\psi}_i(\theta_0)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt.
$$

Therefore, we have

$$n^{-1}\,\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\hat{\theta}) = n^{-1}\,\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0) + o_p(1).$$

Let

$$\boldsymbol{Q}(\theta_0) \;=\; E\left\{\int_a^\tau \left[\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\gamma},i}(t) - \boldsymbol{k}(t,\theta_0)\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)\right] w_i(t)\boldsymbol{\psi}_i(\theta_0)\boldsymbol{D}_{\boldsymbol{\gamma},i}(t)\,dt\right\}.$$

By the law of large numbers,

$$n^{-1}\,\boldsymbol{\mathcal{I}}_{\boldsymbol{\gamma}}(\theta_0) \xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{\gamma}}(\theta_0).$$

□

Proof of 4.35

We consider the third term of (4.34). By the law of large numbers,

$$n^{-1}\sum_{i=1}^{n}\boldsymbol{D}^{\mathsf{T}}_{\boldsymbol{\eta},i}(t)w_i(t)\boldsymbol{A}_i(\theta_0)\left\{\frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t)\right\}$$

has mean zero.

By (4.10) and (A.1), we have

$$\begin{aligned}
& n^{-\frac{1}{2}} \sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t)(\boldsymbol{\psi}_i(\hat{\theta}) - \boldsymbol{\psi}_i(\theta_0)) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t) \right\} \\
= \;& n^{-\frac{1}{2}} \sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \left(\boldsymbol{A}_i(\theta_0)(\hat{\theta} - \theta_0) + o_p(n^{-\frac{1}{2}}) \right) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t) \right\} \\
= \;& n^{\frac{1}{2}}(\hat{\theta} - \theta_0) \times n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \boldsymbol{A}_i(\theta_0) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t) \right\} \\
+ \;& o_p(1) \times n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \left\{ \frac{\Delta_i N_i(t)}{G(T_i)} - \boldsymbol{F}_{1i}(t) \right\} \\
= \;& O_p(1) \times o_p(1) + o_p(1) \\
= \;& o_p(1).
\end{aligned} \tag{A.10}$$

□

Proof of 4.36

With similar argument, the fourth term of (4.34) is

$$\begin{aligned}
& n^{-\frac{1}{2}} \sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t)(\boldsymbol{\psi}_i(\hat{\theta}) - \boldsymbol{\psi}_i(\theta_0)) \left\{ \frac{\Delta_i N_i(t)}{\widehat{G}(T_i)} - \frac{\Delta_i N_i(t)}{G(T_i)} \right\} \\
= \;& n^{-\frac{1}{2}} \sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \left(\boldsymbol{A}_i(\theta_0)(\hat{\theta} - \theta_0) + o_p(n^{-\frac{1}{2}}) \right) \Delta_i N_i(t) \left\{ \frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i) G(T_i)} \right\} \\
= \;& n^{\frac{1}{2}}(\hat{\theta} - \theta_0) \times n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \boldsymbol{A}_i(\theta_0) \Delta_i N_i(t) \left\{ \frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i) G(T_i)} \right\} \\
+ \;& o_p(1) \times n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \Delta_i N_i(t) \left\{ \frac{G(T_i) - \widehat{G}(T_i)}{\widehat{G}(T_i) G(T_i)} \right\}
\end{aligned} \tag{A.11}$$

Similar to (4.13), by the consistency of $\widehat{G}(t)$, the fourth term of (4.34) is equal to

$o_p(1)$ uniformly in $t \in [0, \tau]$. $\square$

Proof of 4.41

Let

$$\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t, \theta_0) = \sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \psi_i(\theta_0) \boldsymbol{D}_{\boldsymbol{\eta},i}(t).$$

By using (4.10) and the consistency of $\hat{\theta}$, we have

$$\begin{aligned} n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t, \hat{\theta}) &= n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \psi_i(\theta_0) \boldsymbol{D}_{\boldsymbol{\eta},i}(t) \\ &\quad + (\hat{\theta} - \theta_0) \times n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \boldsymbol{A}_i(\theta_0) \boldsymbol{D}_{\boldsymbol{\eta},i}(t) + o_p(n^{-\frac{1}{2}}) \\ &= n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \psi_i(\theta_0) \boldsymbol{D}_{\boldsymbol{\eta},i}(t) + o_p(1) + o_p(n^{-\frac{1}{2}}). \end{aligned} \tag{A.12}$$

Therefore, we have

$$n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t, \hat{\theta}) = n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t, \theta_0) + o_p(1).$$

Let

$$\boldsymbol{Q}_{\boldsymbol{\eta}}(t, \theta_0) = E\{\boldsymbol{D}_{\boldsymbol{\eta},i}^{\mathsf{T}}(t) w_i(t) \psi_i(\theta_0) \boldsymbol{D}_{\boldsymbol{\eta},i}(t)\}.$$

By the law of the large numbers, we have

$$n^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}(t, \theta_0) \xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{\eta}}(t, \theta_0).$$

$\square$

Proof of 4.60

By the theory of Robins, Rotnizky and Zhao (1994), we have the fact that $\widehat{\eta}(t)$ is consistent estimator of $\eta_0(t)$. By the missing at random assumption,

$$n^{-1}\int_0^\tau \sum_{i=1}^n \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^\mathsf{T} w_i(t)(1-\psi_i(\theta_0))V_{x,i}\,dt$$

has mean zero. Then, by the central limit theorem and Lemma A.1, we have

$$n^{-\frac{1}{2}}\int_0^\tau \sum_{i=1}^n \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^\mathsf{T} w_i(t)(1-\psi_i(\theta_0))V_{x,i}\,dt = O_p(1).$$

It follows by Slutsky's theorem that

$$\begin{aligned}
& -n^{-\frac{1}{2}}\int_0^\tau \sum_{i=1}^n \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^\mathsf{T} w_i(t)(1-\psi_i(\theta_0))V_{x,i}\,\{\widehat{\eta}(t)-\eta_0(t)\}\,dt \\
= & -\{\widehat{\eta}(t)-\eta_0(t)\}\,n^{-\frac{1}{2}}\int_0^\tau \sum_{i=1}^n \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^\mathsf{T} w_i(t)(1-\psi_i(\theta_0))V_{x,i}\,dt \\
= & \; o_p(1)\times O_p(1) \\
= & \; o_p(1). \qquad \text{(A.13)}
\end{aligned}$$

□

Proof of 4.61

With similar argument in the proof of (4.60), we have

$$\begin{aligned}
n^{-\frac{1}{2}}\int_0^\tau \sum_{i=1}^n \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^\mathsf{T} w_i(t)(1-\psi_i(\theta_0))\left\{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma}\right\} dt &= O_p(1), \\
n^{-1}\int_0^\tau \sum_{i=1}^n \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^\mathsf{T} w_i(t)(1-\psi_i(\theta_0))\,dt &= o_p(1),
\end{aligned}$$

and also have the fact that $\hat{\gamma}$ is consistent estimator of γ_0.

By the taylor expansion of $g(\widehat{\gamma}, Z_i, t)$ around the true value γ_0, we have

$$\begin{aligned}
& -n^{-\frac{1}{2}} \int_0^{\tau} \sum_{i=1}^{n} \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0)) \{g(\widehat{\gamma}, Z_i, t) - g(\gamma_0, Z_i, t)\} \, dt \\
= & -n^{-\frac{1}{2}} \int_0^{\tau} \sum_{i=1}^{n} \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0)) \\
& \qquad \times \left\{ \frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma} (\hat{\gamma} - \gamma_0) + o_p(n^{-\frac{1}{2}}) \right\} dt \\
= & -(\hat{\gamma} - \gamma_0) \times n^{-\frac{1}{2}} \int_0^{\tau} \sum_{i=1}^{n} \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0)) \left\{ \frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma} \right\} dt \\
- & \ o_p(1) \times n^{-1} \int_0^{\tau} \sum_{i=1}^{n} \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\}^{\mathsf{T}} w_i(t)(1-\psi_i(\theta_0)) \, dt \\
= & \ o_p(1) \times O_p(1) + o_p(1) \\
= & \ o_p(1).
\end{aligned}$$

□

Proof of 4.68

With similar argument in the proof of (4.61), we have

$$-n^{-\frac{1}{2}} \sum_{i=1}^{n} V_{x,i}^{T} w_i(t)(1-\psi_i(\theta_0)) \{\frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0}\} = O_p(1). \tag{A.14}$$

It follows that

$$
\begin{aligned}
& -n^{-\frac{1}{2}} \sum_{i=1}^{n} V_{x,i}^{T} w_i(t)(1-\psi_i(\theta_0)) \left\{ g(\hat{\gamma}, Z_i, t) - g(\gamma_0, Z_i, t) \right\} \\
= & -n^{-\frac{1}{2}} \sum_{i=1}^{n} V_{x,i}^{T} w_i(t)(1-\psi_i(\theta_0)) \left\{ \frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0} (\hat{\gamma} - \gamma_0) + o_p(n^{-\frac{1}{2}}) \right\} \\
= & -(\hat{\gamma} - \gamma_0) \times n^{-\frac{1}{2}} \sum_{i=1}^{n} V_{x,i}^{T} w_i(t)(1-\psi_i(\theta_0)) \{ \frac{\partial g(\gamma_0, Z_i, t)}{\partial \gamma_0} \} + o_p(1) \\
= & \; o_p(1) \times O_p(1) + o_p(1) \\
= & \; o_p(1). \qquad \text{(A.15)}
\end{aligned}
$$

□

Proof of 4.69

With similar argument to (4.60), it can be proven to be equal to $o_p(1)$. □