STATISTICAL METHODS FOR DATA FROM CASE-COHORT STUDIES

Poulami Maitra

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2018

Approved by:

Jianwen Cai

Chirayath Suchindran

Donglin Zeng

Haibo Zhou

Jeannette Bensen

ABSTRACT

Poulami Maitra: Statistical Methods for Data from Case-Cohort Studies
(Under the direction of Jianwen Cai)

In epidemiological studies and disease prevention trials, interest often lies in the relationship between certain disease endpoint and some exposure of interest. When the event is rare and/or some of the covariate information are quite expensive to collect for the entire cohort, case-cohort designs are widely used to reduce the financial cost of the study while achieving the same study goals. The case-cohort sampling scheme entails the random sampling of individuals, called the sub-cohort, along with all the cases. In the situation when the event rate is not low but resources are limited, the generalized case-cohort design is more appropriate, where only a fraction of cases are sampled along with the sub-cohort. In this dissertation, we consider two aspects of case-cohort studies. One is for statistical methods for the analysis of recurrent events and the other concerns power/sample size calculation for interaction test.

Many methods for the analysis of data from case-cohort studies have been proposed in the literature. However, most of these methods are for either a single event or multitude of events of different types on the same subject. There has not been much work on the recurrent events data under case-cohort sampling scheme. Valid statistical methods that take into account the correlation between the events from the same individual needs to be developed. In this dissertation, the first two topics are related to recurrent events. We consider modeling the recurrent events using the rate model under the original and generalized case-cohort designs. The first topic considers the multiplicative rates model and the second topic considers additive rates models. For both types of the rate models, we propose weighted estimating equation approach for the parameter estimates for both sampling designs. We showed that the proposed estimators are consistent and asymptotically normally distributed. We conducted simulation studies to examine the performance of our proposed estimators in finite samples and they performed well. For the multiplicative rates model, we illustrated the proposed method to assess the relationship between prior measles infection and acute lower-respiratory-infections (ALRI) in a double-blinded randomized clinical trial, conducted in Brazil. We illustrated our proposed method for additive rates model to study the effect of FEV_1 on the recurrence of pulmonary exacerbation in patients with cystic fibrosis.

In the third topic, we address another aspect of the case-cohort design. All the previous work in the literature concern sample size and power calculation in case-cohort data for a dichotomized main effect. However, in certain situations, one might be interested in the association of a covariate and time to event response in different biomarker groups, which may be expensive to measure. We extend the existing idea for the single binary main effect to the interaction between the variable and the dichotomized biomarker in the presence of a rare event. We propose different power formulas based on the simplification of a generalized log-rank test for the case-cohort design. A cost efficiency formula comparing the case-cohort design to a simple random sample is derived. We examine the performance of the bounds based on the same test. Simulation studies are conducted to illustrate the efficiency for the case-cohort design. We illustrate the use of the formula based on information from the pooled databases of Lung Adjuvant Cisplatin Evaluation (LACE) and Cancer and Leukemia Group B (CALGB) 9633.

To the people who are there only in my memories.

I miss you and I hope I made you proud.

ACKNOWLEDGEMENTS

Knowledgeable people of the internet have blogged about what to write in one's thesis acknowledgment (yes, you read that right) and so, following them, I am going to (try to) make it short and just mention the few things and people that affected and shaped me during my time at UNC. If I had to describe my PhD experience at UNC in a single sentence, I would quote Glen Hansard, "It's gonna be a long one/ I'll be working all night long/ It's gonna be a long one/ But I'm paying my way". Even though my diploma would say PhD (Biostatistics), I've gotten a degree in life lessons.

Well, we are all social animals and I am no exception. These five years were spent in the presence of some amazing people, who have helped me grow as a person and a researcher. The least I can do, is mention their names. The first would be my adviser, Dr. Jianwen Cai. I am forever grateful to her for helping me with everything, whether it is my dissertation, or dealing with frustrations, or even allowing me to stay in Charlotte with Pourab. I am also indebted to all my other committee members (Dr. Suchindran, Dr. Zeng, Dr. Zhou and Dr. Bensen) for being so helpful with the dissertation. I would like to thank my parents and my parents-in-law for freaking out about every little thing in my (our) life, trying to solve them from eight thousand miles away and irrevocably supporting me through everything.

Coming to the group of people who made my life interesting and this journey, a little less arduous, the first name that pops up is Neha. Another soul in the department who joined UNC the same year I did and have been climbing the PhD wall with me. The board game nights we have with her and Raul are one of the very few activities that I look forward to. Ipsita (my longest apartment-mate) who was always there whenever I needed advise or vent out. Jyotishka and Shalini for the amazing conversations we've had and their 6 year old beagle-jack russel terrier, Max for the amazing month we spent with him. I am extremely thankful to Sayan, Ritwik, Abhishek, Sujatro, Monica, Suman, Arkopal and Priyam for the happy memories, school and college friends whose conversations have always been like a breath of fresh air and Birgitte, for being there. And to our 6 month old puppy Luna, thank you for teaching me patience and that love is unconditional. Finally (almost), I would like to thank my baby sister for keeping me updated in almost all things gossip and keeping my tastes young (literally, from her birth).

Lastly, I would like to thank the person without whom I would not have been here. Thank you Pourab for believing in me and not running away, even after realizing how crazy I am. As Oscar Wilde said, “To live is the rarest thing in the world. Most people exist, that is all.” Thank you for teaching me to live.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Large epidemiologic cohort studies or disease prevention trials are expensive as they require the follow-up of several thousand individuals for a long period of time before producing valuable results (Prentice 1986). The cost mainly arises from the culmination of the raw materials for covariate data ascertainment for all the cohort members. Such raw materials may include blood serum samples, tissue specimens, occupational exposure records, etc. When the disease is rare, much of the covariate data on the disease-free subjects is inessential. Prentice (1986) cited the multiple risk factor intervention trial (MRFIT Research Group, 1982), a randomized trial of 12,866 subjects, who were followed-up for an average of seven years and reportedly cost US $100 million. Such studies provide a primary application area for the case-cohort design. In the prevention trial context, a case-cohort design would reduce the cost of assembling the cohort history while allowing for the tracking of the history for the subcohort on an on-going basis.

The case-cohort study design was first proposed by Prentice (1986). It is a retrospective study design nested within a prospective cohort. Specifically, the entire cohort is followed for the disease of interest over time. At a certain time during follow-up, a case-cohort sample is drawn. The case-cohort sample consists of a random sample from the cohort and all subjects who had developed the disease by that time. The expensive or hard to measure covariates are then collected on subjects in the case-cohort sample. Prentice (1986) studied the proportional hazards model for case-cohort data and obtained an estimating equation using pseudo-likelihood approach. In this dissertation, we propose methods to analyze recurrent event data from case-cohort studies.

Recurrent events are common in biomedical research. Andersen and Gill (1982) extended the Cox proportional hazards model to include recurrent events by modeling the intensity process of the multivariate counting process. Prentice et al. (1981), Wei et al. (1989), Lawless and Nadeau (1995), Sun et al. (2004), among others considered marginal rates (or mean) model to evaluate the effect of the risk factors on recurrent event data. Some examples include the occurrence of new tumours in patients with superficial bladder cancer (Byar, 1980), recurrent seizures in epileptic patients (Albert 1991), rejection episodes in patients receiving kidney transplants (Cole et al. 1994), repeated infections in HIV-patients (Li and Lagakos 1997) and repeated cardiovascular events in patients (Cui et al. 2008). The two primary frameworks to study the association between the risk factors and the disease recurrence are the additive and multiplicative rates models. Most modern analysis of survival data address multiplicative models for relative risk (rates) using proportional rates models, mostly due to desirable theoretical properties along with the easy interpretation of results (Pepe and Cai 1993, Lawless 1995, Lin et al. 2000, Schaubel et al. 2006, Kang and Cai 2009a). However, researchers may be interested in the risk (rate) difference, rather than the relative measure, attributed to the exposure. Further, the risk difference is more relevant to the public health as it translates directly to the number of disease cases that may be avoided by eliminating the exposure (Kulich and Lin 2000). Consequently, the additive rates models can be considered as an alternative to the multiplicative model (Schaubel et al. 2006, Yin and Cai 2004, Zeng and Cai 2010, Liu et al. 2013, Kang et al. 2013, He et al. 2013).

All the aforementioned articles considered the full cohort. However, in the three decades from Prentice's (1986) seminal paper on case-cohort data, there has been many theoretical developments and application to different studies in the literature (e.g. Self and Prentice (1988), Lin and Ying (1993), Barlow (1994), Borgan et al. (1995), Chen (2001), Zeng et al. (2006), Kang and Cai (2009a), Dong et al. (2014)). Estimation procedures have been proposed to model single-event data or clustered failure times data, arising from case-cohort studies but methodologies to address the rates model for such data are limited. Lu and Tsiatis (2006) and Zhang et al. (2011) developed marginal models for clustered failure times data, in which the clusters are usually formed by the dependent subjects. Chen and Chen (2014) extended Prentice's (1986) idea to recurrent events with certain clustering feature, which was represented by properly modified Cox-type self-exciting intensity model. However, there has not been much work on modeling the marginal rates/mean model for recurrent events under such sampling scheme. Using the intensity model (Andersen and Gill 1982) to analyze recurrent events assumes that all the influence of the prior events on future recurrence is through only the possibly time-varying covariates (Lin et al. 2000). Since, this may not be the case in practice, less restrictive methods need to be developed for such data from case-cohort studies. Motivated by these, we propose statistical methods for modeling recurrent events data from case-cohort studies. We will consider both the multiplicative and additive rates models in analyzing recurrent events data from case-cohort studies.

Another aspect of case-cohort studies is the design of such studies. In this new era of cancer research, large amounts of data are often available after the completion of Phase III trials. There is a growing interest in using this available data to discover new biomarkers that can be helpful in predicting the best treatment for a particular patient. The advantages of using information collected in completed Phase III trials are : (a) the data is already available for further analysis and (b) even if the clinical trail may indicate that the treatment is not effective, the collected information can still be used for biomarker discovery. In developing personalized treatment program, genomic biomarkers are often used which are often quite expensive to measure and maybe time-consuming. In such cases, case-cohort study design can be applied if only a very small proportion of the cohort experiences the disease endpoint of interest. The selection of the subset in the second phase can be flexible, utilizing information collected in the first phase, such as the patient's disease status, treatment effects and any auxiliary information. Power and/or sample size will need to be considered to design such a study. Cai and Zeng (2004) and Cai and Zeng (2007)) proposed simple formula to calculate the power for the main effect under case-cohort studies and the bounds under generalized case-cohort design, respectively. In this dissertation, we will develop power/sample size formula for testing the interaction between an exposure of interest and some expensive biomarker on a rare event under case-cohort design.

CHAPTER 2: LITERATURE REVIEW

In this Chapter, we review the literature on the statistical methods for : (i) univariate failure time data arising from case-cohort studies, (ii) correlated failure time data, more specifically, recurrent events from prospective studies and their marginal analysis, (iii) sample size calculation for effects in presence of case-cohort studies when the event is rare. We review the literature on statistical methods for univariate failure time data in Section 2.1, case-cohort studies in Section 2.2, recurrent events analysis from prospective studies in Section 2.3 (including marginal models using multiplicative models and frailty models), additive rates models in Section 2.4, and Power/sample size calculation for case-cohort data in Section 2.5.

2.1 Univariate Failure Time Models

The Cox proportional hazards model (Cox 1972) has been one of the most widely used procedures to study the effects of covariates on failure time. This model assumes that the effect of the covariate on the hazard function is constant over time. A more general version of the Cox model assumes that the hazard function of the failure time, T, associated with the covariates Z, is given by

$$\lambda(t \mid Z(t)) = \lambda_0(t) exp\left\{\beta_0' Z(t)\right\}, \tag{2.1}$$

where $\lambda_0(t)$ is the unspecified baseline hazard function and β_0 is a p-dimensional vector of unknown parameters.

Let C denote the censoring time and X = min(T, C) denote the observed time. Let N(t) be the counting process and $Y(t) = \mathbf{I}(X \geqslant t)$ be the at-risk process. Further, let us define $\Delta = \mathbf{I}(T \leqslant C)$ as the indicator of failure. Each failure time is assumed to be subject to independent right-censorship. Let $(X_i, \Delta_i, \mathbf{Z}_i); i = 1, 2, \ldots, n$ denote the i-th independent copy of $(X, \Delta, \mathbf{Z})$ and τ denotes the study end point. The regression parameter can be estimated by the partial likelihood score function introduced by (Cox 1975) :

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left(Z_i(t) - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right) dN_i(t), \tag{2.2}$$

where $S^{(0)}(\beta, t) = \frac{1}{n} \sum_{j=1}^{n} Y_j(t) exp(\beta Z_j(t))$ and $S^{(1)}(\beta, t) = \frac{1}{n} \sum_{j=1}^{n} Y_j(t) Z_j(t) exp(\beta Z_j(t))$. Under some regularity conditions, the maximum partial likelihood estimator, $\hat{\beta}$, defined as the solution to the score equation, $U(\beta) = 0$ converges to a normal distribution as $n \to \infty$ with mean β_0 and a variance which can be consistently estimated by $- \left[\frac{\delta}{\delta \beta'} U(\beta) \mid_{\beta = \hat{\beta}} \right]^{-1}$ (Andersen and Gill 1982).

2.2 Case-Cohort Studies

Epidemiologic studies are considered to be one of the most reliable methods for assessing the variation in rate of mortality and to study the effect of covariates on the rate in the population. When the event of interest is rare or the relationship that is of interest is complex, cohort studies would require large number of subjects and/or long periods of follow-up in order to accumulate enough failures to have sufficient statistical power to make meaningful conclusions. However, this would increase the cost of collecting such covariate information on all subjects, if feasible. Prentice (1986) proposed the case-cohort design to reduce the number of subjects for whom the covariate information is collected, hence reducing the overall cost. This method involves the selection of a random sample from the entire cohort, which is called the sub-cohort and the assembly of the covariate information on the individuals in this sub-cohort as well as the subjects who experienced the event of interest during the follow-up period. The sub-cohort also provides a basis for monitoring the covariates during the follow-up of the cohort. Studying the relative risk process is quite natural in understanding the effect of the covariate history on the event rates. Based on the relative risk regression model (Cox 1972), the hazard function of the i-th subject, at time t, is modeled by

$$\lambda(t \mid Z_i) = \lambda_0(t) r \left\{ \beta_0 Z_i(t) \right\}, \tag{2.3}$$

where r(.) is a known function with r(0) = 1. The pseudo-likelihood function for the estimation of β_0 has the form :

$$\tilde{L}(\beta_0) = \prod_{i=1}^{n} \left[\frac{r_{ii}}{\sum_{l \in \tilde{R}(t_i)} r_{li}} \right]^{\Delta_i}, \tag{2.4}$$

where $\tilde{R}(t) = D(t) \cup$ Sc, Sc is the set of all individuals in the sub-cohort, $D(t)$ is the set of all individuals who have observed a failure at time t. It can be written as $D(t) = \{i \mid N_i(t) \neq N_i(t-)\}$ and $r_{li} = Y_l(t_i) r\{\beta_0 Z_l(t_i)\}$. Covariate information is assumed to be available only for the set, $K(t) \cup$ Sc at time t, where $K(t) = \{i \mid N_i(t) = 1\}$. The maximum pseudo-likelihood estimate, $\hat{\beta}_{PL}$ is defined by $\tilde{U}(\hat{\beta}_{PL}) = 0$ where $\tilde{U}(\beta)$ is defined as

$$\tilde{U}(\beta) = \frac{\delta}{\delta\beta'} log \tilde{L}(\beta) = \sum_{i=1}^{n} \tilde{U}(\beta) = \sum_{i=1}^{n} \Delta_i \left(c_{ii} - \sum_{l \in \tilde{R}(t_i)} \frac{b_{li}}{\sum_{l \in \tilde{R}(t_i)} r_{li}} \right), \tag{2.5}$$

where $b_{li} = Y_l(t_i) Z_l(t_i) r\{\beta_0 Z_l(t_i)\}$; $c_{ii} = b_{ii} r'\{\beta_0 Z_i(t_i)\}$ and $r'(t) = \frac{\delta}{\delta\beta'} r(t)$. Prentice (1986) showed that the variance of $n^{-1/2} U(\beta_0)$ is given by

$$\tilde{V}(\beta) = I(\beta) + 2 \sum_{i=1}^{n} \Delta_i \tilde{\Delta}(t_i) \sum_{k \mid t_k < t_i} \Delta_k v_{ki},$$

where $I(\beta) = -\frac{\delta}{\delta\beta'} U(\beta), v_{ki} = -\sum_{j \in \tilde{R}(t_i)} \left(\frac{B_k + b_{ik} - b_{jk}}{R_k + r_{ik} - r_{jk}} \right)' \left(c_{ji} - \frac{B_i}{R_i} \right) r_{ji} R_i^{-1}, R_i = \sum_{k \in \tilde{R}(t_i)} r_{ki}$, $B_i = \sum_{k \in \tilde{R}(t_i)} b_{ki}$ and $\tilde{\Delta}(t) = 1$ if $\tilde{R}(t) \neq$ Sc, 0, o.w. Hence, by Taylor series expansion, we have the variance of $n^{1/2} \left(\hat{\beta}_{PL} - \beta_0 \right)$ is given by $I(\beta_0)^{-1} \tilde{V}(\beta_0) I(\beta_0)^{-1}$. A natural estimator of the cumulative baseline rate is proposed as

$$\hat{\Lambda}_0(t) = \tilde{n} n^{-1} \int_0^t \left[\sum_{l \in \text{Sc}} Y_l(u) r \left\{ \hat{\beta}'_{PL} Z_l(u) \right\} \right]^{-1} d\bar{N}(u),$$

where $\bar{N}(t) = \sum_{i=1}^{n} \bar{N}_i(t)$ and $\tilde{n}$ is the size of the random sub-cohort. Self and Prentice (1988) noted that the efficiency of the relative risk parameter estimation depends on the number of subjects experiencing the event.

Self and Prentice (1988) developed the asymptotic distribution theory for the case-cohort maximum pseudo-likelihood estimator and related quantities, with slightly different pseudolikelihood and variance estimator from the ones proposed by Prentice (1986). In their formulation of the risk set, only the sub-cohort individuals were considered, whereas in the original Prentice (1986) paper, the risk set included a non-subcohort individual that fails at a particular time point. In Self and Prentice (1988), they considered a similar relative risk regression model as (2.3). The maximum pseudolikelihood estimator, $\hat{\beta}_{sp}$, is defined as the solution to $U(\beta) = \frac{\delta}{\delta\beta'} log\tilde{L}(\beta) = 0$, where we have

$$log\tilde{L}(\beta) = \sum_{i \in C} \left[\int_0^\tau log\left[r\left\{\beta' Z_i(t)\right\}\right] dN_i(t) - \int_0^\tau log\left\{ \sum_{j \in \mathrm{Sc}} Y_j(t) r\left(\beta' Z_j(t)\right)\right\}\right] dN_i(t), \tag{2.6}$$

where C is the cohort and Sc is the subcohort of size $\tilde{n}$. Under some regularity conditions, they proved that $\hat{\beta}_{sp}$ converges in probability to β_0 and $n^{-1/2}U(\beta_0)$ converges to a Normal distribution with mean zero and variance $\Sigma(\beta_0)+\Delta(\beta_0)$, where $\Sigma(\beta) = -lim_{n\to\infty}\frac{1}{n}\frac{\partial^2}{\partial\beta\partial\beta'}log\tilde{L}(\beta)$ and $\Delta(\beta)$ consists of the contributions of the covariance among the components induced by the random sampling and have a complicated expression. Hence, by the Taylor series expansion, $n^{1/2}\left(\hat{\beta}_{sp} - \beta_0\right)$ converges to a Gaussian distribution with mean 0 and covariance matrix $\Sigma(\beta_0)^{-1} + \Sigma(\beta_0)^{-1}\Delta(\beta_0)\Sigma(\beta_0)^{-1}$. Self & Prentice (1988), in their paper, propose consistent estimators for $\Sigma(\beta_0)$ and $\Delta(\beta_0)$. For the cumulative hazard function, $\Lambda_0(t)$, the proposed estimator is given by

$$\Lambda_0(t) = \tilde{n}n^{-1} \int_0^t \left[\sum_{l \in \mathrm{Sc}} Y_l(u) r\left\{\hat{\beta}'_{sp} Z_l(u)\right\}\right]^{-1} d\bar{N}(u). \tag{2.7}$$

$n^{1/2}\left(\hat{\beta}_{sp} - \beta_0\right)$ and $n^{1/2}\left(\tilde{\Lambda}_0(t) - \Lambda_0(t)\right)$ are shown to converge weakly and jointly to Gaussian random variables with zero mean and appropriate limiting covariances. It is also shown that the Prentice's (1986) estimator, $\hat{\beta}_{PL}$, and the Self and Prentice (1988) estimator, $\hat{\beta}_{sp}$, are asymptotically equivalent provided an individual's contributions to $S^{(0)}$ and $S^{(1)}$ are asymptotically negligible. The variance estimate that is proposed by Prentice (1986) is somewhat different from the one proposed by Self and Prentice (1988) but it has been shown to converge to $\Sigma(\beta_0)^{-1} + \Sigma(\beta_0)^{-1}\Delta(\beta_0)\Sigma(\beta_0)^{-1}$.

The variance estimators proposed in these two papers (Prentice 1986, Self and Prentice 1988) are quite complicated. There have been many methods proposed to estimate the variance of the estimators from the pseudo-likelihood. Wacholder et al. (1989) proposed a bootstrap estimate of the variance of the covariate effect estimator. Their method echoes the original case-cohort sampling scheme by resampling cases and sub-cohort controls separately. This method is quite intensive computationally and would be quite time-consuming for large studies but it avoids the direct computation of the covariance among score components.

Barlow (1994) proposed a robust estimator of the variance based on the influence of an individual observation on the overall score. In the paper, the author assumed the standard Cox Proportional Hazard regression model for the relative risk as seen in equation (2.3). The conditional probability of failure at time t_j is given by

$$p_i(t_j) = \frac{Y_i(t_j)w_i(t_j)r_i(t_j)}{\sum_{k=1}^{n} Y_k(t_j)w_k(t_j)r_k(t_j)},$$

where the weight of the i-th subject at time t is given by

$$w_i(t) = \begin{cases} 1 & \text{if } dN_i(t) = 1, \\ \frac{m(t)}{\tilde{m}(t)} & \text{if } dN_i(t) = 0 \text{ and } i \in \text{Sc}, \\ 0 & \text{if } dN_i(t) = 0 \text{ and } i \notin \text{Sc} \end{cases}$$

where $m(t)$ is the number of disease-free individuals in the full-cohort who are at risk at time t and $\tilde{m}(t)$ is the number of disease-free individuals in the random sub-cohort who are at risk at time t; $r_i(t) = exp\{\beta_0' Z_i(t)\}$. One can easily note that Prentice (1986) used the weights as binary indicator, it being $= 1$ for the i-th individual at time t if $dN_i(t) = 1$ or $i \in$Sc and zero, otherwise. The Self & Prentice method used only the denominator summed over the subcohort members in the likelihood. The estimation of the unknown parameter follows directly from the log-likelihood of the conditional probability, i.e., $\sum_{i=1}^{n} \int_0^{\tau} log\,(p_i(t))\, dN_i(t) =$

$\sum_t \sum_{i=1}^n dN_i(t) log\left(p_i(t)\right)$. The robust variance estimator that was proposed in the paper, using the infinitesimal jackknife estimator, is given by $\hat{V}(\tilde{\beta}) = \frac{1}{n}\sum_{i=1}^n \hat{e}_i \hat{e}_i'$ where $\hat{e}_i = \tilde{\beta} - \tilde{\beta}_{-(i)} = I^{-1}(\tilde{\beta})\hat{c}_i(t_0)$, $c_i(t_0)$ is defined as the influence of an observation on the overall score for a particular individual i at time t_0 and $\hat{e}_i$ is the change in $\tilde{\beta}$ if the i-th observation is deleted. Further, $c_i(t_0)$ is given by

$$c_i(t_0) = \int_0^{t_0} Y_i(t)\left[dN_i(t) - \lambda_i(t)\right]\left(Z_i(t) - E_Z(t)\right) d\bar{N}(t),$$

where $E_Z(t) = \sum_{i=1}^n p_i(t) Z_i(t)$. $I(\tilde{\beta})$ is the information matrix generated by the pseudo-likelihood function. We can estimate $c_i(t_0)$ by $\hat{c}_i(t_0) = \int_0^{t_0} Y_i(t)\left[dN_i(t) - \hat{p}_i(t)\right]\left(Z_i(t) - \hat{E}_Z(t)\right)$ $\times d\bar{N}(t)$ and $I(\tilde{\beta}) = \sum_t \sum_i \hat{p}_i(t)\left[z_i(t) - \hat{E}(t)\right]\left[z_i(t) - \hat{E}(t)\right]'$.

Lin and Ying (1993) tackled the problem of missing covariate data under Cox Proportional Hazards regression model, of which the case-cohort design was a special case. They proposed an approximated partial-likelihood score function for the estimation of the regression parameters. A new variance-covariance estimator which is much easier to calculate than that Prentice (1986) and Self and Prentice (1988) was also proposed. The standard Cox PH regression model was assumed, as given in equation (2.1). Suppose that the data consist of iid random quintuplets $(X_i, \Delta_i, Z_i(.), H_{0i}(.), \mathbf{H}_i(.))$ where $Z_i(.) = \{Z_{i1}(.), Z_{i2}(.), \ldots, Z_{1p}(.)\}'$ may not be fully observed and $H_{0i}(.)$ is an indicator function and $\mathbf{H}_i(.)$ is a $p \times p$ matrix with the indicator functions $H_{1i}(.), H_{2i}(.), \ldots, H_{pi}(.)$ being the diagonal elements. Considering the original case-cohort design, we have $\mathbf{H}_i(.) = \mathbf{I}_p$ which is the $p \times p$ identity matrix and $H_{0i}(.)$ is 1 if the i-th subject belongs to the sub-cohort and zero, otherwise. The approximate partial likelihood score function can be written as

$$\tilde{U}_H(\beta) = \sum_{i=1}^n \Delta_i \mathbf{H}_i(X_i)\{Z_i(X_i) - E_H(\beta, X_i)\},$$

where $E_H(\beta, X_i) = \frac{S_H^{(1)}(\beta, X_i)}{S_H^{(0)}(\beta, X_i)}$; $S_H^{(d)}(\beta, t) = \frac{1}{n}\sum_{i=1}^n H_{0i}(t) Y_i(t) exp\{\beta' Z_i(t)\} Z_i(t)^{\otimes d}$. Let us define the root of the estimating equation, $\tilde{U}_H(\beta) = 0$, by $\tilde{\beta}_H$. Under certain regularity conditions, $n^{1/2}\left(\tilde{\beta}_H - \beta_0\right)$ can be shown to converge weakly to a Gaussian random variable with mean zero and variance given by $A^{-1}(\beta_0)V(\beta_0)A^{-1}(\beta_0)$, where $A_n(\beta) = -\frac{1}{n}\frac{\delta}{\delta\beta'}\tilde{U}_H(\beta)$, $lim_{n\to\infty}A_n(\beta) = A(\beta)$, $V(\beta) = E\left(W_1(\beta)^{\otimes 2}\right)$,

$$W_i(\beta) = \Delta_i \mathbf{H}_i(X_i)\{Z_i(X_i) - e_H(\beta, X_i)\} - \int_0^{X_i} \frac{\mathbf{h}(t)}{h_0(t)} H_{0i}(t)\{Z_i(t) - e_H(\beta, t)\} exp(\beta' Z_i(t)) d\Lambda_0(t),$$

where $e_H(\beta, t) = \frac{s_H^{(1)}(\beta,t)}{s_H^{(0)}(\beta,t)}$, $s_H^{(d)}(\beta, t) = E(S_H^{(d)}(\beta, t))$ for d = 0, 1; $\mathbf{h}(t) = E(\mathbf{H}_i(t))$, $h_k(t) = E(H_{ki}(t))$ for all $k = 0, 1, \dots, p$. Therefore, The covariance matrix can be approximated by $A_n^{-1}(\tilde{\beta}_H)\hat{V}(\tilde{\beta}_H)A_n^{-1}(\tilde{\beta}_H)$ where $\hat{V}(\beta) = \frac{1}{n}\sum_{i=1}^n \hat{W}_i(\beta)^{\otimes 2}$ and

$$\hat{W}_i(\beta) = \Delta_i \mathbf{H}_i(X_i)\{Z_i(X_i) - E_H(\beta, X_i)\}$$

$$-\frac{1}{n}\sum_{l=1}^n \Delta_l Y_i(X_l) H_{0i}(X_l)\mathbf{H}_l(X_l) exp\{\beta' Z_i(X_l)\} \times \frac{(Z_i(X_l) - E_H(\beta, X_l))}{S_H^{(0)}(\beta, X_l)}.$$

For the case-cohort design, the variance estimator $A_n^{-1}(\tilde{\beta}_H)\hat{V}(\tilde{\beta}_H)A_n^{-1}(\tilde{\beta}_H)$ is much easier to calculate, even when we have time-dependent covariates, than Prentice (1986) and Self and Prentice (1988). Another advantage of the proposed method is that incomplete covariate information on the covariates is allowed. Furthermore, the form of the proposed estimator remains unchanged under multiple sub-cohort augmentations. A natural estimator of the cumulative hazard function is proposed as

$$\tilde{\Lambda}(\tilde{\beta}_H, t) = \sum_{i=1}^n \frac{I(X_i \leqslant t) H_{0i}(X_i)\Delta_i}{nS_H^{(0)}(\tilde{\beta}_H, X_i)} = \int_0^t \frac{H_{0i}(u) dN_i(u)}{nS_H^{(0)}(\tilde{\beta}_H, u)}.$$

The process $n^{1/2}\left(\tilde{\Lambda}(\tilde{\beta}_H, t) - \Lambda_0(t)\right)$ converges weakly to a Gaussian process with mean zero and covariance function

$$\psi(t, s) = \int_0^{t\wedge s} \frac{d\Lambda_0(u)}{s^{(0)}(\beta_0, u)} + J'(t)A^{-1}(\beta_0)V(\beta_0)A^{-1}(\beta_0)'J(s) - J'(s)A^{-1}(\beta_0)G(t) - J'(t)A^{-1}(\beta_0)G(s), \tag{2.8}$$

where $J(t) = \int_0^t \frac{s^{(1)}(\beta_0,u)d\Lambda_0(u)}{s^{(0)}(\beta_0,u)}$ and $G(t) = E[\int_0^\infty \int_0^{t\wedge v} \frac{H_{0i}(u)exp\{\beta_0' Z_i(u)\}d\Lambda_0(u)}{s^{(0)}(\beta_0,u)}$ $\times \left(\mathbf{H}_i(v) - \frac{H_{0i}(v)}{h_0(v)}\mathbf{h}(v)\right)(Z_i(v) - e_H(v)) \times Y_i(v) exp\{\beta_0' Z_i(v)\}]$. $\psi(t, s)$ can be consistently estimated by simply replacing the parameters by the sample counterparts. The authors noted that this might not be the best choice if $H_{0i}(X_i) = 0$ for most of the non-zero Δ_i's. Hence, for the original case-cohort design, the authors recommend to use the formula proposed by Self and Prentice (1988) as seen in equation (2.7) rather than this.

Chen and Lo (1999) proposed a class of estimating equations for case-cohort design based on the partial likelihood score function, which lead to simple estimators that improve Prentice's pseudolikelihood estimator. The authors explored the usual Cox PH model (2.3), with r(.) being replaced by exp(.) and did not include any time-dependent covariates. They observed the triplets (X_i, Δ_i, Z_i) for the i-th individual. One can note that

$$\begin{aligned} E(Z \mid X = t, \Delta = 1) &= \frac{E\left(Ze^{\beta' Z}\mathbf{I}(X \geqslant t)\right)}{E\left(e^{\beta' Z}\mathbf{I}(X \geqslant t)\right)} \\ &= \frac{pE\left(Ze^{\beta' Z}\mathbf{I}(X \geqslant t) \mid \Delta = 1\right) + (1-p)E\left(Ze^{\beta' Z}\mathbf{I}(X \geqslant t) \mid \Delta = 0\right)}{pE\left(e^{\beta' Z}\mathbf{I}(X \geqslant t) \mid \Delta = 1\right) + (1-p)E\left(e^{\beta' Z}\mathbf{I}(X \geqslant t) \mid \Delta = 0\right)}, \end{aligned} \tag{2.9}$$

where p = P($\Delta = 1$). Let F^1 and F^0 be the conditional joint distributions of (X, Y) given $\Delta = 1$ and $\Delta = 0$ respectively. Further, suppose that $\mathcal{R}^1$ is the index set of random sample of k_1 cases and $\mathcal{R}^0$ is the index set of random sample of k_0 censored individuals. Then, one can estimate F^l, l = 0, 1 by the respective empirical counterpart, $\mathcal{R}^l$. Hence, replacing the population quantities by the empirical analogues, the authors proposed the following estimating function :

$$U(\beta) = \sum_{i \in \mathcal{R}^1} \int_0^\infty \left[Z_i - \frac{(\hat{p}/k_1)\sum_{j \in \mathcal{R}_t^1} Z_j e^{\beta' Z_i} + \{(1-\hat{p})/k_0\}\sum_{j \in \mathcal{R}_t^0} Z_j e^{\beta' Z_i}}{(\hat{p}/k_1)\sum_{j \in \mathcal{R}_t^1} e^{\beta' Z_i} + \{(1-\hat{p})/k_0\}\sum_{j \in \mathcal{R}_t^0} e^{\beta' Z_i}} \right] dN_i(t). \tag{2.10}$$

The authors considered several cases with different estimates of p depending on how much information one has about the data and considered the consistent estimator of the unknown parameters, β_0, and the corresponding covariance terms for the asymptotic distribution of $n^{1/2}\left(\hat{\beta}_{\text{CL}} - \beta_0\right)$ with mean zero. If the entire cohort is well-defined with n as the total cohort size, n_1 is the number of cases in the cohort and n_0 is the number of censored subjects and n_0^* is the number of censored subjects in the sub-cohort. Then the estimating function reduces to

$$U(\beta)=\sum_{i=1}^{n}\int_{0}^{\infty}\left[Z_i-\frac{\sum_{j\in\mathcal{R}_t^1}Z_je^{\beta'Z_i}+(n_0/n_0^*)\sum_{j\in\mathcal{R}_t^0}Z_je^{\beta'Z_i}}{\sum_{j\in\mathcal{R}_t^1}e^{\beta'Z_i}+(n_0/n_0^*)\sum_{j\in\mathcal{R}_t^0}e^{\beta'Z_i}}\right]dN_i(t)=0. \tag{2.11}$$

The solution to this equation, $\hat{\beta}_{cl}$, is consistent for β_0 and $n^{1/2}\left(\hat{\beta}_{cl}-\beta_0\right)$ converges to a Gaussian distribution with mean zero and variance given by

$$\sigma^2=\frac{1}{n}\Sigma_1^{-1}+\left(\frac{1}{n_0^*}-\frac{1}{n}\right)\Sigma_1^{-1}\left[\Sigma_0^{-1}-pV_1-\frac{1-p}{p}E_0^{\otimes 2}\right]\Sigma_1^{-1},$$

where $\Sigma_1 = var(W^*)$, $\Sigma_0 = var(W)$, $V_1 = var(W \mid \Delta = 1)$, $E_0 = E(W \mid \Delta = 0)$, $W^* = \{Z - m(Y)\}\Delta$, $W = \int_0^\infty\{Z - m(t)\}e^{\beta Z}\mathbf{I}(Y \geqslant t)d\Lambda_0(t)$, $m(t) = E(Z \mid Y = t, \Delta = 1)$. Kulich and Lin (2000) demonstrated the use of case-cohort data in estimating the regression parameter of an additive hazards regression model, where the conditional hazard function given a set of the covariates is the sum of an arbitrary baseline hazard and a function of the unknown regression parameter and the covariates. We have discussed this in details when discussing the literature for additive rates models.

Chen (2001) proposed a more efficient estimator by using local averages type of weights. The paper focuses on a unified approach for the parameter estimation of Cox PH model under different cohort sampling schemes, like nested case control, case-cohort and classical case-cohort methods. The sample reuse approach proposed via local averaging leads to more efficient estimators and this method is applicable to more complex sampling designs. The estimating equation they propose is the following :

$$U(\beta)=\sum_{i=1}^{n}\int_{0}^{\infty}[\mathbf{h}_i(t)-\frac{\sum_{j=1}^{n}W_j^*(t)\mathbf{h}_j(t)exp(\beta'Z_j(t))Y_j(t)}{\sum_{j=1}^{n}W_j^*(t)exp(\beta'Z_j(t))Y_j(t)}]W_i^*(t)dN_i(t)=0, \tag{2.12}$$

where $Y_j(t) = \mathbf{1}(T_j \wedge C_j \geq t)$ when one has only one event, $Z_j(t)$ is the vector of covariates corresponding to the j-th subject, $\mathbf{h}_i(.)$ are the sample analogues of a covariate-related process and $W_j^*(t)$ is the weight at time t. Chen & Lo considered the weights to be $W_j^*(t) = \Delta_j + (1 - \Delta_j)\delta_j n_0/n_0^*$, where Δ_j is the $\mathbf{1}$(event occurs for the j-th subject),

$\delta_j = \mathbf{1}$(j-th individual is sampled), n_0 & n_0^* is the number of censored individuals in the cohort and the subcohort, respectively. Chen (2001) proposed the following weight function, $W_j^*(t) = W_j^* = \delta_j / r_n(X_j, \delta_j)$, where $X_j = T_j \wedge C_j$ (when we have only one event) and

$$\begin{aligned} r_n(t,d) &= \frac{\sum_{l=1}^{n} \delta_l \Delta_l \mathbf{1}(x_l \in [t_{i-1}, t_i])}{\sum_{l=1}^{n} \delta_l \mathbf{1}(x_l \in [t_{i-1}, t_i])} \qquad \text{if d =1 and } t \in [t_{i-1}, t_i] \\ &= \frac{\sum_{l=1}^{n} (1-\delta_l) \Delta_l \mathbf{1}(x_l \in [s_{j-1}, s_j])}{\sum_{l=1}^{n} (1-\delta_l) \mathbf{1}(x_l \in [s_{j-1}, s_j])} \qquad \text{if d =0 and } t \in [s_{j-1}, s_j], \end{aligned}$$

where $0 = t_0 \leq t_1 \leq \ldots \leq t_{a_n} = \tau$ and $0 = s_0 \leq s_1 \leq \ldots \leq s_{b_n} = \tau$ are two partitions which satisfy certain conditions, resulting in more efficient estimation of the parameters. The author noted that the existing work on case-cohort sampling on survival analysis is highly dependent on the specific sampling design and the methods result in inefficient estimates of the regression parameters. Under the regularity conditions, the solution to equation (2.12), $\hat{\beta}_h$, is consistent for β_0 and the asymptotic distribution of $n^{1/2}\left(\hat{\beta}_h - \beta_0\right)$ is Gaussian with mean zero and variance $\Sigma_{h,Z}^{-1}\left(\Sigma_{h,h} + \Sigma_h^*\right)\Sigma_{h,Z}^{-1}$ with $\Sigma_{h,Z} = cov(M_{\tilde{h}}, M_{\tilde{Z}})$, $\Sigma_{h,h} = cov(M_{\tilde{h}}, M_{\tilde{h}})$, $\Sigma_h^* cov(W_{\tilde{h}}, M_{\tilde{h}})$, $M_h = \int_0^\tau \mathbf{h}(t) dM(t)$, $W_{\tilde{h}} = (1/\pi - 1)(M_{\tilde{h}} - M_{\tilde{h}}^o)$, $M_{\tilde{h}}^o = E(M_{\tilde{h}} \mid Y, \Delta)$, $\tilde{h}(t) = h(t) - E(h(t))$. Kulich and Lin (2004) developed a class of weighted estimating equations with time-dependent weight functions for stratified case-cohort designs. The authors in this paper considered the Cox (1972) regression model as described in (2.3) with exp(.) as the function r(.). They considered a cohort of n subjects who can be divided into K mutually exclusive strata based on a discrete random variable V. They assumed that V affects the failure time only through the covariates, i.e., T is independent of V given Z(.). The data observed is the following : $(X_{ki} = T_{ki} \wedge C_{ki}, \Delta_{ki}, Z_{ki}(t), 0 \leqslant t \leqslant \tau, V_{ki}, \xi_{ki})$ for the individuals in the sub-cohort and $(X_{ki}, \Delta_{ki}, Z_{ki}(X_{ki}))$ for all the cases outside the sub-cohort. The estimating equation is given by

$$U(\beta) = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^\tau \left[Z_{ki}(t) - E_C(\beta, t)\right] dN_{ki}(t) = 0, \tag{2.13}$$

where $E_C(\beta,t) = \frac{S_C^{(1)}(\beta,t)}{S_C^{(0)}(\beta,t)}$, $S_C^{(d)}(\beta,t) = \frac{\sum_{k=1}^{K} \sum_{i=1}^{n_k} \varrho_{ki}(t) Z_{ki}(t) Y_{ki}(t) e^{\beta' Z_{ki}(t)}}{\sum_{k=1}^{K} \sum_{i=1}^{n_k} \varrho_{ki}(t) Y_{ki}(t) e^{\beta' Z_{ki}(t)}}$, $\varrho_{ik}(t) = \Delta_{ki} + (1 - \Delta_{ki})\xi_{ki}\hat{\alpha}_k(t)$, which is based on the weights proposed by Kalbfleisch and Lawless (1988). ξ_i

is the indicator that the individual is included in the sub-cohort and $\hat{\alpha}_k(.)$ is a function of the controls (because of the separation between the cases and the controls). The authors proposed a doubly weighted estimator

$$\hat{\alpha}_k(t) = \left\{ \sum_{i=1}^{n_k} (1 - \Delta_{ki}) A_{ki}(t) \right\}^{-1} \left\{ \sum_{i=1}^{n_k} \xi_{ki} (1 - \Delta_{ki}) A_{ki}(t) \right\},$$

where $A_{ki}(t)$ is a diagonal matrix with m potentially different random processes on the diagonal. Redefine the at-risk covariate average as $E_{DW}(\beta, t) = \{S_{DW}^{(0)}(\beta, t)\}^{-1} S_{DW}^{(1)}(\beta, t)$ where $S_{DW}^{(1)}(\beta, t)$ and $S_{DW}^{(0)}(\beta, t)$ are the matrix counterparts of $S_C^{(1)}(\beta, t)$ and $S_C^{(0)}(\beta, t)$, respectively. Based on these, the authors showed that, under certain regularity conditions, $\hat{\beta}_{DW}$ is a consistent estimator of β_0 and $n^{1/2}\left(\hat{\beta}_{DW} - \beta_0\right)$ converges to a Gaussian distribution with mean 0 and variance $\mathbf{I}_F^{-1} + \mathbf{I}_F^{-1} \Sigma_{DW} \mathbf{I}_F^{-1}$, where $\mathbf{I}_F = \int_0^\tau \left[\frac{s^{(2)}(\beta,t)}{s^{(0)}(\beta,t)} - e(\beta, t)^{\otimes 2} \right] s^{(0)}(\beta, t) d\Lambda_0(t)$, $s^{(d)}(\beta, t) = E(Z_i(t) Y_i(t) e^{\beta' Z_i(t)})$, d = 0,1,2; $e(\beta, t) = \frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}$, $\Sigma_{DW} = \sum_{k=1}^K q_k \frac{1-\alpha_k}{\alpha_k}$ $E_k \left\{ (1 - \Delta_i) \int_0^\tau \left(\mathbf{R}_{ki}(t) - \mu_k^{-1}(t) A_{ki}(t) \psi_k(t) \right) d\Lambda_0(t) \right\}^{\otimes 2}$ with $q_k = P(V = k)$, $\alpha_k = P(\xi = 1 \mid V = k)$, $\mathbf{R}_{ki}(t) = (Z_{ki}(t) - e(\beta, t)) e^{\beta' Z_{ki}(t)} Y_{ki}(t)$, $\mu_k = E\left[(1 - \Delta_{ki}) A_{ki}(t)\right]$ and $\psi_k(t) = E\left[(1 - \Delta_{ki}) \mathbf{R}_{ki}(t)\right]$.

Nan (2004) developed semi-parametric efficient estimators, obtained by solving the the efficient score equation with the help of the Newton-Raphson algorithm. Lu and Tsiatis (2006) proposed a general class of semi-parametric transformation model. They considered weighted estimating equations for simultaneous estimation of the regression parameters and the transformation function. The semi-parametric linear transformation model is specified by

$$H(T) = -\beta_0' Z + \epsilon,$$

where T is the survival time, H is an unknown monotone increasing function, β_0 is the p-dimensional unknown regression parameters and ϵ is the error term which has a continuous distribution, independent of the censoring process and covariates. The corresponding estimating equations are given by

$$\sum_{i=1}^{n} \pi_i \left[dN_i(t) - Y_i(t)d\Lambda\{H(t) + \beta' Z_i\}\right] = 0 \quad (t \geqslant 0), \quad H(0) = -\infty,$$

$$\sum_{i=1}^{n} \int_0^{\infty} Z_i \pi_i \left[dN_i(t) - Y_i(t)d\Lambda\{H(t) + \beta' Z_i\}\right] = 0.$$

The authors showed that the resulting regression estimators are asymptotically normal, with a closed form variance-covariance matrix that can be consistently estimated by the usual plug-in estimators. Lu and Shih (2006) extended the case-cohort design to clustered data, where each clustered comprises of correlated individuals, whose structure of dependence is kept unspecified. The estimating equation is defined as

$$U(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[Z_{ij}(X_{ij}) - \bar{E}(\beta, X_{ij})\right] \delta_{ij},$$

where $\bar{E}(\beta, t) = \frac{\bar{S}^1(\beta,t)}{\bar{S}^0(\beta,t)}$, $\bar{S}^d(\beta, t) = \sum_{i=1}^{n} \sum_{j}^{m} \Delta_{ij} Y_{ij}(t) e^{\beta' Z_{ij}(t)} Z_{ij}(t)^{\otimes d}$, Δ_{ij} is an individual indicator that equals 1 if individual (i, j) is selected for the subcohort, and 0, otherwise and δ_{ij} is the indicator of whether the j-th individual in the i-th cluster observes the event or is censored. Kang and Cai (2009a) considered marginal hazards model for case-cohort data with multiple disease outcomes. Time to different events are modeled simultaneously in order to compare the effect of a risk factor on different types of diseases. They have proposed valid statistical methods that take the correlations among the outcomes from the same subject into account. They have considered the multiplicative intensity process model, which is specified by

$$\lambda_{ik}(t \mid Z_{ik}(t)) = Y_{ik}(t) \lambda_{0k}(t) e^{\beta_0' Z_{ik}(t)},$$

where $\lambda_{0k}(t)$ is an unspecified baseline hazard function for disease outcome k and β_0 is the vector of unknown parameters. Defining the estimating equation, similar to (2.13), it is shown that the parameter estimators are consistent and asymptotically normal and the corresponding variance-covariance term can be easily replaced by the sample counterparts. Zhang et al.

(2011) developed estimating equation for clustered failure time data assuming a marginal hazards model, with a common baseline hazard and common regression coefficient across clusters. One of the key advantages of the case-cohort design is its ability to use the same sub-cohort for several diseases or for sub-types of diseases (Prentice 1986, Wacholder et al. 1989, Langholz and Thomas 1990, Wacholder 1991). The availability of the case-cohort sub-cohort may be useful for study monitoring and can be considered to be a natural comparison group at all disease occurrence times for all of the different diseases. Chen and Chen (2014) extended the case-cohort design to recurrent events with specific clustering feature using a modified Cox-type self- exciting intensity model. Under this model, conditional on the covariates and history of events upto time t, the intensity function of N(.) (simple point process) is given by

$$\lambda(t) = \mu(t) exp\{\Psi(t,\theta)\},$$

where $\Psi(t,\theta) = \beta' Z(t) + \phi(t)$, $\mu(t)$ is the unspecified baseline intensity, $Z(t)$ is the time-varying p-dimensional vector of covariate, β is the vector of regression coefficients, $\phi(t)$ is a self-exciting term depending on past events of the process and θ is the combined finite dimensional parameters to be specified. The pseudo-likelihood score equation for θ is given by

$$\log L(\theta) = \sum_{i=1}^{n} \int_0^{C_i} \left[\psi_i(t,\theta) - \frac{\sum_{j=1}^{n} \epsilon_j \psi_j(t,\theta) Y_j(t) exp\{\Psi_j(t)\}}{\sum_{j=1}^{n} \epsilon_j Y_j(t) exp\{\Psi_j(t)\}} \right] dN_i(t) = 0, \qquad (2.14)$$

where $\psi(t,\theta) = \frac{\partial}{\partial\theta}\Psi(t,\theta)$. The authors went onto show the asymptotic properties of the solution to the above equation, under certain regularity conditions.

2.3 Recurrent Events Data

In many longitudinal studies and for several medical conditions, subjects experience repeated or recurrent events. Some examples include the occurrence of new tumors in patients with superficial bladder cancer (Byar, 1980), recurrent seizures in epileptic patients (Albert 1991), rejection episodes in patients receiving kidney transplants (Cole et al. 1994), repeated infections in HIV-patients (Li and Lagakos 1997), repeated cardiovascular events in patients (Cui et al. 2008). In this section, we look at the different ways one can model the recurrent event data and the assumptions that are attached to it. In the following subsections, we will summarize the marginal models (more specifically, intensity models and rate & mean models) which do not assume anything about the nature of the dependence among the recurrent events and the frailty model which explicitly specifies the relationship among the repeated events.

2.3.1 Marginal Models using Multiplicative Models

Prentice et al. (1981) considered two stratified proportional hazards type of regression models for modeling the intensity function in terms of the covariates and the failure time history. The main diference between the two models is that one of them specifies the baseline intensity function as a function of time from the beginning of study, while the other specifies it as a function of time from the subject's immediately preceding failure. Let $Z(t) = \{z(u) \mid u \leqslant t\}$ be the covariate process upto time t and $N(t) = \{n(u) \mid u \leqslant t\}$, where n(u) is the number of failures on a study subject prior to time t. The counting process N(t) is equivalent to random failure time, $T_1, T_2, \ldots, T_{n(t)}$ in $[0, t)$. The authors define the hazard or the intensity function at time t as the instantaneous rate of failure at time t given the covariate and counting process.

$$\lambda\{t \mid N(t), Z(t)\} = \lim_{\delta t \to 0} \frac{P(t \leqslant T_{n(t)+1} \leqslant t + \delta t \mid N(t), Z(t))}{\delta t}$$

$$\Rightarrow \lambda\{t \mid N(t), Z(t)\} = \lambda_{0s}(t) e^{\beta_{0s} z(t)}, \quad \text{time from the beginning of study} \tag{2.15}$$

$$\text{or} \quad \lambda\{t \mid N(t), Z(t)\} = \lambda_{0s}(t - t_{n(t)}) e^{\beta_{0s} z(t)}, \quad \text{time from the immediately preceeding failure,} \tag{2.16}$$

where $\lambda_{0s}(t)$ $(s = 1, 2, \ldots)$ are arbitrary baseline intensity functions and the stratification variable, $s = s(N(t), Z(t))$ may change as a function of time for a particular subject and β_{0s} is the stratum specific regression coefficient. The authors propose two different likelihood functions for the two quite different models considered. The likelihood function for the model where they studied the time from the beginning of the study is given by

$$L(\beta_0) = \prod_{s \geqslant 1} \prod_{i=1}^{d_s} \frac{exp[\beta_0' z_{si}(t_{si})]}{\sum_{l \in \mathcal{R}(t_{si},s)} exp[\beta_0' z_l(t_{si})]},$$

where $\mathcal{R}(t, s)$ is the risk set at time t for the s-th stratum. For the time from the last observed failure, we have

$$L(\beta_0) = \prod_{s \geqslant 1} \prod_{i=1}^{d_s} \frac{exp[\beta_0' z_{si}(t_{si})]}{\sum_{l \in \tilde{\mathcal{R}}(u_{si},s)} exp[\beta_0' z_{t_l+u+si}(t_{si})]},$$

where $\tilde{\mathcal{R}}(t, s)$ is the risk set at gap-time t stratum-s. t_l is the last failure time on subject l prior to entry into stratum s. The authors note that ordinary asymptotic likelihood methods can be applied to both the likelihood functions (Cox 1975) though consideration should be given to the size and to the number of failures in each stratum. Andersen and Gill (1982) explored the Cox PH regression model (as in equation (2.1)). The authors extended the Cox PH model for a single failure time data where the effect of the covariate is proportional on the hazard, to the multivariate counting process, where the effect of the covariate process is proportional on the intensity function. They showed that $M_i(t) = N_i(t) - \int_0^t d\Lambda_i(t)$ are local square-integrable martingales on the time interval [0, 1]. The solution to the estimating equation $U(\beta) = 0$ is denoted by $\hat{\beta}$, where $U(\beta)$ is defined as in (2.2), with $Y_i(t)$ defined as an indicator function showing whether the individual is still under observation or not. They showed that $\hat{\beta}$ is consistent for β_0 and $n^{-1/2}U(\beta_0)$ converges to a Gaussian distribution with mean zero and variance, $\Sigma(\beta_0)$, where $\Sigma(\beta) = \int_0^1 \left(\frac{s^{(2)}(\beta,t)}{s^{(0)}(\beta,t)} - \frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}^{\otimes 2} \right) s^{(0)}(\beta, t)\lambda_0(t)dt.$

Wei et al. (1989) noted that many survival studies record the times to two or more failures per subject, where the failure types might be completely different or they may be repetitions of the same type. The authors proposed to model the marginal distribution of each failure time variable with a Cox-type PH regression model. In this approach, no structure is imposed on the dependence of the different failure times for each subject. The hazard function for the k-th type of failure time of the i-th subject is given by

$$\lambda_k(t \mid Z_{ki}(t)) = \lambda_{k0}(t)e^{\beta_{0k} Z_{ki}(t)}, \quad \text{k} = 1, 2, \ldots, K, \quad \text{i} = 1, 2, \ldots, n. \tag{2.17}$$

The k-th failure-specific partial likelihood (Cox, 1975) is defined as

$$L_k(\beta_0) = \prod_{i=1}^{n} \frac{exp[\beta_0' Z_{ki}(X_{ki})]}{\sum_{l \in \mathcal{R}(X_{ki})} exp[\beta_0' Z_{li}(X_{ki})]},$$

where $\mathcal{R}(t) = \{i : X_{ki} \geqslant t\}$ is the set of individuals who are at risk of observing the k-th type of failure just prior to time t. Under certain regularity conditions, the solution to the partial likelihood equation, $\frac{\partial L_k(\beta)}{\partial \beta} = 0$, given by $\hat{\beta}_k$, will be consistent for the unknown parameter, β_{k0}. $n^{1/2}(\hat{\beta} - \beta_0) = n^{1/2}(\hat{\beta}_1 - \beta_{01}, \hat{\beta}_2 - \beta_{02}, \ldots, \hat{\beta}_K - \beta_{0K})$ converges asymptotically to a Normal distribution with zero-mean and variance Q with elements equal to $D_{ij}(\hat{\beta}_i, \hat{\beta}_j)$, i, j $= 1, 2, \ldots, K$ and $D_{ij}(\hat{\beta}_i, \hat{\beta}_j)$ can be consistently estimated by $\hat{A}_i(\hat{\beta}_i)^{-1}\hat{B}(\hat{\beta}_i, \hat{\beta}_j)\hat{A}_j(\hat{\beta}_j)^{-1}$ with $\hat{A}_i(\beta_i)^{-1} = \frac{1}{n}\sum_k \Delta_{ik}\left[\frac{S^{(2)}(\beta_i, X_{ik})}{S^{(0)}(\beta_i, X_{ik})} - (\frac{S^{(1)}(\beta_i, X_{ik})}{S^{(0)}(\beta_i, X_{ik})})^{\otimes 2}\right]$, $\hat{B}(\hat{\beta}_i, \hat{\beta}_j) = \frac{1}{n}\sum_k W_{ik}(\hat{\beta}_i)W_{jk}(\hat{\beta}_j)'$, $W_{ik}(\beta_k) = \left[\Delta_{ik} - \sum_{m=1}^{n} \frac{\Delta_{im} Y_{ik}(X_{im}) e^{\beta_k' Z_{ik}(X_{im})}}{S^{(0)}(\beta_i, X_{ik})}\right] \times \left[Z_{ik}(X_{im}) - \frac{S^{(1)}(\beta_i, X_{ik})}{S^{(0)}(\beta_i, X_{ik})}\right]$,
$S^{(d)}(\beta, t) = \frac{1}{n}\sum_{i=1}^{n} Z_i(t) Y_i(t) e^{\beta' Z_i(t)}$.

Pepe and Cai (1993) looked at different methods to display and analyze multiple failure time data. The authors looked at the rate function. They defined the rate functions as $r^F(t) = \lim_{\Delta \to 0} \frac{1}{\Delta} P[\text{event occurred in } (t, t+\Delta) \mid \text{at risk and no event observed at } t]$ (rate of first infection) and $r^R(t) =$
$\lim_{\Delta \to 0} \frac{1}{\Delta} P[\text{event occurred in } (t, t+\Delta) \mid \text{at risk and event previously observed at } t]$ (rate of recurrent infection). Suppose that r^F and r^R are functions of time that involve certain parameters, α and γ as r^F_α and r^R_γ respectively. The likelihood for α is given by

$$L(\alpha) = \prod_{t_i \in D} r^F_\alpha(t_i) exp\left\{-\int_0^{t_i} r^F_\alpha(u)du\right\} \times \left[\prod_{t_j \in D} r^F_\alpha(t_j) exp\{-\int_0^{t_j} r^F_\alpha(u)du\}\right],$$

where D is the set of times when the event occurred for the first time; O is the set of observation times for subjects censored or lost to follow-up due to competing risk events. The corresponding score function is given by

$$S^F(\alpha) = \int f_\alpha(t)\left\{dN^F(t) - Y^F(t) r^F_\alpha(t)dt\right\}, \tag{2.18}$$

where $f_\alpha(t) = \frac{\partial log r^F_\alpha(t)}{\partial \alpha}$, $N^F(t)$ = the number of individuals who had the event observed by time t, and $Y^F(t)$ = the number of individuals under observation and who had not observed any prior event at t. By definition, $E\left(dN^F(t)/Y^F(t) \mid Y^F(t)\right) = r^F_\alpha(t)dt$. The authors note that under mild regularity conditions, the solution to the estimating equation (2.18) should be a consistent estimator of α, even when $f_\alpha(t)$ is somewhat different from $\frac{\partial log r^F_\alpha(t)}{\partial \alpha}$. They further noted that setting $f_\alpha(t)$ at this quantity would yield the most efficient estimator. The authors also observed that likelihood-based estimation of r^R_γ would require specifying a model for the joint distribution of recurrence times within the same individual, but one can get consistent estimates of γ from an estimating equation like (2.18).

$$S^R(\alpha) = \sum_{i=1}^{n} S^R_i(\alpha) = \sum_{i=1}^{n} \int I_\gamma(t)\left\{dN^R_i(t) - Y^R_i(t) r^R_\alpha(t)dt\right\} = 0, \qquad (2.19)$$

where $N^R_i(t)$ is the number of events observed by time t for the i-th individual with previously observed events, $Y^R_i(t)$ is the indicator that the i-th individual, with previously observed events, is at-risk of an event at t, and $I_\gamma(t)$ is some bounded deterministic vector-valued function whose dimension is the same as the dimension of γ. The authors showed that the solution to the estimating equation (2.19), $\hat{\gamma}$, is consistent estimator of γ and $n^{-1/2}S^R(t) \xrightarrow{D} N(O, V(t))$, where $V(t)$ is the variance- covariance matrix for $S^R_i(t)$ and one can get the asymptotic distribution of $\hat{\gamma}$ from Taylor series expansion.

Lawless and Nadeau (1995) presented non-parametric methods and regression models to estimate the cumulative mean function (CMF), defined as $M(t) = E(N_i(t))$, where $N_i(t)$ is the number of events occurring in time [0, t]. The main objective is to estimate $M(t)$ based on the observed times times, which are defined as $t_{i1} \leqslant t_{i2} \leqslant \ldots \leqslant t_{ir_i}$ for the i-th individual over the interval [0, τ_i] (i = 1,2,. . ., k) and the individuals are independent. For simplicity, the authors presented the results in a discrete-time framework. Define $\delta_i(t) = 1$, if $t \leqslant \tau_i$, 0 o.w., to denote whether an individual observed an event. Further, let $n_i(t) \geqslant 0$ denote the number of events that occur at time 't' for the i-th individual so that $m(t) = E\{n_i(t)\}$ and $M(t) = \sum_{s=0}^{t} m(s)$. The maximum likelihood estimator the authors proposed is given by $\hat{M}(t) = \sum_{s=0}^{t} \frac{n_.(s)}{\delta_.(s)}, 0 \leqslant t \leqslant \tau = \max_i(\tau_i)$, where $n_.(t) = \sum_{i=1}^{k} \delta_i(t) n_i(t)$ and $\delta_.(t) = \sum_{i=1}^{k} \delta_i(t)$. The authors proposed the variance of $\hat{M}(t)$ as $var(\hat{M}(t)) = \sum_{i=1}^{k} \sum_{s=0}^{t} \sum_{u=0}^{t} \frac{\delta_i(s)\delta_i(u)}{\delta_.(s)\delta_.(u)} cov(n_i(s), n_i(u))$. The authors further extended the idea to develop similar approach for a flexible family of regression models. The following regression model

$$E(n_i(t)) = m_i(t) = m_0(t)P_i(t)g(z_i(t), \beta_0)$$

was considered, where β_0 is a p-dimensional vector of regression parameters, g is a positive valued function, $m_0(t) \geqslant 0$ is the baseline mean function and $P_i(t)$ is some known function. Noting that $m_0(t) = \frac{n.(t)}{R(t,\beta_0)}$ and $R(t, \beta_0) = \sum_{i=1}^{k} \delta_i(t) g_i(t)$, the estimating equation for β is given by $\sum_{j=1}^{k} \sum_{l \in D_j} \left\{ \frac{\partial log g_l(t_j)}{\partial \beta} - \frac{\partial R(t_j,\beta)}{\partial \beta} \right\} = 0$, where D_j represents the set of individuals with events at t_j, including repetitions for an individual at that time. Under mild conditions, the solution to this estimating equation, $\hat{\beta}$, is consistent for β_0. Defining

$$W_i(\beta, s) = \frac{\partial log g_i(s)}{\partial \beta} - \frac{\partial log R(s, \beta)}{\partial \beta},$$

$$\hat{B}_{1i}(t) = \sum_{s=0}^{t} \delta_i(s) W_i(\hat{\beta}, s)[n_i(s) - \hat{g}_i(s)\hat{m}_0(s)],$$

where $\hat{g}_i(s) = P_i(s)g(x_i, \hat{\beta})$, $\hat{B}_1 = \frac{1}{k} \sum_{i=1}^{k} \hat{B}_{1i}\hat{B}'_{1i}$ and $\hat{A}_1 = \frac{1}{k} \sum_{i=1}^{k} \sum_{s=0}^{\tau} \delta_i(s)\hat{m}_0(s) \frac{\partial \hat{g}_i(s)}{\partial \hat{\beta}}$ $\times W_i(\hat{\beta}, s)'$, we have $k^{1/2}\left(\hat{\beta} - \beta_0\right)$ converges asymptotically to a Gaussian distribution with mean 0 and a sandwich variance term $v\hat{a}r(\hat{\beta}) = \hat{A}_1^{-1}\hat{B}_1(\hat{A}_1^{-1})'$. Lawless (1995) discussed different ways of introducing covariates in the analysis of recurrent events and the distinction between rate (and mean) function and intensity functions event process characterizations. The author defines the rate of occurrence function is $r(t) = \frac{\partial E(N(t))}{\partial t}$ and the mean function is given by

$$E\left[N(t)\right] = \int_0^t r(u)du \tag{2.20}$$

where r(t) can be the intensity function (but usually they are quite different). Covariates can be introduced into the rate or mean functions as $r(t; z) = \rho(t)\phi(z)$. Note that this type of marginal model is different from Wei et al. (1989), where they have modeled the distribution of time to the j-th event (j = 1,2,...) and the hazard function is given by $h_j(t; z) = h_{oj}(t)e^{\beta z}$.

Renewal type of models based on intervals between successive events would be characterized by

$$\lambda(t; H_{it}, z_i) = h_j(t - t_{ij}) exp(\beta_j z_i),$$

where $N_i(t-) = j$ (Gail et al. 1980) specifies that for each subject the times to the first event and between successive events are independent random variables and they do not necessarily have the same distribution; $H_t = \{N(s) \mid s < t\}$ is the history of the process up to time t; t_{ij} is the j-th observed event time for the i-th individual and β_j are the regression parameters. The authors used Lawless and Nadeau's (1995) approach to estimate β, yielding $R_0(t) = \int_0^t \frac{dN(u)}{\sum_{i=1}^n \delta_i(u)\phi_i(z_i,\beta)}$ and $N(u)$ is the total number of events observed at u. The authors further proposed the joint distribution of $n^{1/2}\left(\hat{\beta} - \beta, \hat{R}_0(t) - R_0(t)\right)$. They noted that the marginal analyses for means are quite easy to interpret and may be made robust to other assumptions about the recurrent event process, provided that the observation period $[0, \tau_i]$ for the i-th individual (for all i) are independent of the corresponding event process.

Cook et al. (1996) investigated robust non-parametric tests, in the sense, that these methods do not rely on distributional assumptions of the event-generating process, for recurrent event data. They have further explored a family of generalized pseudo-score statistics (Lawless and Nadeau 1995) in which weight functions may be chosen to generate tests sensitive to various types of departure from the null hypothesis (which states that the mean functions for the treatment and control groups are identical). Denote $N_i(t)$ as the number of events experienced by the i-th subject and t is the time on the study; $dN_i(t) = \lim_{\delta t \to 0}[N_i(t) - N_i(t - \delta t)]$. $Y_i(t)$ be an indicator variable which takes value 1 if subject i is observed to be at risk at time t, and is zero, otherwise and $\Lambda_i(t) = E\{N_i(t)\}$. The authors note that Nelson (1988) and Lawless and Nadeau (1995) have stated the well known non-parametric Nelson-Aalen estimator,

$$\hat{\Lambda}(t) = \sum_{i=1}^{n} \int_0^t \frac{Y_i(u) dN_i(u)}{\sum_{j=1}^n Y_j(u)} = \int_0^t \frac{dN_.(u)}{Y_.(u)}.$$

Under mild conditions, the results in Lawless & Nadeau (1995) can be extended to get

$$\sigma(t,s) = n\sum_{i=1}^{n}\left[\int_0^t\int_0^s \frac{Y_i(u)Y_i(v)}{Y_.(u)Y_.(v)} cov(dN_i(u), dN_i(v))\right],$$

where $\sigma(t,s) = ncov(\hat{\Lambda}(t), \hat{\Lambda}(s))$, $Y_.(u) > 0, 0 < u \leqslant \max(t,s)$. Assuming that $Y_.(u) \to \infty$ as $n \to \infty$ for all $u \in (0, \max(t,s)]$ and $\sigma(t,s)$ converges to a positive definite limit, $\hat{\Lambda}(t)$ is consistent for $\Lambda(t)$ and $\sigma(t,s)$ would be estimated by

$$\hat{\sigma}(t,s) = n\sum_{i=1}^{n}\left[\int_0^t\int_0^s \frac{Y_i(u)Y_i(v)}{Y_.(u)Y_.(v)}(dN_i(u) - d\hat{\Lambda}(u))(dN_i(v) - d\hat{\Lambda}(v))\right].$$

The authors showed that under some additional conditions, $n^{1/2}\{\hat{\Lambda}(t) - \Lambda(t)\}, t > 0$ converges to a mean zero Gaussian process with covariance function equal to $\sigma(t,s)$ over some interval $0 < t, s, \leqslant \tau$. The authors also considered stratified designs and the forms are natural extensions of the above. Chang and Wang (1999) focused on conditional regression analysis, given the ordinal nature of recurrent events per subject. A semi-parametric hazards model, including the structural and episode specific parameters considered as stratification variable, has been proposed for modeling the data. Spiekerman and Lin (1999) proposed a Cox-type regression model to model the marginal distribution of multivariate failure time data. They used different baseline hazard functions for different failure types and proved that the maximum quasi-partial-likelihood estimator of the regression parameters are consistent and asymptotically normal. Lin et al. (1999) proposed a simple non-parametric estimator for the multivariate distribution function of the gap times between successive events of the same type, with each individual experiencing multiple repetitions of the same, when the follow-up time is subject to right censoring.

Lin et al. (2000) provided the rigorous justification for the methods outlined in Pepe and Cai (1993) and Lawless and Nadeau (1995) papers. The authors note that the main difference between the Andersen and Gill (1982) paper and the above mentioned papers is the following condition which implies that all the influence of prior events on future recurrence is through the time varying covariates at time, t.

$$E(dN^*(t) \mid \mathcal{F}_{t-}) = E(dN^*(t) \mid Z(t)),$$

where $N^*(t)$ is the number of events that occurred over the interval [0, t], Z(t) is the covariate process and $\mathcal{F}_t$ is the σ-field generated by $\{N(s), Z(s) \mid 0 \leqslant s \leqslant t\}$. The authors recommended to relax this assumption because the dependence of the recurrent events may not be adequately captured by the time-varying covariates and there was no method at the time to verify the assumption. Defining $E(dN^*(t) \mid Z(t)) = d\mu_z(t)$, the model that the authors worked with is

$$d\mu_z(t) = exp\{\beta_0' Z(t)\} d\mu_0(t), \tag{2.21}$$

where $\mu_0(t)$ is a known continuous function. $N^*(t)$ would not be fully observed as the subjects are followed for a limited amount of time. Let C denote the censoring variable and it is assumed that the censoring mechanism is independent in the sense that $E(dN^*(t) \mid Z(t), C \geqslant t) = E(dN^*(t) \mid Z(t))$. $N(t) = N^*(t \wedge C)$ and $Y(t) = \mathbf{I}(C \geqslant t)$. Hence for each individual, the observable data is $(N_i(.), Y_i(.), Z_i(.))$. The partial likelihood score function is given by

$$U(\beta, t) = \sum_{i=1}^{n} \int_0^t \{Z_i(u) - E(\beta, t)\} dN_i(u),$$

where $E(\beta, t) = \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}$, $S^{(d)}(\beta, t) = \frac{1}{n}\sum_{i=1}^{n} Y_i(t) Z_i(t) exp\{\beta' Z_i(t)\}$, d = 0,1,2. Note that, the estimating equation would be given by $U(\beta, \tau) = 0$. It is shown by using empirical processes theory that the solution to the estimating equation, $\hat{\beta}$, converges almost surely to β_0 and $n^{1/2} U(\beta_0, t)$ $(0 \leqslant t \leqslant \tau)$ converges weakly to a zero-mean Gaussian process with covariance function

$$\Sigma(t, s) = E\left[\int_0^t \{Z_i(u) - e(\beta_0, u)\} dM_i(u) \int_0^s \{Z_i(v) - e(\beta_0, v)\}' dM_i(v)\right],$$

where $dM_i(t) = \mathbf{I}(C_i \geqslant t)\left[dN_i^*(t) - exp\{\beta_0' Z_i(t)\} d\mu_0(t)\right]$ and $e(\beta, t) = \frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}$, $s^{(d)}(\beta, t) = E(S^{(d)}(\beta, t))$. Further, $n^{1/2}(\hat{\beta} - \beta_0)$ converges weakly to Gaussian distribution with mean zero and variance $\Gamma = A^{-1}\Sigma A^{-1}$, where $A = E\left\{\int_0^\tau (Z_i(t) - e(\beta, t))^{\otimes 2} Y_i(t) exp\{\beta_0' Z_i(t)\} d\mu_0(t)\right\}$, which is also positive definite. The Aalen-Breslow type estimator is used to estimate $\mu_0(t)$ and is defined as $\hat{\mu}_0(t) = \int_0^t \frac{d\bar{N}(u)}{nS^{(0)}(\hat{\beta}, u)}, t \in [0, \tau]$ with $\bar{N}(t) = \sum_{i=1}^n N_i(t)$. The covariance matrix would then be estimated by $\hat{\Gamma} = \hat{A}^{-1}\hat{\Sigma}\hat{A}^{-1}$, $\hat{A} = -\frac{1}{n}\frac{\partial U(\hat{\beta},\tau)}{\partial \beta'} = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{Z_i(t) - E(\hat{\beta}, t)\} d\hat{\mu}_0(t)$, $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{Z_i(u) - E(\hat{\beta}, u)\} d\hat{M}_i(u) \int_0^\tau \{Z_i(v) - E(\hat{\beta}, v)\}' d\hat{M}_i(v)$ and $\hat{M}_i(t)$ $= N_i(t) - \int_0^t Y_i(u) exp(\hat{\beta}' Z_i(u)) d\mu_0(u)$. The authors proceeded further to incorporate a random weight function, $\hat{Q}(\beta, t)$, which is non-negative, bounded and monotone in t and converges almost surely to a continuous deterministic function, q(t), $t \in [0, \tau]$. The terms are similar to the ones shown with a q(.) term in them. Further, it is shown that $n^{1/2}\{\hat{\mu}_0(t) - \mu_0(t)\}$ is asymptotically equivalent to $n^{-1/2}\sum \Psi_i(t)$ where

$$\Psi_i(s) = \int_0^s \frac{dM_i(t)}{s^{(0)}(\beta_0, t)} - h'(\beta_0, t)A^{-1}\int_0^\tau \{Z_i(t) - e(\beta_0, t)\} dM_i(t),$$

$h(\beta_0, t) = \int_0^t e(\beta_0, u) d\mu_0(u)$, and the covariance function is given by $\xi(t, s) = E(\Psi(t)\Psi(s))$. Peña et al. (2001) derived Nelson-Aalen and Kaplan-Meier type of estimators for the non-parametric estimator of the cumulative distribution function of the time to occurrence of recurrent events, in presence of censoring.

Cai and Schaubel (2004) considered modeling the rate function for the recurrent events, when there are different types of events. The rate function corresponding to the k-th event type is $E(dN_{ik}^*(t) \mid Z_{ik}) = g(\beta_0' Z_{ik}) d\mu_{0k}(t)$, where $N_{ik}^*(t) = \int_0^t dN_{ik}^*(u)$ is the number of events of type k at time t for the i-th subject. Further, let C_{ik} and $Y_{ik}(t) = \mathbf{I}(C_{ik} \geqslant t)$ be indicator functions denoting the event-type-specific censoring time and at-risk function, respectively. Hence, the observed process is denoted by $N_{ik}(t) = \int_0^t Y_{ik}(u) dN_{ik}^*(u)$. They proposed a slightly different estimating equation, analogous to generalized estimating equation (Liang and Zeger 1986).

$$U(\beta) = \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau Z_{ik}(t) \frac{g^{(1)}(\beta' Z_{ik}(t)}{g(\beta' Z_{ik}(t)} \left[dN_{ik}(t) - Y_{ik}(t) g(\beta' Z_{ik}(t)) d\mu_{0k}(t)\right] = 0 \qquad (2.22)$$

and $\sum_{i=1}^{n} \int_0^s [dN_{ik}(t) - Y_{ik}(t)g(\beta' Z_{ik}(t))d\mu_{0k}(t)] = 0 \; \forall s \in (0, \tau]$, where $P(Y_{ik}(\tau) = 1) > 0$ $\forall k = 1, 2, \ldots, K$. The authors showed that the solution to the estimating equation, $\hat{\beta}$, would be a consistent estimator of β_0 and $n^{1/2}\left(\hat{\beta} - \beta_0\right)$ converges weakly to a Gaussian distribution with mean zero and variance $\Sigma(\beta_0)$, where $\Sigma(\beta) = A(\beta)^{-1}V(\beta)A(\beta)^{-1}$ with $A(\beta) = \sum_{k=1}^{K} \int_0^\tau v_k(bz, t)s_k^{(0)}(\beta_0, t)d\mu_{0k}(t)$, where $v_k(\beta, t) = \frac{s_k^{(2)}(\beta,t)}{s_k^{(0)}(\beta,t)} - (\frac{s_k^{(1)}(\beta,t)}{s_k^{(0)}(\beta,t)})^{\otimes 2}$, $s_k^{(d)}(\beta, t) = E(S_k^{(d)}(\beta, t))$ and $S_k^{(d)}(\beta, t) = \frac{1}{n}\sum_{i=1}^{n} Y_{ik}(t)Z_{ik}(t)e^{\beta' Z_{ik}(t)}$, for d = 0,1,2; $V(\beta) = E\left[(\sum_{k=1}^{K} U_{1k}(\beta))^{\otimes 2}\right]$ with $U_{i,k}(\beta) = \int_0^\tau \left(Z_{ik}(t)\frac{g^{(1)}(\beta' Z_{ik}(t)}{g(\beta' Z_{ik}(t)} - \frac{s_k^{(1)}(\beta,t)}{s_k^{(0)}(\beta,t)}\right) dM_{ik}(t, \beta)$, $dM_{ik}(t, \beta) = dN_{ik}(t) - Y_{ik}(t)g(\beta' Z_{ik}(t))d\mu_{0k}(t)$. Schaubel and Cai (2005) proposed methods of estimating parameters in two semi-parametric proportional rates/means models for recurrent events with clustering among subjects. One of the models considered the baseline rate function to be common across clusters, while the other model had cluster-specific baseline rates. All these methods have looked at the marginal distribution of the rate function or the mean function (and sometimes, the intensity function). Huang and Chen (2003) investigated a marginal proportional hazards model for gaps between recurrent events. They also established a connection between the gap times and the clustered data with informative cluster size.

2.3.2 Frailty Models

The marginal models discussed in the previous section does not explicitly model the intra-subject correlation. When one in interested in the effect of the risk factors on the failure times, rather than the correlation among the events for each subject, the marginal model approach suits this purpose quite well. However, in some situations, the strength and nature of dependencies within each individual's recurrent event times might be of interest and for such cases, the frailty models have been developed and studied extensively in the literature. The frailty model explicitly formulated the underlying dependence structure through an unobservable random variable. This unknown quantity is called the frailty. The key assumption is that the individual failure times are independent conditional on the frailty term for each individual. Considering the Cox PH model, the intensity of the i-th individual for the k-th event, conditional on the frailty term, w_i would be given by

$$\lambda_i(t \mid w_i) = w_i \lambda_0(t) exp(\beta_0' Z_i(t)), \tag{2.23}$$

where the frailty terms are independent and identically distributed, having some parametric distribution. Various choices are possible for the density of the frailty model, including the positive stable distribution, inverse-Gaussian distribution, log-normal distribution, the most common being the gamma distribution, because of mathematical convenience.

The parameter estimates are obtained through the EM algorithm, using the partial likelihood in the maximization step (Klien 1992). Therneau and Grambsch (2001) showed that another approach to estimate the distribution of the shared frailty would be to use penalized partial likelihood. One important thing to note is that in equation (2.23), β_0 is to be interpreted conditional on the frailty term. There has been extensive debate over which method is more naturally suited for correlated data. The marginal modeling approach does not depend on any models for the underlying correlation structure and would result in more robust results under misspecification of the correlation structure. On the other hand, if it is possible to learn more about the structure, one might prefer the frailty model to get more efficient estimators. When the main purpose of the analysis is to model the effect of the covariates and the dependence can be treated as a nuissance, it is more preferable to use the marginal models, whereas if the interest lies in the nature of the dependency and quantifying that, the conditional frailty model would be much more sensible. Hence, the choice of the type of model depends on the goal of the study.

2.4 Additive Rates Models for Marginal Analysis

The models that have been discussed thus far assumes multiplicative risk models as the effect of the covariates is multiplicative. In epidemiological studies, one might be interested in the risk difference, rather than the risk ratio. The risk difference can be translated directly into the number of disease that would be avoided by removing a particular exposure (Kulich and Lin 2000). In certain settings, for example, in studies of health care utilization, absolute covariate effects are of direct interest. Based on an additive model, the predicted change in rate attributable to a covariate can be easily predicted without information on the baseline cost. The cost savings associated with a proposed intervention program in health intervention studies can be directly calculated from an additive model, whereas information on baseline cost is required if a multiplicative model is fitted. Further, the effect is not directly represented by the regression coefficient in the multiplicative model. When one is interested in the risk difference as the measure of interest, the additive rates model provides a good alternative to the widely studied Cox (1972) proportional hazards model. Since temporal effects are not assumed to be proportional for each covariate, the additive risk model would be more useful in providing information about the temporal influence of each covariate which is not available from the Cox PH model. Specifically, in studies of excess risk, where the background risk and the excess risk can have very different temporal forms, additive risk models seem to be biologically more plausible than proportional hazard models (Huffer and McKeague 1991). Under the additive risk model, the hazard function for the failure time, T associated with Z(.) takes the form

$$\lambda(t \mid Z) = \lambda_0(t) + \beta_0' Z(t), \tag{2.24}$$

where $\lambda_0(t)$ is the unspecified baseline hazard function and β_0 is the p-dimensional vector of regression parameters. Lin and Ying (1994) were the first to develop semi-parametric estimating equation for equation (2.24) along with the asymptotic distribution. They proposed to estimate β_0 using the following estimating equation which mimics the partial likelihood score function for the additive hazards model.

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left[Z_i(t) - \bar{Z}(t)\right] \left\{dN_i(t) - Y_i(t)\beta' Zit\right\}, \tag{2.25}$$

where $\bar{Z}(t) = \frac{\sum_{i=1}^{n} Y_i(t) Z_i(t)}{\sum_{i=1}^{n} Y_i(t)}$. One can easily obtain $\hat{\beta}$ by solving equation $U(\beta) = 0$ and it is given by

$$\hat{\beta} = \left[\sum_{i=1}^{n} \int_0^{\tau} Y_i(t) \left\{Z_i(t) - \bar{Z}(t)\right\}^{\otimes 2} dt\right]^{-1} \left[\sum_{i=1}^{n} \int \left\{Z_i(t) - \bar{Z}(t)\right\} dN_i(t)\right]. \tag{2.26}$$

Under some regularity conditions, $\hat{\beta}$ is a consistent estimator of β_0 and the asymptotic distribution of $n^{1/2}(\hat{\beta} - \beta_0)$ is p-variate Gaussian with mean zero and sandwich variance term which can be consistently estimated by $A^{-1}BA^{-1}$ where

$$A = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{Z_i(t) - \bar{Z}(t)\right\}^{\otimes 2} dt, \quad B = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \left\{Z_i(t) - \bar{Z}(t)\right\}^{\otimes 2} dN_i(t).$$

The estimators for the baseline cumulative hazard function, $\Lambda_0(t)$, and the survival function, $S(t, Z)$, were proposed and their asymptotic properties were explored. To ensure monotonicity, modified estimators of the above quantities, $\hat{\Lambda}_0^*(t) = \max_{s \leqslant t} \hat{\Lambda}_0(s)$ and $\hat{S}_0^*(t, Z) = \max_{s \leqslant t} \hat{S}_0(s, Z)$ have also been proposed. These estimators were shown to be asymptotically equivalent to the original estimators.

Kulich and Lin (2000) studied additive hazards models in case-cohort design scheme. They proposed a weighted estimating equation by modifying (2.25) as

$$U_W(\beta) = \sum_{i=1}^{n} \rho_i \int_0^{\tau} \left(Z_i(t) - \bar{Z}_W(t)\right)\left[dN_i(t) - Y_i(t)\beta' Z_i(t)dt\right], \tag{2.27}$$

where

$$\bar{Z}_W(t) = \frac{\sum_{i=1}^{n} \rho_i Y_i(t) Z_i(t)}{\sum_{i=1}^{n} \rho_i Y_i(t)}, \quad \rho_i = \Delta_i + (1 - \Delta_i)\frac{\xi_i}{p_i}, \quad p_i = P(\xi_i = 1).$$

The resulting estimator, $\hat{\beta}_H$, has the following closed form :

$$\hat{\beta}_W = \left[\sum_{i=1}^{n} \rho_i \int_0^{\tau} Y_i(t)\left\{Z_i(t) - \bar{Z}_W(t)\right\}^{\otimes 2} dt\right]^{-1} \left[\sum_{i=1}^{n} \int \left\{Z_i(t) - \bar{Z}_W(t)\right\} dN_i(t)\right].$$

Under certain regularity conditions, they showed that $\hat{\beta}_W$ is consistent for β_0 and $n^{1/2}(\hat{\beta}_W - \beta_0)$ has a Gaussian distribution with mean zero and sandwich variance term, corresponding to two different sampling situations. They further proposed an estimator of the cumulative hazard function as

$$\hat{\Lambda}_{W0}(t) = \int_t \frac{\sum_{i=1}^{n} dN_i(u)}{\sum_{i=1}^{n} \rho_i Y_i(u)} - \int_0^t \hat{\beta}_W \bar{Z}_W(t) dt.$$

$n^{1/2}(\hat{\Lambda}_{W0}(t) - \Lambda_0(t))$ converges asymptotically to a Gaussian process on $[0, \tau]$ with mean 0 and appropriate variance term. The additive hazards model has been applied to interval censored data (Lin et al. 1998, Martinussen and Scheike 2002), competing risk analysis of case-cohort studies (Sun et al. 2004), and the recurrent gap times (Sun et al. 2006). Yin and Cai (2004) proposed additive hazards regression model to multivariate failure time data where they considered correlation between different time points for the individuals. Schaubel et al. (2006) proposed the semi-parametric additive rates model to fit recurrent events. The authors noted that the model can be used for any process with non-negative increments. The model assumes that the covariates affect the unspecified baseline rate additively. As in the multiplicative model, the additive rates model is defined as

$$E\left[dN_i^*(t) \mid Z_i(t)\right] = d\mu_0(t) + \beta_0' Z_i(t)dt, \tag{2.28}$$

where $dN_i^*(t) = N_i^*(t+dt) - N_i^*(t)$ $(dt \downarrow 0)$ equals the increment in $N_i^*(t)$ over the small interval $(t, t+dt]$ and $E\left[dN_i^*(t) \mid Z_i(t), C_i > t\right] = E\left[dN_i^*(t) \mid Z_i(t)\right]$. Note that unlike intensity functions, the expectation is not considered, conditional on the entire history. In general, the intensity and rate functions can be related by

$$E\left[dN_i^*(t) \mid Z_i(t)\right] = E\left[E\left\{dN_i^*(t) \mid \mathcal{F}_i(t)\right\} \mid Z_i(t)\right].$$

The estimator of β_0 can be easily obtained by solving (2.25) where $dN_i(t) = \mathbf{I}(C_i > t)dN_i^*(t)$ and $Y_i(t) = \mathbf{I}(C_i > t)$. Let us define $M_i(t, \beta) = N_i(t) - \int_0^t Y_i(u)\{d\mu_0(u) + \beta' Z_i(u)du\}$ which has mean zero at $\beta = \beta_0$. Following Liang and Zeger (1986)'s generalized estimating equations the authors defined the following estimating functions for β_0 and $\mu_0(t)$.

$$\sum_{i=1}^{n} \int_0^t \mathbf{I}(C_i > u)dM_i(u, \beta) = 0. \tag{2.29}$$

$$\sum_{i=1}^{n} \int_0^t \mathbf{I}(C_i > u)Z_i(u)dM_i(u, \beta) = 0. \tag{2.30}$$

Solving these two equations, we get $\hat{\beta}$, which is explicitly defined in (2.26) with the asymptotic distribution of $n^{1/2}\{\hat{\beta} - \beta_0\}$ to be of the same form as the sandwich estimator described earlier.

The estimator of the baseline mean is given by

$$\hat{\mu}_0(t,\beta) = \int_0^t \frac{\sum_{i=1}^n Y_i(u)\left\{dN_i(u) - \beta' Z_i(u)du\right\}}{\sum_{i=1}^n Y_i(u)}.$$

Similarly, as before, to ensure monotonicity, the baseline mean is estimated as $\tilde{\mu}_0(t,\hat{\beta}) = \max_{0\leqslant s\leqslant t}\hat{\mu}_0(s,\hat{\beta})$. Note that, it is possible for the increments in $\hat{\mu}_0(s,\hat{\beta})$ be negative. Rather than constraining $\hat{\beta}$ to force the baseline rate estimator to be positive, the baseline mean is constrained to be monotone non-decreasing.

Ma (2007) explored additive risk model with case-cohort data, with weights similar to Chen and Lo (1999). A class of estimating equations have been proposed, each depending on a different prevalence ratio estimate. Asymptotic properties of the proposed estimators and inference based on the m out of n nonparametric bootstrap are investigated. Liu et al. (2010) considered a more flexible additive-multiplicative rates model for analysis of recurrent event data, in which some covariate effects are additive while others have multiplicative effect on the rate function. The estimating equation for the regression parameters is given by

$$E\left[dN_i^*(t) \mid W_i(t)\right] = g\left(\gamma_0' Z_i(t)\right)dt + h\left(\beta_0' X_i(t)\right)d\mu_0(t),$$

where $W_i(t) = (Z_i(t), W_i(t))$ is the set of covariates, $\mu_0 = (\gamma_0', \beta_0')'$ is p-dimensional of unknown regression parameters, g and h are known link functions, $\mu_0(.)$ is an unspecified continuous baseline mean function for subjects with covariates $Z_i(t)$ and $X_i(t)$ such that $g\left(\gamma_0' Z_i(t)\right) = 0$ and $h\left(\beta_0' X_i(t)\right) = 1$. The authors propose estimating equation, mimicking the generalized estimating equations.

$$\sum_{i=1}^n \int_0^t dM_i(u,\theta) = 0, \quad \sum_{i=1}^n \int_0^\tau Q_i(u,\theta)dM_i(u,\theta) = 0,$$

where $Q_i(t,\theta)$ is a smooth p-dimensional vector-valued function of $W_i(t)$ and β, not involving $\mu_0(t)$; $M_i(t,\theta) = N_i(t) - \int_0^t Y_i(u)\left\{g\left(\gamma' Z_i(t)\right)dt + h\left(\beta' X_i(t)\right)d\mu_0(t)\right\}$. The estimators for these regression parameters from these equations are shown to be consistent and asymptotically normally distributed under appropriate regularity conditions. Kang et al. (2013) studied marginal additive hazards model for case-cohort studies with multiple disease outcomes. The estimating equation is very similar to (2.27) with another summation, for the disease category. He et al. (2013) proposed semi-parametric additive rates model, noting that individuals within a cluster might not be independent and that creates another level of complexity in recurrent events data. Liu et al. (2013) proposed an additive transformation model for modeling recurrent events. The rate function is defined by

$$\mu(t \mid Z) = \mu_0(t) + Q(t, \beta_0' Z(t)),$$

where β_0 is the vector of regression parameters, $\mu_0(.)$ is an unspecified non-negative function with $\mu_0(0) = 0$, and $Q(t, x)$ is a pre-specified non-negative link function with $Q(0, x) = 0$ for any x. Yu et al. (2014) and Cao and Yu (2017) propose different estimating equations to model the additive hazards model for generalized case-cohort (GCC) sampling and optimal GCC sampling respectively.

The additive and multiplicative risk models provide two major frameworks for studying the association between risk factors and the event occurrence (e.g., some disease of interest or death). Most survival analyses focus on multiplicative models for relative risk using the proportional hazards theory, mainly due to the easy interpretability of the risk ratio, desirable theoretical properties and the availability of computer programs to fit the models readily. However, in many biomedical studies the researchers might be interested in risk difference, rather than the relative risk or the PH assumptions are not valid which might lead to biased results if the multiplicative models are fit, ignoring the conditions (O'Neill, 1986). In such situations, it is recommended to use the additive rates model than the more easily available multiplicative model.

2.5 Sample Size Calculation for Case-Cohort Design

When the event is rare or the cost of collecting the covariate data for the entire cohort is high, there has been a lot of work done in analyzing case-cohort data. Previous sections have addressed the analyses of case-cohort data; this section would focus on the design aspect of the case-cohort sampling scheme. Sample size and power calculation are important components at the design stage of any study. Specifically, for the case-cohort studies, it is very helpful in determining the size of the subcohort needed for the study. It is also of interest to know how much more power may be gained by using a case-cohort design, compared with a simple random sample design conditional on the costs involved in the study.

Cai and Zeng (2004) proposed methods for computing sample size and power for case-cohort studies. The authors presented two log-rank type test statistics and derived the sample size and power for them. Denote the number of subjects in the j-th group (1,2) by n_j, T_{ij} is the potential failure time and C_{ij} is the censoring time and one observes $X_{ij} = \min(T_{ij}, C_{ij})$ with $\Delta_{ij} = \mathbf{I}(T_{ij} \leqslant C_{ij})$, n = $n_1 + n_2$. ξ_{ij} is the indicator that the i-th subject from the j-th group is selected in the sub-cohort. $\Lambda_j, j = 1, 2$ is the cumulative hazard function of the failure time in group j and the null hypothesis is $H_0 : \Lambda_1(t) = \Lambda_2(t)$. The test statistic would be given by

$$T_{sp} = \sum_{i=1}^{n_1} \frac{\Delta_{i1} w(X_{i1}) \tilde{Y}_2(X_{i1})}{\tilde{Y}_1(X_{i1}) + \tilde{Y}_2(X_{i1})} - \sum_{i=1}^{n_2} \frac{\Delta_{i2} w(X_{i2}) \tilde{Y}_1(X_{i2})}{\tilde{Y}_1(X_{i2}) + \tilde{Y}_2(X_{i2})}, \tag{2.31}$$

where $w(.)$ is the weight function and $\tilde{Y}_k(X_{ik}) = q\bar{Y}_k(X_{ik})$, k = 1,2, where q is the sampling fraction for the sub-cohort. This statistic is exactly the score function of the pseudo-partial likelihood function given by Self and Prentice (1988). Using that, the distribution of this would be easily derivable. Instead of using $\tilde{Y}_k$, the authors also proposed another way of approximating the total risk set by

$$Y_j^*(t) = \frac{n}{\tilde{n}} \sum_{i=1}^{n_j} \mathbf{I}(X_{ij} \geqslant t, \xi_{ij} = 1, \Delta_{ij} = 0) + \sum_{i=1}^{n_j} \mathbf{I}(X_{ij} \geqslant t, \Delta_{ij} = 1).$$

As in (2.31), one can get another test statistic by simply replacing $\tilde{Y}_j(t)$ by $Y_j^*(t)$. Furthermore, the asymptotic variance of $n^{-1/2} T_{sp}$ can be estimated by

$$\hat{\sigma}_{sp}^2 = \hat{\sigma}^2 + \frac{2(1-\hat{q}}{n} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \left[\frac{\Delta_{ij} w(X_{ij}) \tilde{Y}_1(X_{ij}) \tilde{Y}_2(X_{ij})}{\left(\tilde{Y}_1(X_{ij}) + \tilde{Y}_2(X_{ij})\right)^2} \left\{ \sum_{j'=1}^{2} \sum_{i'=1}^{n_{j'}} \frac{\Delta_{i'j'} w(X_{i'j'}) \mathbf{I}(X_{i'j'} \leqslant X_{ij})}{\tilde{Y}_1(X_{i'j'}) + \tilde{Y}_2(X_{i'j'})} \right\} \right]$$

$$- \frac{(1-\hat{q}}{n} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \frac{\Delta_{ij} w(X_{ij}) \tilde{Y}_1(X_{ij}) \tilde{Y}_2(X_{ij})}{\left(\tilde{Y}_1(X_{ij}) + \tilde{Y}_2(X_{ij})\right)^3},$$

where $\hat{\sigma}^2 = \frac{1}{n} \left[\sum_{i=1}^{n_1} \frac{\Delta_{i1} w(X_{i1}) \tilde{Y}_2(X_{i1})^2}{\left(\tilde{Y}_1(X_{i1}) + \tilde{Y}_2(X_{i1})\right)^2} + \sum_{i=1}^{n_2} \frac{\Delta_{i2} w(X_{i2}) \tilde{Y}_1(X_{i2})^2}{\left(\tilde{Y}_1(X_{i2}) + \tilde{Y}_2(X_{i2})\right)^2} \right]$ is the consistent estimate for the variance of the log-rank test based on the entire data set, $\hat{q} = \frac{n}{\tilde{n}}$. The asymptotic variance of the other test statistic was estimated by $\hat{\sigma}^2_{sp}$ as well. The authors compared the sample size and power based on these statistics and the log-rank test based on the entire cohort and that based on only the sub-cohort. Cai and Zeng (2007) extended this idea to generalized case-cohort design, when the event is not rare. As described earlier, in a generalized case-cohort design, one randomly samples without replacement a subcohort from the full cohort in the first step and then in the second step, another sample is randomly taken without replacement from the remaining failures. The observed data in this case would be, $(X_i, (\xi_i + (1-\xi_i)\Delta_i \eta_i), Z_i, \Delta_i), i = 1, 2, \ldots, n$. Assuming that Z is a dichotomous variable, one can construct the following test statistic :

$$W_n = \int_0^\tau \frac{w(t)\tilde{Y}_1(t)\tilde{Y}_2(t)}{\tilde{Y}_1(t) + \tilde{Y}_2(t)} \left\{ \frac{d\tilde{N}_1(t)}{\tilde{Y}_1(t)} - \frac{d\tilde{N}_2(t)}{\tilde{Y}_2(t)} \right\},$$

where τ is the study duration, w(t) is the weight function, $\tilde{Y}_j(t) = \sum_{i=1}^{n} \mathbf{I}(X_i \geqslant t, Z_i = j) \times \left[\xi_i \Delta_i + \frac{\xi_i(1-\Delta_i)}{p} + \frac{\Delta(1-\xi_i)\eta_i}{q} \right]$ and $\tilde{N}_j(t) = \sum_{i=1}^{n} \Delta_i \mathbf{I}(X_i \geqslant t, Z_i = j) \left\{ \xi_i + \frac{\xi_i(1-\Delta_i)}{p} \right\}$. Denoting $\alpha(t) = \frac{P(Z=1, C \geqslant t)}{P(C \geqslant t)}$, one can see that $n^{-1/2} W_n$ converges weakly to a Gaussian distribution with mean zero and variance $\sigma^2_W = \sigma^2 + \frac{1-p}{p} Var(\nu(X,Z)(1-\Delta)) + \frac{P(\Delta=1)(1-p)(1-q)}{q} Var(\mu(X,Z) \mid \Delta = 1, \xi = 0)$ with $\nu(X,Z) = -\int_0^X w(t)(1-\alpha(t))d\Lambda(t)\mathbf{I}(Z=1) + \int_0^X w(t)\alpha(t)d\Lambda(t)\mathbf{I}(Z=2)$ and $\mu(X,Z) = \nu(X,Z) + w(Z)(1-\alpha(Z))\mathbf{I}(X=1) - w(Z)\alpha(Z)\mathbf{I}(Z=2)$. Stratified sampling is commonly used in the survey sampling to improve the estimation precision for the population quantity of interest. Hu et al. (2014) derived the sample size/power calculation for a stratified case-cohort (SCC) design, based on a stratified test statistic. The test statistic looks very similar to (22) and can be defined as

$$T_{sp} = \sum_{l=1}^{L} T_{sp,l} = \sum_{i=1}^{n_{l1}} \frac{\Delta_{li1} w(X_{li1}) \tilde{Y}_{l2}(X_{li1})}{\tilde{Y}_{l1}(X_{li1}) + \tilde{Y}_{l2}(X_{li1})} - \sum_{i=1}^{n_{l2}} \frac{\Delta_{li2} w(X_{li2}) \tilde{Y}_{l1}(X_{li2})}{\tilde{Y}_{l1}(X_{li2}) + \tilde{Y}_{l2}(X_{li2})},$$

where L is the number of groups. Since each of the groups are independent among themselves, the variance term would be similar to σ_{sp}^2, summed over all possible strata. The authors investigated and compared the proportional, balanced, and optimal sampling design methods and derived the corresponding sample size calculation formula. In the next chapter, we discuss our proposed work.

CHAPTER 3: MULTIPLICATIVE RATES MODEL FOR RECURRENT EVENTS IN CASE-COHORT DATA

3.1 Introduction

In large epidemiological studies and disease prevention trials, the majority of the effort and cost arises from the assembling of the covariate measurements and follow-up information on all the individuals. When the disease incidence is low and some exposures are expensive to measure, it is not feasible and not cost effective to measure the expensive variable on all individuals in the cohort. To reduce cost and achieve the same study goals as the cohort study, Prentice (1986) proposed the case-cohort study design. Under this design, a random sample is selected from the entire cohort, named sub-cohort, and covariate information is collected on only this sub-cohort and the individuals who experience the event.

Development of statistical methods for data from case-cohort studies is an active research area. For univariate failure time data, Self and Prentice (1988), Wacholder et al. (1989), and Barlow (1994) considered efficient and robust estimation of the variance of the case-cohort estimator. Borgan et al. (1995) considered a more general sampling frame whereas Lin and Ying (1993) viewed the case-cohort design as a special case of the missing data problem. Borgan et al. (2000) developed methods for the analysis of exposure stratified case-cohort design and Breslow and Wellner (2006) considered weighted likelihood for two-phase stratified samples. Chen (2001) and Kulich and Lin (2004) developed sample reuse methods via local averaging leading to more efficient estimation. Nonetheless, correlated failure time data are quite common in biomedical and public health research. Lu and Shih (2006) and Zhang et al. (2011) developed estimating equation for clustered failure time data assuming a marginal hazards model, accounting for correlation within clusters, which are formed by correlated subjects. Kang and Cai (2009a) considered marginal hazards model for case-cohort data with multiple disease outcomes. However, methods for analyzing recurrent events data from case-cohort studies are scarce.

Recurrent events are commonly encountered in biomedical research. Our motivating example is from a doubly-blind, placebo-controlled community trial conducted in northeastern Brazil in a cohort of children aged between 6 to 48 months (Barreto et al. 1994). The primary objective of this study was to evaluate the effect of high doses of vitamin A on acute-lower-respiratory-tract infection (ALRI). One thousand two hundred and four children were randomized to receive vitamin A supplement or placebo. They were followed for 1 year. An episode of ALRI was defined as cough plus a respiratory rate of 50 breaths per min or higher for children under 12 months, and 40 breaths per min or higher for older children(Barreto et al. 1994). About 15.38% of the children had at least one ALRI episode during their follow-up period. The number of episodes ranged from 1 to 6, resulting in a total of 305 episodes. As a secondary objective, it is of interest to examine whether the child ever had measles is related to ALRI. It can be expensive to verify the measles information because it is based on the parents' acknowledgement. With the relatively low ALRI rate, a case-cohort sampling design can be more cost-effective in this situation.

Various methods have been proposed for analyzing recurrent event data from the full cohort. These include modeling the intensity functions of the recurrent event process (Andersen and Gill 1982), rate/mean function (Pepe and Cai 1993, Lawless and Nadeau 1995, Cook et al. 2009, Lin et al. 2000), and the gap times between each recurrence (Huang and Chen 2003, Schaubel and Cai 2004). However, methods for analyzing recurrent events data from case-cohort studies is limited. Chen and Chen (2014) extended the case-cohort design to recurrent events with specific clustering feature using a modified Cox-type self-exciting intensity model. Such model makes the assumption that the dependence of the recurrent events is captured by some time-varying covariates. This assumption may not be easily verifiable. An alternative is to model the marginal rate or mean function. The marginal rates or means model does not require such assumption (Lin et al. 2000) and the parameters in this model have population average interpretation, which is desirable in many population studies. However, analysis methods for marginal rates model have not been investigated for recurrent events data from case-cohort study design.

The main goal of this article is to propose case-cohort designs for recurrent events data and the estimation procedures for data from such designs. We considered two different situations. One is when the recurrent events are not very common in which case we will include into the case-cohort sample all individuals who developed events during the follow-up. The second situation is when the recurrent events are relatively common, when the proportion of subjects who experienced at least one event is about $20\% - 50\%$. In such case, we propose to only include into the case-cohort study a sample of those who developed events during the follow-up. We refer to the first situation as traditional case-cohort design and the second as the generalized case-cohort design.

In this chapter, we propose weighted estimating equations for estimating the parameters in the marginal rates regression model for recurrent events in case-cohort studies. The chapter is organized as follows. In section 3.2, the design of the study and the estimation procedure are proposed. The asymptotic properties of the estimators are studied in section 3.3. The finite sample properties are investigated by simulations in section 3.4. In section 3.5, we illustrate the proposed method on a case-cohort study based on the ALRI data on children in Brazil. In the section 3.6, we provide some final remarks.

3.2 Model and Estimation

Suppose there are n independent individuals in the cohort. Let $N_i^*(t)$ be the number of recurrent events for the individual i over the time interval $[0, t)$, C_i is the censoring time. $Z_i(.) = (Z_i^E(.)', Z_i^C(.)')'$ is the p-dimensional covariate of interest for the individual i, where $Z_i^E(.)$ is the set of expensive-to-measure variables and $Z_i^C(.)$ is the set of all other covariates. Let T_{ij}^* denote the j-th recurrent event time for the individual i. The observed time is $T_{ij} = T_{ij}^* \wedge C_i$, $j = 1, 2, \dots, n_i + 1$, where n_i is the number of events that are observed for individual i, and $N = \sum_{i=1}^n n_i$ is the total number of observed events. Let $Y_i(t) = \mathbf{1}(C_i \geq t)$, $N_i(t) = N_i^*(t \wedge C_i) = \sum_{j=1}^{n_i} \mathbf{I}(T_{ij} \leq t) = \sum_{j=1}^{n_i} \mathbf{I}(T_{ij}^* \leq t) Y_i(t)$, $\Delta_i = 1 - \mathbf{I}(\min_j(T_{ij}^*) \geq C_i)$ which is the indicator that individual i experienced at least one event, and τ denote the study ending time. The rate function for an individual is denoted as $E\left(dN^*(t) \mid Z(t)\right) = d\mu_z(t)$. We assume the following proportional rates model:

$$d\mu_Z(t) = e^{\theta_0' Z^E(t) + \gamma_0' Z^C(t)} d\mu_0(t) \Rightarrow \mu_Z(t) = \int_0^t e^{\theta_0' Z^E(u) + \gamma_0' Z^C(u)} d\mu_0(u), \tag{3.1}$$

where $\mu_0(.)$ is an unspecified continuous baseline mean function and θ_0 and γ_0 are the vectors of unknown parameters. Denoting $\beta_0 = (\theta_0', \gamma_0')'$, we can rewrite the rates model as $d\mu_Z(t) = e^{\beta_0' Z(t)} d\mu_0(t)$ and the mean function is given by $\mu_0(t) = \int_0^t e^{\beta_0' Z(u)} d\mu_0(u)$ for all $t \in [0, \tau]$. Note that the covariates are allowed to be time-dependent. We assume that the possibly time-dependent covariates are external (Kalbfleisch and Prentice (2002)), i.e., they are not affected by the recurrent event process.

3.2.1 Case-cohort study design for recurrent events

In this subsection, we introduce two sampling schemes for the recurrent event data. The first deals with the situation that the event is not common in the population. In this case, we draw a random sample from the full cohort and supplement that with all the cases. We call this sampling scheme the original case-cohort design. The second sampling scheme is for the situation that the event is relatively common and we cannot afford to sample all individuals with events. An example for the common recurrent event is the randomized double-blinded trial conducted by Genentech Inc. in the early 1990's to study the effect of rhDNase on pulmonary exacerbations among patients with cystic fibrosis (Therneau and Hamilton 1997). Even though, the pulmonary exacerbation rate is $\sim 40\%$, the variable FEV_1 may be quite expensive to measure. Under such situation, we propose to sample only a fraction of those who have events for the case-cohort sampling. We call this sampling scheme the generalized case-cohort design with recurrent events.

3.2.2 Estimation under the original case-cohort design

Under the case-cohort sampling, we select a sub-cohort from the entire cohort by simple random sampling. Let ξ_i denote the indicator function for individual i being selected into the subcohort; $\tilde{\alpha} = \frac{\tilde{n}}{n}$ is the subcohort proportion where $\tilde{n}$ is the number of individuals selected in the sub-cohort and n is the number of individuals in the full cohort. We call an individual a case if the individual experienced at least one event and an individual a non-case if the individual did not have an event during the study period. Hence, the observable information for individual i is $\{\mathbf{T}_i, \Delta_i, \xi_i, Z_i^C(t), Z_i^E(t),\ t \in [0, \tau]\}$ if individual i is in the case-cohort sample. In other words, we have $\{\mathbf{T}_i, \Delta_i, \xi_i, Z_i(t),\ 0 \le t \le \tau\}$ if $\Delta_i = 1$ or $\xi_i = 1$ and $\{\mathbf{T}_i, \Delta_i, \xi_i, Z_i^C(t), t \in [0, \tau]\}$ when $\Delta_i = 0$ and $\xi_i = 0$. When information on the covariates for all the individuals are available, one can consider the following estimating equation for the full cohort data (Lin et al. (2000)):

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ Z_i(t) - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right\} dN_i(t) = 0, \tag{3.2}$$

where $S^{(d)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)}$, $\forall d = 0, 1$. One can easily solve the estimating equation by some iterative algorithm, for example, the Newton-Raphson iteration method. However, because the data are not complete in case-cohort studies, (3.2) cannot be used directly. We consider a weighted estimating equation approach based on the idea of inverse probability of selection weighting. The estimating equation considered for estimating β_0 is the following:

$$\hat{U}^I(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ Z_i(t) - \frac{\tilde{S}^{(1)}(\beta, t)}{\tilde{S}^{(0)}(\beta, t)} \right\} dN_i(t) = 0, \tag{3.3}$$

where $\tilde{S}^{(d)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} w_i^I(t) Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)}$, $\forall d = 0, 1$, $w_i^I(t) = \Delta_i + (1 - \Delta_i) \frac{\xi_i}{\hat{\alpha}(t)}$, where $\hat{\alpha}(t) = \frac{\sum_i (1-\Delta_i)\xi_i Y_i(t)}{\sum_i (1-\Delta_i) Y_i(t)}$ is the estimator of the true sampling parameter, α. The weight is 1 for all the cases and is $\hat{\alpha}(t)^{-1}$ for the non-cases in the sub-cohort. Similar idea for the weights was considered by Kalbfleisch and Lawless (1988).They considered the time-invariant version of $\hat{\alpha}(t)$, which was given by $\tilde{\alpha}$. Borgan et al. (2000) used a similar idea for univariate failure time data from stratified case-cohort studies. We denote the solution to this equation by $\hat{\beta}^I$. Our proposed Breslow-Aalen type estimator of the baseline mean function is given by

$$\hat{\mu}_0(\hat{\beta}^I, t) = \int_0^t \frac{\sum_{i=1}^{n} dN_i(u)}{n \tilde{S}^{(0)}(\hat{\beta}^I, u)}. \tag{3.4}$$

3.2.3 Estimation under the generalized case-cohort design

For the generalized case-cohort design, we sample a fraction of cases outside of the sub-cohort. Let η_i be an indicator for individual i who is a case but outside the sub-cohort being sampled. Let $\tilde{q} = \frac{n_1^*}{n_1 - \tilde{n}_1}$ denote the sampling proportion for the additional cases, where n_1^*, n_1 and $\tilde{n}_1$ are the number of selected individuals who have experienced at least one event but are not in the subcohort, individuals who experienced at least one event in the full cohort and those who were in the subcohort respectively. Under this design, the covariate information is available for the subcohort members and the selected cases ($\eta_i = 1$). Hence, the observable information for individual i is $\{\mathbf{T}_i, \Delta_i, \xi_i, \eta_i, Z_i(t) : t \in [0, \tau]\}$ when $\xi_i = 1$ or $\eta_i = 1$ and $\{\mathbf{T}_i, \Delta_i, \xi_i, Z_i^C(t), t \in [0, \tau]\}$ if $\xi_i = 0$ and $\eta_i = 0$. Using the inverse of probability of being sample as the weight, our proposed estimating equation for the Generalized case-cohort sampling scheme is

$$\hat{U}^{II}(\beta) = \sum_{i=1}^{n} \int_0^{\tau} w_i^{II}(t) \left\{ Z_i(t) - \frac{\hat{S}^{(1)}(\beta, t)}{\hat{S}^{(0)}(\beta, t)} \right\} dN_i(t) = 0, \tag{3.5}$$

where $\hat{S}^{(d)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} w_i^{II}(t) Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)}$, $\forall d = 0, 1$, and the weight function is given by $w_i^{II}(t) = \frac{(1-\Delta_i)\xi_i}{\hat{\alpha}(t)} + \Delta_i \xi_i + \frac{\Delta_i(1-\xi_i)\eta_i}{\hat{q}(t)}$ where $\hat{q}(t) = \frac{\sum_i \Delta_i(1-\xi_i)\eta_i Y_i(t)}{\sum_i \Delta_i(1-\xi_i) Y_i(t)}$ is the estimator of the true sampling parameter q. We denote the solution of this equation by $\hat{\beta}^{II}$. The Breslow-Aalen type estimator of the baseline mean function is $\tilde{\mu}_0(\hat{\beta}^{II}, t) = \int_0^t \frac{\sum_{i=1}^{n} w_i^{II}(u) dN_i(u)}{n \hat{S}^{(0)}(\hat{\beta}^{II}, u)}$.

3.3 Asymptotic properties

In this section, we investigate the asymptotic properties of the estimators. Define the following terms:

$$e(\beta, t) = \frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)}, \quad \vartheta(\beta, t) = \frac{s^{(2)}(\beta, t) s^{(0)}(\beta, t) - s^{(1)}(\beta, t)^{\otimes 2}}{(s^{(0)}(\beta, t))^2}, \quad \tilde{Z}_i(t) = Z_i(t) - e(\beta, t),$$

$$dM_i(t) = dN_i(t) - Y_i(t) e^{\beta_0' Z(t)} d\mu_0(t), \quad M_{\tilde{Z},i}(\beta) = \int_0^{\tau} \tilde{Z}_i(t) dM_i(t), \quad dM_{\tilde{Z},i}(\beta, t) = \tilde{Z}_i(t) dM_i(t),$$

$$A(\beta) = \int_0^{\tau} \vartheta(\beta, t) s^{(0)}(\beta, t) d\mu_0(t).$$

We define the norm for the vector m, matrix M, and function f as the following: $\|m\| = \max_i |m_i|$, $\|M\| = \max_{i,j} |M_{ij}|$, $\|f\| = \sup_t |f(t)|$. The estimator under the traditional case-cohort sampling scheme is a special case of the generalized case-cohort sampling scheme, so its asymptotic property is a special case of the generalized case-cohort sampling scheme. Hence, in the Appendix, we focus the proofs on the asymptotic properties of the estimators under the generalized case-cohort design, $\hat{\beta}^{II}$ and $\hat{\mu}_0^{II}(\hat{\beta}^{II}, t)$. The regularity conditions and the outline of the proofs are provided in the Appendix. The asymptotic properties are summarized in the following theorems.

Theorem 1. *Under the regularity conditions in the Appendix, for $k = I$ or II, $\hat{\beta}^k$ is a consistent estimator of β_0. $n^{1/2}\{\hat{\beta}^k - \beta_0\}$ converges to a Gaussian distribution with mean zero and variance given by*

$$\Sigma^k(\beta_0) = A(\beta_0)^{-1}\left[Q(\beta_0) + \frac{1-\alpha}{\alpha}V^I(\beta_0) + \boldsymbol{I}(k = II)(1-\alpha)\frac{1-q}{q}P(\Delta_1 = 1)V^{II}(\beta_0)\right]A(\beta_0)^{-1},$$

where

$$Q(\beta) = E\left(M_{\tilde{Z},1}(\beta)\right)^{\otimes 2},$$

$$V^I(\beta_0) = var\left((1-\Delta_1))\int_0^\tau \left[R_1(\beta_0, t) - \frac{Y_1(t)E((1-\Delta_1)R_1(\beta_0,t))}{E((1-\Delta_1)Y_1(t))}\right]d\mu_0(t)\right),$$

$$V^{II}(\beta_0) = var\left(\int_0^\tau \left[dM_{\tilde{Z},1}(\beta_0,t) - \frac{Y_1(t)E(dM_{\tilde{Z},1}(\beta_0,t) \mid \Delta_1 = 1, \xi_1 = 0)}{E(Y_1(t) \mid \Delta_1 = 1)}\right] \mid \Delta_1 = 1, \xi_1 = 0)\right),$$

$$A(\beta) = \int_0^\tau \vartheta(\beta,t)s^{(0)}(\beta,t)d\mu_0(t), \quad R_i(\beta,t) = Y_i(t)\tilde{Z}_i(t)e^{\beta Z_i(t)}.$$

Each of these terms, $A(\beta_0)$, $Q(\beta_0)$, $V^I(\beta_0)$ and $V^{II}(\beta_0)$ can be estimated respectively by their sample counterparts, $\hat{A}(\hat{\beta}^k)$, $\hat{Q}(\hat{\beta}^k)$, $\hat{V}^I(\hat{\beta}^k)$ and $\hat{V}^{II}(\hat{\beta}^k)$. $\hat{A}(\beta) = -\frac{1}{n}\frac{\partial U(\beta)}{\partial \beta}$, $\hat{Q}(\beta) = \frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\tilde{\alpha}}\left(M_{\tilde{Z},i}(\beta)^{\otimes 2}\right)$, $\hat{M}_{\tilde{Z},i}(\beta) = \Delta_i \sum_{j=1}^{n_i}\left(Z_i(T_{ij}) - \frac{\hat{S}^{(1)}(\beta,T_{ij})}{\hat{S}^{(0)}(\beta,T_{ij})}\right) - \frac{1}{n}\sum_{j=1}^n \sum_{k=1}^{n_j} \frac{\Delta_j Y_i(T_{jk})e^{\beta Z_i(T_{jk})}}{\hat{S}^{(0)}(\beta,T_{jk})} \times \left(Z_i(T_{jk}) - \frac{\hat{S}^{(1)}(\beta,T_{jk})}{\hat{S}^{(0)}(\beta,T_{jk})}\right)$, $\hat{V}^I(\beta) = \frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\tilde{\alpha}}\left[\frac{1-\Delta_i}{n}\sum_{j=1}^n\sum_{k=1}^{n_j}\frac{\Delta_j}{\hat{S}^{(0)}(\beta,T_{jk})}\left\{\hat{R}_i^I(\beta,T_{jk}) - \frac{Y_i(T_{jk})\hat{E}\big((1-\Delta_1)R_1(\beta,T_{jk})\big)}{\hat{E}\big((1-\Delta_1)Y_1(T_{jk})\big)}\right\}\right]^{\otimes 2}$, $\hat{V}^{II}(\beta) = \frac{1}{\sum_{i=1}^n \Delta_i(1-\xi_i)}\sum_{i=1}^n \frac{\eta_i}{\tilde{q}}\Delta_i(1-\xi_i)\left[\left\{M_{\tilde{Z},i}(\beta) - \right.\right.$

$\sum_{j=1}^{n}\sum_{k=1}^{n_j} \frac{Y_i(T_{jk})\hat{E}\left(dM_{\tilde{Z},j}(\beta,T_{jk})|\Delta_j=1,\xi_j=0\right)}{\hat{E}\left(Y_1(T_{jk})|\Delta_j=1\right)}\Big\}\Big]^{\otimes 2}$, $d\hat{M}_{\tilde{Z},i}(\beta,T_{ij}) = \left(Z_i(T_{ij}) - \frac{\hat{S}^{(1)}(\beta,T_{ij})}{\hat{S}^{(0)}(\beta,T_{ij})}\right)$ $\left(\Delta_i - Y_i(T_{ij})e^{\beta Z_i(T_{ij})}d\hat{\mu}_0^k(\beta,T_{ij})\right)$, $\hat{R}_i^I(\beta,t) = \left\{Z_i(t) - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right\} Y_i(t)e^{\beta' Z_i(t)}$, $\hat{E}\left((1-\Delta_1) R_1(\beta,t)\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i}{\tilde{\alpha}}(1-\Delta_i)\hat{R}_i^I(\beta,t)$, $\hat{E}\left(dM_{\tilde{Z},1}(\beta,t) \mid \Delta_i = 1, \xi_i = 0\right) = \frac{1}{\sum_{i=1}^{n}\Delta_i(1-\xi_i)}\sum_{i=1}^{n} \frac{\eta_i}{\tilde{q}}\Delta_i(1-\xi_i)d\hat{M}_{\tilde{Z},i}(\beta,t)$, $\hat{E}\left((1-\Delta_1)Y_1(\beta,t)\right) = \frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)Y_i(t)$, $\hat{E}\left(Y_1(\beta,t) \mid \Delta_1 = 1\right) = \frac{1}{\sum_{i=1}^{n}\Delta_i}\sum_{i=1}^{n}\Delta_i Y_i(t)$.

To obtain the asymptotic distribution of the baseline mean function $\hat{\mu}_0(t)$, we need to first define the metric space, $\mathcal{D}[0,\tau]$, consisting of right continuous functions f(t) with left-hand limits and $f:[0,\tau]\to R$. The metric for this space is defined by, $d(f,g) = \sup_{t\in[0,\tau]}\{\mid f(t) - g(t)\mid\}$, $f(t), g(t) \in \mathcal{D}[0,\tau]$. The following theorem summarizes the asymptotic properties of $\hat{\mu}_0^k(t)$.

Theorem 2. *Under the regularity conditions, for $k = I$ or II, we can show that $\hat{\mu}_0^k(\beta,t)$ converges in probability to $\mu_0(t)$ uniformly in $t\in[0,\tau]$. Further, defining $W_n(t) = \hat{\mu}_0^k(\hat{\beta}^k,t) - \mu_0(t)$, we have $n^{1/2}W_n(t)$ converges to a Gaussian distribution with mean zero. The variance-covariance function between $W_n(t)$ and $W_n(s)$ is given by*

$$\phi^k(t,s)(\beta_0) = E\left(\nu_i(\beta_0,t)\nu_i(\beta_0,s)\right) + \frac{1-\alpha}{\alpha}E\left(\psi_i(\beta_0,t)\psi_i(\beta_0,s)\right) + \boldsymbol{I}(k=II)(1-\alpha)\frac{1-q}{q}P(\Delta_1 = 1) \times E\left(\zeta_i(\beta_0,t)\zeta_i(\beta_0,s) \mid \Delta_i = 1, \xi_i = 0\right),$$

where

$$\nu_i(\beta,t) = r(\beta,t)'A(\beta)^{-1}M_{\tilde{Z},i}(\beta,t) + \int_0^t \{s^{(0)}(\beta,u)\}^{-1}dM_i(u),$$

$$\psi_i(\beta,t) = \Bigg[r(\beta,t)'A(\beta)^{-1}(1-\Delta_i)\int_0^\tau \left\{R_i(\beta,u) - \frac{Y_i(u)E((1-\Delta_i)R_i(\beta,u))}{E((1-\Delta_i)Y_i(u))}\right\}d\mu_0(u)$$
$$+(1-\Delta_i)\int_0^t Y_i(u)\left(e^{\beta' Z_i(u)} - \frac{Y_i(u)E((1-\Delta_i)Y_i(u)e^{\beta' Z_i(u)})}{E((1-\Delta_i)Y_i(u))}\right)\frac{d\mu_0(u)}{s^{(0)}(\beta,u)}\Bigg],$$

$$\zeta_i(\beta,t) = \Bigg[r(\beta,t)'A(\beta)^{-1}\int_0^\tau \left(dM_{\tilde{Z},i}(\beta_0,t) - \frac{Y_i(t)E\left[dM_{\tilde{Z},1}(\beta_0,t)(\beta_0,t) \mid \Delta_1 = 1, \xi_1 = 0\right]}{E\left[Y_1(t) \mid \Delta_1 = 1, \xi_1 = 0\right]}\right)$$
$$+\int_0^t \frac{1}{s^{(0)}(\beta,u)}\left(dM_i(u) - \frac{Y_i(t)E\left[dM_1(u)(\beta_0,t) \mid \Delta_1 = 1, \xi_1 = 0\right]}{E\left[Y_1(t) \mid \Delta_1 = 1, \xi_1 = 0\right]}\right)\Bigg],$$

$r(\beta,t) = -\int_0^t e(\beta,u)s^{(0)}(\beta,u)d\mu_0(u)$.

Similarly, each of these terms can be consistently estimated by their sample counterparts. $\hat{\phi}^k(t,s)(\beta) = \frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\tilde{\alpha}}\left(\hat{\nu}_i(\beta,t)\hat{\nu}_i(\beta,s)\right) + \mathbf{I}(k = I)\frac{1}{n}\frac{1-\tilde{\alpha}}{\tilde{\alpha}}\sum_{i=1}^n \frac{\xi_i}{\tilde{\alpha}}\left(\hat{\psi}_i(\beta,t)\hat{\psi}_i(\beta,s)\right) + \mathbf{I}(k = II)\frac{1}{n}\frac{1-\tilde{q}}{\tilde{q}}\hat{P}(\Delta_1 = 1)(1-\tilde{\alpha})\frac{1}{\sum_{i=1}^n \Delta_i(1-\xi_i)}\sum_{i=1}^n \frac{\eta_i}{\tilde{q}}\left(\hat{\zeta}_i(\beta,t)\hat{\zeta}_i(\beta,s)\right)$, $\hat{\nu}_i(\beta,t) = \hat{r}(\beta,t)'\hat{A}(\beta)^{-1}$ $\hat{M}_{\tilde{Z},i}(\beta,t) + \int_0^t \{\hat{S}^{(0)}(\beta,u)\}^{-1} d\hat{M}_i(u)$, $\hat{r}(\beta,t) = -\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^{n_i}\frac{\mathbf{I}(T_{ij}\leq \min(t,C_i))\hat{S}^{(1)}(\beta,T_{ij})}{\hat{S}^{(0)}(\beta,T_{ij})^2}$, $\int_0^t \frac{d\hat{M}_i(u)}{\hat{S}^{(0)}(\beta,u)} = \sum_{l=1}^{n_i}\frac{\mathbf{I}(T_{il}\leq \min(t,C_i))}{\hat{S}^{(0)}(\beta,T_{il})} - \frac{1}{n}\sum_{j=1}^n\sum_{l=k}^{n_j}\frac{\Delta_j \mathbf{I}(T_{jk}\leq \min(t,C_j))Y_i(T_{jk})e^{\beta' Z_i(T_{jk})}}{\hat{S}^{(0)}(\beta,T_{jk})^2}$, $\hat{\psi}_i(\beta,t) = \left[\hat{r}(\beta,t)'\hat{A}(\beta)^{-1}(1-\Delta_i)\int_0^\tau \left\{\hat{R}_i(\beta,u) - \frac{Y_i(u)\hat{E}((1-\Delta_i)Y_i(u)R_i(\beta,u))}{\hat{E}((1-\Delta_i)Y_i(u))}\right\} d\hat{\mu_0}^k(u) + (1-\Delta_i)\int_0^t Y_i(u)\left(e^{\beta' Z_i(u)} - \frac{Y_i(u)\hat{E}((1-\Delta_i)Y_i(u)e^{\beta' Z_i(u)})}{\hat{E}((1-\Delta_i)Y_i(u))}\right)\frac{d\hat{\mu_0}^k(u)}{\hat{S}^{(0)}(\beta,u)}\right]$, $\hat{\zeta}_i(\beta,t) = \left[\hat{r}(\beta,t)'\hat{A}(\beta)^{-1}\Delta_i(1-\xi_i)\left\{M_{\tilde{Z},i}(\beta) - \int_0^\tau \frac{Y_i(u)\hat{E}(dM_{\tilde{Z},1}|\Delta_1=1,\xi_i=0)}{\hat{E}(Y_1(u)|\Delta_1=1)}\right\} + \Delta_i(1-\xi_i)\int_0^t \frac{1}{\hat{S}^{(0)}(\beta,t)}\left(dM_i(u) - \frac{Y_i(u)\hat{E}(dM_1(u)|\Delta_1=1,\xi_1=0)}{\hat{E}(Y_1(u)|\Delta_1=1)}\right)\right]$, $\hat{E}((1-\Delta_i)Y_i(u)e^{\beta' Z_i(u)}) = \frac{1}{n}\sum_{i=1}^n(1-\Delta_i)\frac{\xi_i}{\tilde{\alpha}}Y_i(t)exp\{\beta' Z_i(t)\}$.

Studying the variance components, we can identify three sources of variation in $\Sigma^{II}(\beta_0)$ and $\phi^{II}(t,s)(\beta_0)$. The three components correspond to the variation due to the different sampling present in the data: one from the cohort, one because of the sampling of the subcohort from the cohort, and the last is from the sampling of the cases outside the subcohort. Further, note that for the original case-cohort design, since no randomness arises from sampling cases outside the random sub-cohort, the third term does not arise in the variance term of $\hat{\beta}^I$.

3.4 Simulation Results

We have conducted extensive simulation studies to examine the finite sample properties of the proposed estimators. To generate the recurrent event times, we have adopted Jahn-Eimermacher et al. (2015)'s algorithm. We consider the following random-effects intensity model to generate the recurrent events:

$$\lambda(t \mid Z(t), \vartheta) = \vartheta e^{\beta_0 Z(t)}\lambda_0(t) \tag{3.6}$$

where ϑ is an unobserved unit-mean positive random variable that is independent of Z. We derived the functions, $\tilde{\Lambda}_t$ and $\tilde{\Lambda}_t^{-1}$, from the intensity process using the formula $\tilde{\Lambda}_t(u \mid Z(t),\vartheta) = \Lambda(u+t \mid Z(t),\vartheta) - \Lambda(t \mid Z(t),\vartheta)$. Independent random numbers, a_i, are drawn from a uniform distribution on [0,1]. The following recursive algorithm is applied to obtain the recurrent event data for individual i: $t_{i1} = \Lambda^{-1}(-log(a_{i1}) \mid Z_i(t_{i1}), \vartheta)$, $t_{i,j+1} = t_{i,j} + \tilde{\Lambda}^{-1}_{t_{i,j}}(-log(a_{i,j+1}) \mid Z_i(t_{i,j+1}), \vartheta)$, $j = 1, 2, \ldots, n_i$. We assumed that ϑ has a Gamma distribution with mean 1 and variance σ^2. We considered binary covariate generated from Bernoulli (0.5) and continuous covariate from Uniform (0,1). We considered different cohort sizes: 1000, 2000 and 4000 and the number of simulated data sets being considered is 1000. We considered σ^2 such that mean recurrence is 3 for those who had at least one event. We considered β_0 to be 0.5 or 0.

Table 3.1 summarizes simulation results for situations that the proportion of individuals who experienced at least one event was low (5%, 10%, 20%). For the case-cohort sampling, the sub-cohort sampling proportion was 25% and all cases were sampled. The simulation results show that the coefficient estimates are approximately unbiased for all the situations we considered. From Table 3.1, we note that the proposed estimated standard errors provide good estimates of the true variability of $\hat{\beta}^I$ in all the situations except for when both the full cohort and the event rate are very small. As the cohort size increases, the performance of the estimated standard error improves. The variance of $\hat{\beta}^I$ decreases as the cohort size and/or the event proportion increases. The coverage rate of the nominal 95% confidence intervals using the proposed method is in the $92 - 95\%$ range in all the situations considered except when the event rate along with the cohort size are small. As the cohort size or event rate increases, the 95% confidence interval coverage rate improves.

Table 3.1: Summary of Simulation Results of $\hat{\beta}^I$ for Multiplicative Model

Z	β_0	Cohort Size	Event proportion	Bias	Model Std. Error	Empirical Std. Dev.	Coverage
Bernoulli(0.5)	0.5	1000	0.05	-0.003	0.474	0.535	0.85
			0.10	-0.031	0.342	0.323	0.91
			0.20	-0.034	0.225	0.228	0.94
		2000	0.05	0.002	0.344	0.349	0.90
			0.10	-0.023	0.227	0.231	0.93
			0.20	-0.013	0.161	0.165	0.94
		4000	0.05	-0.02	0.245	0.245	0.93

			0.10	-0.025	0.161	0.153	0.95
			0.20	-0.023	0.113	0.116	0.93
	0	1000	0.05	-0.0006	0.360	0.368	0.90
			0.10	-0.0002	0.287	0.288	0.94
			0.20	0.008	0.219	0.222	0.94
		2000	0.05	0.008	0.256	0.261	0.92
			0.10	-0.005	0.205	0.208	0.93
			0.20	0.003	0.154	0.154	0.95
		4000	0.05	0.0146	0.181	0.188	0.94
			0.10	0.0013	0.145	0.146	0.95
			0.20	-0.0037	0.109	0.110	0.94
Unif(0,1)	0.5	1000	0.05	-0.05	0.746	0.777	0.86
			0.10	-0.022	0.566	0.589	0.90
			0.20	-0.045	0.392	0.395	0.94
		2000	0.05	-0.036	0.531	0.547	0.90
			0.10	-0.02	0.404	0.416	0.92
			0.20	-0.011	0.279	0.281	0.94
		4000	0.05	0.008	0.381	0.399	0.91
			0.10	-0.025	0.279	0.274	0.95
			0.20	-0.023	0.197	0.200	0.94
	0	1000	0.05	-0.006	0.627	0.636	0.89
			0.10	-0.006	0.509	0.507	0.92
			0.20	0.006	0.382	0.380	0.94
		2000	0.05	0.005	0.443	0.466	0.92
			0.10	-0.007	0.357	0.362	0.92
			0.20	0.0009	0.266	0.263	0.94
		4000	0.05	0.021	0.312	0.322	0.93
			0.10	0.004	0.255	0.254	0.94

			0.20	-0.002	0.189	0.189	0.95

Table 3.2 summarizes the simulation results for situations when the proportion of events is not low (40%, 30% and 25%). We considered generalized case-cohort sampling. The sub-cohort sampling proportion is 10% and sampling proportion for the cases outside the sub-cohort is also 10%. The simulation results show that the coefficient estimates are approximately unbiased, the proposed variance estimator is close to the empirical variance, and the 95% confidence interval coverage is close to the nominal level for all the situations considered.

Table 3.2: Summary of Simulation Results of $\hat{\beta}^{II}$ for Multiplicative Model

Z	β_0	Cohort Size	Event proportion	Bias	Model Std. Error	Empirical Std. Dev.	Coverage
Bernoulli(0.5)	0.5	1000	0.25	-0.017	0.5	0.548	0.93
			0.30	0.018	0.429	0.460	0.92
			0.40	0.014	0.358	0.370	0.93
		2000	0.25	0.005	0.357	0.381	0.93
			0.30	0.01	0.305	0.327	0.93
			0.40	0.001	0.253	0.263	0.94
		4000	0.25	-0.015	0.253	0.254	0.95
			0.30	0.008	0.217	0.219	0.95
			0.40	-0.003	0.178	0.179	0.96
	0	1000	0.25	-0.004	0.5	0.54	0.93
			0.30	0.004	0.434	0.439	0.94
			0.40	-0.002	0.348	0.38	0.93
		2000	0.25	-0.009	0.358	0.373	0.93
			0.30	-0.007	0.307	0.32	0.94
			0.40	0.001	0.249	0.254	0.94
		4000	0.25	0.007	0.253	0.259	0.95

			0.30	-0.004	0.218	0.222	0.94
			0.40	0.006	0.177	0.178	0.95
Unif(0, 1)	0.5	1000	0.25	0.004	0.858	0.966	0.92
			0.30	-0.003	0.736	0.806	0.93
			0.40	0.003	0.610	0.625	0.94
		2000	0.25	-0.026	0.62	0.64	0.94
			0.30	0.0099	0.53	0.55	0.94
			0.40	-0.016	0.436	0.459	0.93
		4000	0.25	-0.01	0.442	0.457	0.94
			0.30	0.0004	0.376	0.391	0.94
			0.40	-0.004	0.312	0.319	0.95
	0	1000	0.25	-0.018	0.864	0.923	0.92
			0.30	0.004	0.742	0.77	0.94
			0.40	0.008	0.60	0.64	0.92
		2000	0.25	-0.015	0.620	0.664	0.92
			0.30	-0.0085	0.531	0.539	0.94
			0.40	-0.0027	0.430	0.431	0.94
		4000	0.25	0.012	0.437	0.457	0.94
			0.30	-0.017	0.378	0.388	0.94
			0.40	0.003	0.305	0.310	0.94

3.5 Application to ALRI data

A doubly-blinded placebo-controlled community trial was conducted in a cohort of 1207 children in northeastern Brazil, who were followed up from December 1990 to December 1991 (Barreto et al. 1994, Amorim and Cai 2015). The primary aim of the original trial was to study the effect of high doses of Vitamin A on diarrhea and acute-lower-respiratory-tract-infections (ALRI). The age of the children at baseline ranged from 6 to 48 months. They were randomly assigned to vitamin A supplement or placebo. For the purpose of our analysis, 1190 children were eligible: sixteen subjects had missing information on one of the variables of interest and one child was shifted to Vitamin A from Placebo. Daily information on respiratory rates were collected (3 times a week) with a recall period of 48 to 72 hours. An episode of ALRI was defined as cough plus a respiratory rate of 50 breaths per min or higher for children under 12 months of age, and 40 breaths per min or higher for older children(Barreto et al. 1994, Amorim and Cai 2015). At these visits, if the child reported cough, then the respiratory rates were measured twice. A new episode of ALRI was defined if there was an interval of 14 or more days (Barreto et al. 1994). Censoring occurred when children were lost to follow-up or the study reached its end. The number of children who had at least one event was 185 with event proportion of 15.37%.

We constructed a case-cohort sample based on this cohort study to illustrate our proposed method. We consider the indicator variable for children ever having measles to be the expensive variable which is only available for the case-cohort sample. We are interested in studying the effect of past occurrence of measles on ALRI. We considered the probability of sub-cohort selection to be 0.2. The total sample size for the case-cohort data is 376 with 238 in the subcohort. The following covariates were considered in the analysis: treatment group (vitamin A vs placebo), child's gender (male vs female), age at baseline (dichotomized based on whether the child is older than 12 months or not), an indicator for the presence of a toilet in the child's house (which is considered as a proxy for hygienic habits), and the indicator for children experiencing measles in their lifetime (based on the information provided by their parents). For the analysis, we considered placebo, female gender, child's age $\leq$ 12 months, no presence of toilet at home, and never experiencing measles as the reference groups. Table 3.3 summarizes the distribution of the baseline variables in the subcohort and the full cohort.

Table 3.3: Baseline Characteristics of the Acute Lower-Respiratory-Tract Infections study

Variables	Full Cohort (n = 1190)	Subcohort ($\tilde{n}$ = 238)
Treatment (Vit. A: 1 vs. Placebo: 0)	0.5017	0.4790
Gender (Boys: 1, Girls: 0)	0.5244	0.5462
Age ($\leq$ 1 yr: 1, $>$ 1 yr: 0)	0.1311	0.1597
Toilet at home (Yes: 1)	0.7361	0.7479

Table 3.3 shows that the distribution of the variables in the subcohort are very similar to that in the full cohort. We applied our proposed method to the case-cohort sample.

Table 3.4: Estimates and standard errors for the multiplicative rates model with data from case-cohort sample from the ALRI study

Effects	Proposed method	
	Case-Cohort ($\tilde{\alpha}$ = 0.2)	
	Estimate	SE
Treatment (Reference: Placebo)	0.0534	0.1995
Gender (Ref: Female)	-0.0056	0.1965
Age (Ref: $>$ 12 months)	1.6859	0.3102
Toilet at home(Ref: Absence)	-0.9033	0.1759
Measles Indicator (Ref: Never)	0.0853	0.3395

Table 3.4 provides results from the model adjusting for covariates. Dichotomized age and presence of toilet at home are significant predictors of the recurrence of ALRI among young children, adjusting for the other variables in the model. Among the other variables, from Table 3.4, high doses of Vitamin A, gender and prior measles indicator are not significantly associated with recurrence of ALRI. Further, controlling for all other variables, children in households with toilets are at a 0.595 times lower risk, of developing ALRI, than the children

living in household without a toilet. Similarly, the risk of ALRI recurrence among children who are younger than 12 months are 5.397 times that of children who are older than 12 months.

In the above analysis, we defined a case as an individual who has experienced at least one event. One can use different definitions of a 'case' for recurrent events, based on the number of events. It is of interest to examine the performance of the estimates based on the different case definition for sampling. We conducted some simulations by sampling from the ALRI study. Sub-cohort proportion was selected as 20% and the number of simulations considered is 1000. We have considered three definitions for 'case': subjects who experienced at least one event, at least two events, or at least three events. Table 3.5 shows the results from the simulations for the different 'case' definitions. We note that the standard errors were the lowest with the at least one case definition. This could be due to the smaller supplementary case samples with the other definitions.

Table 3.5: Estimates and standard errors for different definitions of case for case-cohort sample from the ALRI study

Effects	Case Definition					
	$\geq$ One Event		$\geq$ Two events		$\geq$ Three events	
	Estimate	SE	Estimate	SE	Estimate	SE
Treatment	-0.0272	0.197	-0.0299	0.202	-0.0268	0.205
Gender	-0.118	0.196	-0.121	0.202	-0.132	0.205
Age	1.697	0.290	1.701	0.320	1.702	0.336
Toilet at home	-0.691	0.182	-0.695	0.186	-0.698	0.188
Prior Measles	0.0538	0.339	0.0488	0.361	0.0458	0.365

3.6 Final Remarks

This article proposes methods of fitting marginal multiplicative rates model for both the case-cohort and the generalized case-cohort designs with time-varying weights. The proposed estimators are natural generalizations of the full cohort estimators and has easy interpretation. The proposed estimators are consistent and asymptotically distributed. They perform well in finite samples. Throughout this article, for the case-cohort analysis, we have assumed simple random sampling for the subcohort and all individuals who have experienced the event. Bernoulli sampling of the subcohort is another sampling scheme that can be considered.

In our approach, we do not use all the covariate information that are available for the entire cohort. Developing a more general method taking advantage of those covariates information to improve the efficiency of the estimators is worthy of future research. We have considered event proportion (individuals with atleast one event) around $25\% - 50\%$ to be corresponding to a relatively common event. It can be noted that when the event proportion is $> 50\%$, then simple random sample of the entire cohort should yield sufficient data to achieve the desirable power. Extension of this approach to additive rates model and for multiple types of recurrent events are under further investigation.

CHAPTER 4: ADDITIVE RATES MODEL FOR RECURRENT EVENTS WITH CASE-COHORT DATA

4.1 Introduction

The additive and multiplicative rates models are two different frameworks to study the association between various risk factors and time to event data. Most modern survival analyses focus on the multiplicative model for relative risk using Cox's (1972) proportional hazards model, mainly because of the availability of computer programs and the easy interpretability of the results. However, investigators are sometimes interested in the risk difference because it translates directly into the number of event cases that can be avoided by eliminating the exposure (Kulich and Lin 2000). As a result, depending on the interest, an additive rates model can be considered instead of a multiplicative rates model, especially when the multiplicative model assumption does not hold.

Aalen (1980) first proposed the additive hazards model as an alternative to the Cox (1972) proportional hazards model. The author noted that since the covariate effects are not assumed to be proportional, the additive model is capable of providing information about the temporal influence of each covariate that is not available from the Cox model. Several authors have studied theoretical developments of such models for univariate failure times. These include semi-parametric estimation procedure by Lin and Ying (1994), application to current status data (Lin et al. 1998), modeling in presence of auxiliary covariate information (Jiang and Zhou 2007), among others. Yin and Cai (2004) extended the idea to multivariate failure time data. Schaubel et al. (2006) and He et al. (2013) proposed semi-parametric additive rates models for clustered recurrent event data.

All the aforementioned articles assume that data are fully available for all the members in the cohort. However, in epidemiology studies, cost constraints often make it infeasible to collect information on all the individuals in the cohort, especially when some covariates are expensive to measure. Usually, the majority of the cost comes from the assembling of the covariate data and most of the information on the event-free individuals in the cohort are redundant if the event rate is low. The case-cohort sampling scheme (Prentice 1986) is one of the several study designs that have been proposed to reduce cost in large epidemiological cohort studies. The key principle of this design is to obtain covariate information on only a random subset of the cohort (which is called the sub-cohort) and all individuals in the cohort who experience the event of interest (cases). This leads to the case-cohort studies being considered as an appealing alternative to full-cohort studies for large-scale epidemiological studies where the event is rare or the covariates are quite expensive to measure. When the event rate is not rare, due to financial limitation or technical difficulties, measuring important covariates in the entire cohort may be impractical. In such situation, it is advisable to consider cost-effective sampling designs for selecting subjects from the cohort to reduce the study cost, without losing much of the efficiency. One such study design is the generalized case-cohort design (Kulich and Lin 2004) where covariate information is collected for the sub-cohort and a random sample of the remaining cases. Kulich and Lin (2000), Kang et al. (2013), Yu et al. (2014), Dong et al. (2014), Kim et al. (2016), Cao and Yu (2017) among others developed the semi-parametric inference of additive hazards models for two-phase designs, more specifically case-cohort and generalized case-cohort studies.

Despite the abundance of methodologies in the time-to-event literature for analyzing single event case-cohort data, methods to analyze recurrent events under the case-cohort design have been limited. In this paper, we look at the additive rates model for modeling recurrent events under the generalized case-cohort design, in order to reduce the cost of collecting expensive covariate information. Our illustrating example is a randomized double-blind clinical trial which was carried out by Genentech in 1992 to study the effect of rhDNase on pulmonary exacerbations in patients with cystic fibrosis (Therneau and Hamilton 1997). Six hundred and sixty-five patients were followed up for 169 days and data on multiple exacerbations were collected during this period with about 36.54% of the individuals (rhDNase and Placebo combined) experiencing at least one such event. We will use this data to illustrate our proposed method under generalized case-cohort design.

In the next section, we propose a weighted estimating equation approach for estimating the parameters in the marginal hazard regression model for recurrent event data from case-cohort studies. The asymptotic properties of the proposed estimators are studied in section 4.3. The outlines of the proofs of the asymptotic properties are given in the appendix. In section 4.4, simulation studies were considered to examine the finite sample properties. In section 4.5, we illustrated the proposed method on a case-cohort study based on the rhDNase clinical trial. We conclude the paper with a brief discussion.

4.2 Model and Estimation

Suppose there are n independent individuals in the cohort. Let $N_i^*(t)$ be the number of recurrent events for the individual i over the time interval $[0, t)$, C_i is the censoring time. $Z_i(.)$ $= (Z_i^E(.)', Z_i^C(.)')'$ is the p-dimensional covariate of interest for the individual i, where $Z_i^E(.)$ is the set of expensive-to-measure variables and $Z_i^C(.)$ is the set of all other covariates. Let T_{ij}^* denote the j-th recurrent event time for the individual i. The observed time is $T_{ij} = T_{ij}^* \wedge C_i$, $j = 1, 2, \ldots, n_i + 1$, where n_i is the number of events that are observed for individual i, and $N = \sum_{i=1}^n n_i$ is the total number of observed events. Let $Y_i(t) = \mathbf{1}(C_i \geq t)$, $N_i(t) = N_i^*(t \wedge C_i)$ $= \sum_{j=1}^{n_i} \mathbf{I}(T_{ij} \leq t) = \sum_{j=1}^{n_i} \mathbf{I}(T_{ij}^* \leq t) Y_i(t)$, $\Delta_i = 1 - \mathbf{I}(\min_j(T_{ij}^*) \geq C_i)$ which is the indicator that individual i experienced at least one event, and τ denote the study ending time. The rate function for an individual is denoted as $E\left(dN^*(t) \mid Z(t)\right) = d\mu_z(t)$. We assume the following additive rate model :

$$d\mu_Z(t) = d\mu_0(t) + \left(\theta_0' Z^E(t) + \gamma_0' Z^C(t)\right) dt \Rightarrow \mu_Z(t) = \mu_0(t) + \int_0^t \left\{\theta_0' Z^E(u) + \gamma_0' Z^C(u)\right\} du, \tag{4.1}$$

where $\mu_0(.)$ is an unspecified continuous baseline mean function and θ_0 and γ_0 are the vectors of unknown parameters. Denoting $\beta_0 = (\theta_0', \gamma_0')'$, we can rewrite the rates model as $d\mu_Z(t) =$ $\mu_0(t) + \beta_0' Z(t) dt$ and the mean function is given by $\mu_Z(u) = \mu_0(u) + \int_0^u \beta_0' Z(t) dt$ for all $u \in [0, \tau]$. Note that the covariates are allowed to be time-dependent. We assume that the possibly time-dependent covariates are external (Kalbfleisch and Prentice 2002), i.e., they are not affected by the recurrent event process. If the data were complete, the true regression parameter β_0 can be estimated from the following estimating equations (Schaubel et al. 2006):

$$\sum_{i=1}^{n}\int_0^t Y_i(u)dM_i(u) = 0 \tag{4.2}$$
$$\sum_{i=1}^{n}\int_0^t Y_i(u)Z_i(u)dM_i(u) = 0, \tag{4.3}$$

where $M_i(t) = N_i(t) - \int_0^t Y_i(u)\left\{d\mu_0(u) + \beta_0' Z_i(u)du\right\}$. Combining the above two equations, we have

$$U(\beta) = \sum_{i=1}^{n}\int_0^{\tau}\left(Z_i(t) - \bar{Z}(t)\right)\left[dN_i(t) - Y_i(t)\beta_0' Z_i(t)dt\right] = 0, \tag{4.4}$$

where $\bar{Z}(t) = \frac{\sum_{i=1}^{n} Y_i(t)Z_i(t)}{\sum_{i=1}^{n} Y_i(t)}$. The above estimating equation can be solved explicitly to obtain an estimator of β given by $\hat{\beta} = \left[\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left\{Z_i(t) - \bar{Z}(t)\right\}^{\otimes 2} dt\right]^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{Z_i(t) -\bar{Z}(t)\right\} \times dN_i(t)\Big)$. However, for case-cohort data, since we do not have information on all the individuals, the estimating cannot be applied directly.

4.2.1 Case-cohort study design for recurrent events

In this subsection, we introduce two sampling schemes for the recurrent event data. We call an individual a case if the individual experienced at least one event and an individual a non-case if the individual did not have any event during the study period. The first sampling scheme deals with the situation that the event is not common in the population. An example of such situation is a doubly-blind, placebo-controlled community trial conducted in northeastern Brazil among young children to evaluate the effect of high doses of vitamin A on acute-lower-respiratory-tract infection (ALRI)(Barreto et al. 1994). The event of interest occurs in only about 15% of the cohort. In this case, we draw a random sample from the full cohort and supplement that with all the cases. We call this sampling scheme the original case-cohort design with recurrent events. The second sampling scheme is for the situation described in Section 4.1. In this case, the event is relatively common and we cannot afford to sample all individuals with events. Under such circumstance, we propose to sample only a fraction of those who have events for the case-cohort sampling. We call this sampling scheme the generalized case-cohort design with recurrent events.

4.2.2 Estimation under the original case-cohort design

Under the case-cohort sampling, we select a sub-cohort from the entire cohort by simple random sampling. Let ξ_i denote the indicator function for individual i being selected into the subcohort; $\tilde{\alpha} = \frac{\tilde{n}}{n}$ is the subcohort proportion where $\tilde{n}$ is the number of individuals selected in the sub-cohort and n is the number of individuals in the full cohort. The observable information for individual i is $\{\mathbf{T}_i, \Delta_i, \xi_i, Z_i^C(t), Z_i^E(t), t \in [0,\tau]\}$ if individual i is in the case-cohort sample. In other words, we have $\{\mathbf{T}_i, \Delta_i, \xi_i, Z_i(t), 0 \leq t \leq \tau\}$ if $\Delta_i = 1$ or $\xi_i = 1$ and $\{\mathbf{T}_i, \Delta_i, \xi_i, Z_i^C(t), t \in [0,\tau]\}$ when $\Delta_i = 0$ and $\xi_i = 0$. We considered the following estimating equation to estimate β_0.

$$U^I(\beta) = \sum_{i=1}^{n} \int_0^{\tau} w_i^I(t) \left\{ Z_i(t) - \bar{Z}^I(t) \right\} \left[dN_i(t) - Y_i(t)\beta' Z_i(t) dt \right] = 0, \tag{4.5}$$

where $\bar{Z}^I(t) = \frac{\sum_{i=1}^n w_i^I(t) Y_i(t) Z_i(t)}{\sum_{i=1}^n w_i^I(t) Y_i(t)}$, $w_i^I(t) = \Delta_i + (1-\Delta_i)\frac{\xi_i}{\hat{\alpha}(t)}$, where $\hat{\alpha}(t) = \frac{\sum_i (1-\Delta_i)\xi_i Y_i(t)}{\sum_i (1-\Delta_i) Y_i(t)}$ is the estimator of the true sampling parameter, α. The weight is 1 for all the cases and is $\hat{\alpha}(t)^{-1}$ for the non-cases in the sub-cohort. Similar idea for the weights was considered by Kalbfleisch and Lawless (1988) for a single event. They considered the time-invariant version of $\hat{\alpha}(t)$, which was given by $\tilde{\alpha}$. Borgan et al. (2000) used a similar idea for univariate failure time data from stratified case-cohort studies. We denote the solution to this equation by $\hat{\beta}^I$. Our proposed Breslow-Aalen type estimator of the baseline mean function is given by

$$\hat{\mu}_0(\hat{\beta}^I, t) = \int_0^t \frac{\sum_{i=1}^n w_i^I(u) \left[dN_i(u) - Y_i(u)(\hat{\beta}^I)' Z_i(u) du \right]}{\sum_{i=1}^n w_i^I(u) Y_i(u)}. \tag{4.6}$$

4.2.3 Estimation under the generalized case-cohort design with recurrent events

For the generalized case-cohort design, we sample a fraction of cases outside of the sub-cohort. Let η_i be an indicator of being sampled for individual i who is a case but outside the sub-cohort being sampled. Let $\tilde{q} = \frac{n_1^*}{n_1 - \tilde{n}_1}$ denote the sampling proportion for the additional cases, where n_1^*, n_1 and $\tilde{n}_1$ are the number of selected individuals who have experienced at least one event but are not in the subcohort, individuals who experienced at least one event in the full cohort and those who were in the subcohort respectively. Under this design, the covariate information is available for the subcohort members and the selected cases ($\eta_i = 1$). Hence, the observable information for individual i is $\{\mathbf{T}_i, \Delta_i, \xi_i, \eta_i, Z_i(t) : t \in [0, \tau]\}$ when $\xi_i = 1$ or $\eta_i = 1$ and $\{\mathbf{T}_i, \Delta_i, \xi_i, Z_i^C(t), t \in [0, \tau]\}$ if $\xi_i = 0$ and $\eta_i = 0$. Using the inverse of probability of being sampled as the weight, our proposed estimating equation for the generalized case-cohort sampling scheme is

$$U^{II}(\beta) = \sum_{i=1}^{n} \int_0^{\tau} w_i^{II}(t) \left\{ Z_i(t) - \bar{Z}^{II}(t) \right\} \left[dN_i(t) - Y_i(t)\beta' Z_i(t)dt \right] = 0, \tag{4.7}$$

where $\bar{Z}^{II}(t) = \frac{\sum_{i=1}^{n} w_i^{II}(t) Y_i(t) Z_i(t)}{\sum_{i=1}^{n} w_i^{II}(t) Y_i(t)}$. The weight function is given by $w_i^{II}(t) = \frac{(1-\Delta_i)\xi_i}{\hat{\alpha}(t)} + \Delta_i \xi_i + \frac{\Delta_i(1-\xi_i)\eta_i}{\hat{q}(t)}$ where $\hat{q}(t) = \frac{\sum_i \Delta_i(1-\xi_i)\eta_i Y_i(t)}{\sum_i \Delta_i(1-\xi_i) Y_i(t)}$ is the estimator of the true sampling parameter q. We denote the solution of this equation by $\hat{\beta}^{II}$. The Breslow-Aalen type estimator of the baseline mean function is

$$\hat{\mu}_0(\hat{\beta}^{II}, t) = \int_0^t \frac{\sum_{i=1}^{n} w_i^{II}(u) \left[dN_i(u) - Y_i(u)(\hat{\beta}^{II})' Z_i(u) du \right]}{\sum_{i=1}^{n} w_i^{II}(u) Y_i(u)}. \tag{4.8}$$

Throughout this article, we use superscript I or II depending on whether the quantity corresponds to the original case-cohort design or the generalized case-cohort design respectively.

4.3 Asymptotic properties

In this section, we investigate the asymptotic properties of the estimators. Define the following terms: $e(t) = \frac{E(Y_i(t) Z_i(t))}{E(Y_i(t))}$, $A = E\left[\int_0^{\tau} Y_i(t) \left\{ Z_i(t)^{\otimes 2} - e(t)^{\otimes 2} \right\} d\mu_0(t)\right]$, $\tilde{Z}_i(t) = Z_i(t) - e(t)$, $dM_i(t) = dN_i(t) - Y_i(t)\left[d\mu_0(t) + \beta_0 Z_i(t)\right]$, $M_{\tilde{Z},i}(\beta) = \int_0^{\tau} \tilde{Z}_i(t) dM_i(t)$. Further, the norms of a vector, a, is defined as $\|a\| = \max_i |a_i|$, for a matrix A by $\|A\| = \max_{i,j} \|A_{ij}\|$ and for a function f by $\|f\| = \sup_t |f(t)|$. The estimator under the classical case-cohort sampling scheme is a special case of the generalized case-cohort sampling scheme, so its asymptotic property is a special case of the generalized case-cohort sampling scheme. Hence, in the Appendix, we focus the proofs on the asymptotic properties of the estimators under the generalized case-cohort design, $\hat{\beta}^{II}$ and $\hat{\mu}_0^{II}(\hat{\beta}^{II}, t)$. The regularity conditions and the outline of the proofs are provided in the Appendix. The asymptotic properties are summarized in the following theorems.

Theorem 3. *Under the regularity conditions in the Appendix, for $k = I$ or II, $\hat{\beta}^k$ is a consistent estimator of β_0. $n^{1/2}\{\hat{\beta}^k - \beta_0\}$ converges to a Gaussian distribution with mean zero and variance given by*

$$\Sigma^k(\beta_0) = A^{-1}\left[Q(\beta_0) + \frac{1-\alpha}{\alpha}V^I(\beta_0) + \boldsymbol{I}(k = II)(1-\alpha)\frac{1-q}{q}P(\Delta_1 = 1)V^{II}(\beta_0)\right]A^{-1},$$

where

$$A = E\left[\int_0^\tau Y_i(t)\left\{Z_i(t)^{\otimes 2} - e(t)^{\otimes 2}\right\}dt\right], \quad Q(\beta) = E\left(M_{\tilde{Z},1}(\beta)\right)^{\otimes 2},$$

$$V^I(\beta_0) = E\left((1-\Delta_1)\int_0^\tau\left[dR_1(\beta_0,t) - \frac{Y_1(t)E((1-\Delta_1)dR_1(\beta_0,t))}{E((1-\Delta_1)Y_1(t))}\right]\right)^{\otimes 2},$$

$$V^{II}(\beta_0) = var\left(M_{\tilde{Z},i}(t) - \int_0^\tau \frac{Y_i(t)E(dM_{\tilde{Z},1}(t) \mid \Delta_1 = 1, \xi_1 = 0)}{E(Y_1(t) \mid \Delta_1 = 1)} \mid \Delta_i = 1, \xi_i = 0\right),$$

$$dR_i(\beta, t) = Y_i(t)\tilde{Z}_i(t)\{d\mu_0(t) + \beta_0 Z_i(t)dt\}.$$

Each of these terms, A, $Q(\beta_0)$, $V^I(\beta_0)$ and $V^{II}(\beta_0)$ can be estimated respectively by their sample counterparts, $\hat{A}$, $\hat{Q}(\hat{\beta}^k)$ and $\hat{V}^I(\hat{\beta}^k)$, where $\hat{A} = -\frac{1}{n}\frac{\partial U^k(\beta)}{\partial\beta}$, $\hat{Q}(\beta) = \frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\tilde{\alpha}}\left(M_{\tilde{Z},i}(\beta)^{\otimes 2}\right)$, $\hat{M}_{\tilde{Z},i}(\beta) = \int_0^\tau\{Z_i(t) - \bar{Z}^k(t)\}d\hat{M}_i^k(\beta,t)$, $d\hat{M}_i^k(\beta,t) = dN_i(t) - Y_i(t)\{d\hat{\mu}_0(\beta,t) + \beta' Z_i(t)dt\}$, $\hat{\bar{Z}}^k(\beta,t) = \frac{\sum_{i=1}^n \hat{w_i}^k(t)Z_i(t)Y_i(t)}{\sum_{i=1}^n \hat{w_i}^k(t)Y_i(t)}$, $d\hat{R}_i^{II}(\beta,t) = \left\{Z_i(t) - \bar{Z}^{II}(t)\right\}Y_i(t)\left\{d\hat{\mu}_0^{II}(\beta,t) + \beta' Z_i(t)dt\right\}$, $d\hat{\mu}_0^{II}(\beta,t) = \frac{\sum_i w_i^{II}(t)\{dN_i(t) - Y_i(t)\beta' Z_i(t)dt\}}{\sum_j w_j^{II}(t)Y_j(t)}$, $\hat{E}\left(\Delta_1(1-\xi_1)dR_1(\beta,t)\right) = \frac{1}{n}\sum_{i=1}^n \frac{\eta_i}{\tilde{q}}\Delta_i(1-\xi_i)d\hat{R}_i^{II}(\beta,t)$, $\hat{E}\left(\Delta_1(1-\xi_1)Y_1(\beta,t)\right) = \frac{1}{n}\sum_{i=1}^n \Delta_i(1-\xi_i)Y_i(t)$. Based on these, we have the following estimates of the second and third terms of the variance as

$$\hat{V}^I(\beta) = \frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\tilde{\alpha}}\left[(1-\Delta_i)\sum_{j=1}^n\sum_{k=1}^{n_j}\left\{\hat{R}_i^{II}(\beta, T_{jk}) - \frac{Y_i(T_{jk})\hat{E}\left((1-\Delta_1)dR_1(\beta,T_{jk})\right)}{\hat{E}\left((1-\Delta_1)Y_1(T_{jk})\right)}\right\}\right]^{\otimes 2},$$

$$\hat{V}^{II}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i}{\tilde{q}}\left[\Delta_i(1-\xi_i)\left\{\hat{M}_{\tilde{Z},i}(\beta) - \sum_{k=1}^{n_i}\int_0^{T_{ik}}\frac{\hat{E}(dM_{\tilde{Z},1}(t)\mid\Delta_1=1,\xi_1=0)}{\hat{E}(Y_1(t)\mid\Delta_1=1)}\right\}\right]^{\otimes 2}.$$

To obtain the asymptotic distribution of the baseline mean function $\hat{\mu}_0(t)$, we need to first define the metric space, $\mathcal{D}[0,\tau]$, consisting of right continuous functions f(t) with left-hand limits and $f:[0,\tau]\to R$. The metric for this space is defined by $d(f,g) = \sup_{t\in[0,\tau]}\{\mid f(t)-g(t)\mid\}$ for $f(t), g(t)\in\mathcal{D}[0,\tau]$. The following theorem summarizes the asymptotic properties of $\hat{\mu}_0^k(t)$.

Theorem 4. *Under the regularity conditions in Appendix, for $k = I$ or II, $\hat{\mu}_0^k(\beta,t)$ converges in probability to $\mu_0(t)$ uniformly in $t\in[0,\tau]$. Further, defining $W_n(t) = \hat{\mu}_0^k(\hat{\beta}^k,t) - \mu_0(t)$, $n^{1/2}W_n(t)$ converges to a Gaussian distribution with mean zero. The variance-covariance function between $W_n(t)$ and $W_n(s)$ is given by*

$$\begin{aligned}\phi^k(t,s)(\beta_0) = E\left(\nu_i(\beta_0,t)\nu_i(\beta_0,s)\right) + \tfrac{1-\alpha}{\alpha}E\left(\psi_i^I(\beta_0,t)\psi_i^I(\beta_0,s)\right)\\ +\boldsymbol{I}(k=II)\tfrac{1-q}{q}E\left(\psi_i^{II}(\beta_0,t)\psi_i^{II}(\beta_0,s)\right),\end{aligned}$$

where

$$\nu_i(\beta_0,t) = \left[r(t)'A^{-1}\int_0^{\tau}(Z_i(t)-e(t))\,dM_i(t) + \int_0^t\frac{dM_i(u)}{E(Y_1(u))}\right],$$

$$\begin{aligned}\psi_i^I(\beta_0,t) = (1-\Delta_i)\Bigg[r(t)'A^{-1}\left(R_i(\beta_0) - \int_0^{\tau}\frac{Y_i(t)E((1-\Delta_1)dR_i(\beta_0,t))}{E((1-\Delta_1)Y_i(t))}\right)\\ + \int_0^t\frac{1}{E\left(Y_1(u)\right)}Y_i(u)\left[\beta_0'Z_i(u) - \frac{E\left(\{1-\Delta_1\}Y_1(u)\beta_0'Z_1(u)\right)}{E((1-\Delta_1)Y_1(u))}\right]du\Bigg]\end{aligned}$$

and

$$\begin{aligned}\psi_i^{II}(\beta_0,t) = \Delta_i(1-\xi_i)r(t)'A(\beta_0)^{-1}\int_0^{\tau}\left\{dM_{\tilde{Z},i}(u) - Y_i(u)\frac{E\left[dM_{\tilde{Z},1}(u)\mid\Delta_1=1,\xi_1=0\right]}{E(Y_1(u)\mid\Delta_1=1)}\right\}\\ + \Delta_i(1-\xi_i)\int_0^t\frac{1}{E\left(Y_1(u)\right)}\left\{dM_i(u) - Y_i(u)\frac{E\left[dM_1(u)\mid\Delta_1=1,\xi_1=0\right]}{E(Y_1(u)\mid\Delta_1=1)}\right\}\end{aligned}$$

Each of these terms can be consistently estimated by their sample counterparts. $\hat{\phi}^{II}(t,s)(\beta) = \sum_{i=1}^{n} \frac{\xi_i}{\tilde{\alpha}} \left(\hat{\nu}_i(\beta,t)\hat{\nu}_i(\beta,s)\right) + \frac{1-\tilde{\alpha}^{-1}}{n} \sum_{i=1}^{n} \frac{\xi_i}{\tilde{\alpha}} \left(\hat{\psi}_i^I(\beta,t)\hat{\psi}_i^I(\beta,s)\right) + \frac{1-\tilde{q}^{-1}}{n} \sum_{i=1}^{n} \frac{\eta_i}{\tilde{q}} \left(\hat{\psi}_i^{II}(\beta,t) \times \hat{\psi}_i^{II}(\beta,s)\right)$, $\hat{\nu}_i(\beta,t) = \hat{r}^I(\beta,t)'\hat{A}^{-1}\hat{M}_{\tilde{Z},i}(\beta,t) + \int_0^t \{\hat{E}(Y_1(u))\}^{-1} d\hat{M}_i(u)$, $\hat{\psi}_i^I(\beta,t) = \Big[\hat{r}^I(\beta,t)'\hat{A}^{-1} (1-\Delta_i) \times \Big\{\hat{R}_i^{II}(\beta,u) - \int_0^\tau \frac{Y_i(u)\hat{E}\left((1-\Delta_i)Y_i(u)\tilde{Z}_i(u)\right)d\hat{\mu_0}^I(u)}{\hat{E}((1-\Delta_i)Y_i(u))} - \int_0^\tau \frac{Y_i(u)\hat{E}\left((1-\Delta_i)Y_i(u)\tilde{Z}_i(u)\right)\beta_0' Z_i(u)du}{\hat{E}((1-\Delta_i)Y_i(u))}\Big\} + (1-\Delta_i)\int_0^t \left(\beta' Z_i(u) - \frac{\hat{E}((1-\Delta_i)Y_i(u)\beta' Z_i(u))}{\hat{E}((1-\Delta_i)Y_i(u))}\right) \times \frac{Y_i(u)d\mu_0(u)}{\hat{E}(Y_i(u))}\Big]$, $\hat{\psi}_i^{II}(\beta,t) = \Delta_i(1-\xi_i)\Big[\hat{r}^I(\beta,t)'\hat{A}^{-1} \int_0^\tau \Big\{d\hat{M}_{\tilde{Z},i}(u) - Y_i(u)\frac{\hat{E}\left[dM_{\tilde{Z},1}(u)|\Delta_1=1,\xi_1=0\right]}{\hat{E}(Y_1(u)|\Delta_1=1)}\Big\} + \int_0^t \frac{1}{\hat{E}(Y_1(u))}\Big\{d\hat{M}_i(u) - Y_i(u)\frac{\hat{E}[dM_1(u)|\Delta_1=1,\xi_1=0]}{\hat{E}(Y_1(u)|\Delta_1=1)}\Big\}\Big]$, $\hat{E}((1-\Delta_i)Y_i(u)\beta' Z_i(u)) = \frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\xi_i}{\tilde{\alpha}} Y_i(t)\beta' Z_i(t)$, $\hat{E}(Y_i(u)) = \frac{1}{n}\sum_{i=1}^{n} Y_i(u)$, $\hat{r}^I(\beta,t) = -\int_0^\tau \bar{Z}^{II}(t)dt$, $\hat{E}(\Delta_i(1-\xi_i)Y_i(u)\beta' Z_i(u)) = \frac{1}{n}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\frac{\eta_i}{\tilde{q}} Y_i(t)\beta' Z_i(t)$.

4.4 Simulation Studies

We conducted simulation studies to examine the finite sample properties of the proposed estimators. For subject i, the $(j+1)$-th event time is generated by

$$T_{i,j+1} = T_{i,j} - \frac{log(1-U_{i,j+1})}{Q_i + m_0 + \beta_0 Z_i},$$

where m_0 is the baseline rate, β_0 is the unknown parameter, $U_{i,j+1}$ are independent Uniform(0,1) variates and $T_{i,0} = 0$. The frailty variable, Q_i, induces positive correlation among the within-subject events. It follows a Gamma distribution with mean $E(Q_i) = 1$ and variance σ_Q^2. We consider both discrete and continuous Z for the simulations. We considered binary covariate generated from Bernoulli (0.5) and continuous covariate from Uniform (0,1). We considered different cohort sizes: 1000, 2000 and 4000 and the number of simulation repetition is 1000. We considered σ^2 such that mean recurrence is 3 for those who had at least one event. We considered β_0 to be 0.5 or 0.

Table 4.1 summarizes simulation results for situations that the proportion of individuals who experienced at least one event was low (5%, 10%, 20%). For the case-cohort sampling, the sub-cohort sampling proportion was 25% and all cases were sampled. The simulation results show that the coefficient estimates are approximately unbiased for all the situations we considered. From Table 4.1, we note that the proposed estimated standard errors provide good estimates of the true variability of $\hat{\beta}^I$ in all the situations except for when both the full cohort and the event rate are very small. As the cohort size increases, the performance of the estimated standard error improves. The variance of $\hat{\beta}^I$ decreases as the cohort size and/or the event proportion increases. The coverage rate of the nominal 95% confidence intervals using the proposed method is in the $92 - 95\%$ range in all the situations considered except when both the event rate and the cohort size are small. As the cohort size or event rate increases, the 95% confidence interval coverage rate improves. Specifically, when the event rate is at 5%, the confidence interval coverage is around 80% when the cohort size is 1000 (case cohort size = 288). However, the performance improves as the cohort size or event rate increases. Specifically, for continuous covariate, the confidence interval rate is close to the nominal level when the cohort size is 6000, 4000 or 2000 for event rate of 5% (case cohort size is 1725, 1300 or 800), 10% and 20% respectively. The corresponding required cohort size is smaller for discrete covariate.

Table 4.1: Summary of Simulation Results of $\hat{\beta}^I$ for Additive Model

Z	β_0	Cohort Size	Event proportion	Bias	Model Std. Error	Empirical Std. Dev.	Coverage
Bernoulli(0.5)	0	1000	0.05	0.008	0.712	0.729	0.822
			0.10	0.018	0.555	0.556	0.905
			0.20	-0.011	0.305	0.301	0.928
		2000	0.05	0.020	0.499	0.508	0.872
			0.10	-0.005	0.391	0.390	0.916
			0.20	0.002	0.217	0.223	0.92
		4000	0.05	0.0005	0.345	0.339	0.926
			0.10	-0.005	0.274	0.273	0.951
			0.20	-0.007	0.154	0.151	0.955
		6000	0.05	0.005	0.280	0.269	0.952
			0.10	-0.001	0.221	0.221	0.941
			0.20	-0.004	0.124	0.124	0.94
	0.5	1000	0.05	0.059	0.709	0.725	0.81

			0.10	0.031	0.566	0.569	0.903
			0.20	-0.005	0.319	0.315	0.935
		2000	0.05	0.036	0.497	0.504	0.871
			0.10	0.007	0.400	0.398	0.927
			0.20	0.006	0.227	0.235	0.918
		4000	0.05	0.006	0.344	0.336	0.92
			0.10	-0.002	0.281	0.279	0.950
			0.20	-0.006	0.161	0.162	0.949
		6000	0.05	0.0057	0.279	0.269	0.946
			0.10	0.002	0.227	0.227	0.935
			0.20	-0.005	0.130	0.131	0.94
Unif(0,1)	0	1000	0.05	0.026	1.274	1.300	0.793
			0.10	0.036	0.967	0.961	0.887
			0.20	-0.004	0.530	0.513	0.938
		2000	0.05	0.070	0.860	0.863	0.847
			0.10	0.010	0.683	0.665	0.907
			0.20	-0.001	0.375	0.384	0.922
		4000	0.05	-0.002	0.598	0.595	0.893
			0.10	-0.018	0.477	0.482	0.928
			0.20	-0.013	0.267	0.263	0.944
		6000	0.05	0.009	0.483	0.465	0.927
			0.10	0.004	0.386	0.376	0.948
			0.20	-0.006	0.217	0.213	0.952
	0.5	1000	0.05	0.046	1.271	1.292	0.779
			0.10	0.045	0.986	0.983	0.891
			0.20	0.003	0.555	0.546	0.93
		2000	0.05	0.084	0.857	0.860	0.847
			0.10	0.024	0.697	0.674	0.919

			0.20	0.001	0.394	0.404	0.92
		4000	0.05	0.005	0.596	0.595	0.888
			0.10	-0.012	0.488	0.495	0.929
			0.20	-0.009	0.280	0.278	0.944
		6000	0.05	0.013	0.481	0.463	0.927
			0.10	0.005	0.395	0.385	0.945
			0.20	-0.006	0.227	0.223	0.955

Table 4.2 summarizes the simulation results for situations when the proportion of events is not low (25%, 30% and 40%). We considered generalized case-cohort sampling. The sub-cohort sampling proportion is 10% and sampling proportion for the cases outside the sub-cohort is also 10%. The simulation results show that overall the coefficient estimates are approximately unbiased, the proposed variance estimator is close to the empirical variance, and the 95% confidence interval coverage is close to the nominal level for all the situations considered, except for when both the cohort size (1000 and 2000) and the event rate (25%) are both relatively low. As the cohort size or the event rate increases, the 95% confidence interval coverage improves.

Table 4.2: Summary of Simulation Results of $\hat{\beta}^{II}$ for Additive Model

Z	β_0	Cohort Size	Event proportion	Bias	Model Std. Error	Empirical Std. Dev.	Coverage
Bernoulli(0.5)	0	1000	0.25	0.028	1.199	1.257	0.969
			0.30	0.037	1.249	1.310	0.965
			0.40	0.062	1.129	1.181	0.949
		2000	0.25	-0.031	0.850	0.897	0.963
			0.30	-0.005	0.871	0.892	0.967
			0.40	0.038	0.798	0.796	0.957

		4000	0.25	-0.006	0.612	0.635	0.954
			0.30	-0.021	0.625	0.646	0.960
			0.40	-0.015	0.557	0.586	0.946
		6000	0.25	0.007	0.498	0.510	0.954
			0.30	0.004	0.509	0.514	0.946
			0.40	-0.014	0.456	0.453	0.958
		8000	0.25	0.004	0.609	0.621	0.959
			0.3	-0.0002	0.621	0.639	0.959
			0.4	-0.004	0.564	0.591	0.942
	0.5	1000	0.25	0.114	1.188	1.247	0.972
			0.30	0.046	1.246	1.290	0.959
			0.40	0.004	1.104	1.117	0.950
		2000	0.25	-0.054	0.855	0.865	0.980
			0.30	0.023	0.888	0.887	0.958
			0.40	0.017	0.781	0.792	0.947
		4000	0.25	0.016	0.613	0.611	0.965
			0.30	0.009	0.621	0.635	0.958
			0.40	-0.049	0.562	0.586	0.947
		6000	0.25	0.018	0.500	0.515	0.939
			0.30	-0.017	0.506	0.506	0.954
			0.40	-0.007	0.454	0.453	0.959
		8000	0.25	-0.007	0.613	0.617	0.965
			0.3	0.007	0.624	0.626	0.960
			0.4	-0.033	0.559	0.585	0.941
Unif(0, 1)	0	1000	0.25	0.069	2.093	2.183	0.971
			0.30	0.069	2.171	2.258	0.959
			0.40	0.118	1.94	2.098	0.954
		2000	0.25	-0.029	1.480	1.553	0.971

		0.30	-0.013	1.506	1.536	0.972
		0.40	0.041	1.356	1.364	0.957
	4000	0.25	-0.013	1.059	1.089	0.974
		0.30	-0.044	1.085	1.126	0.954
		0.40	-0.013	0.963	1.018	0.946
	6000	0.25	0.004	0.861	0.861	0.973
		0.30	0.002	0.889	0.895	0.96
		0.40	-0.014	0.784	0.777	0.967
	8000	0.25	0.006	1.05	1.067	0.971
		0.3	-0.001	1.067	1.095	0.960
		0.4	0.014	0.979	1.023	0.946
0.5	1000	0.25	-0.035	2.070	2.263	0.960
		0.30	0.007	2.129	2.255	0.969
		0.40	-0.064	1.910	1.941	0.954
	2000	0.25	0.031	1.498	1.480	0.972
		0.30	0.060	1.520	1.559	0.967
		0.40	0.044	1.387	1.393	0.957
	4000	0.25	-0.043	1.042	1.050	0.97
		0.30	-0.064	1.058	1.087	0.958
		0.40	0.039	0.968	0.994	0.955
	6000	0.25	-0.024	0.872	0.867	0.959
		0.30	-0.013	0.877	0.883	0.960
		0.40	-0.031	0.790	0.787	0.952
	8000	0.25	-0.009	1.049	1.077	0.967
		0.3	-0.003	1.066	1.086	0.963
		0.4	0.017	0.972	0.997	0.954

4.5 Real Data Application

In patients with cystic fibrosis, extracellular DNA (which are released by leukocytes) accumulates in the lung. This leads to exacerbations of respiratory symptoms and progressive deterioration of lung function. Lung disease has been considered as a leading cause of death among cystic fibrosis patients, accounting for over 90% of eventual deaths. DNase I is a human enzyme normally present in the mucus of human lungs that digests extracellular DNA. Genentech Inc. cloned a highly purified recombinant DNase I (rhDNase) in an effort to reduce the viscoelasticity of airway secretions and improving clearance (Therneau and Hamilton 1997). The company conducted a randomized double-blind clinical trial in 1992 comparing rhDNase to placebo for pulmonary exacerbations. For the purpose of the analysis, 645 individuals were eligible who were randomly assigned rhDNase or placebo at the beginning of the study. The primary end-point of the original study was the first pulmonary exacerbation but data was collected on all exacerbation in the follow-up period. We have used information on all the events for this analysis. The largest follow-up time was 170 days, with 37.52% of the cohort experiencing at least one event. Among those individuals who had at least one event, the average number of events is around 1.5. Further, for individuals who experienced a pulmonary exacerbation, they were temporarily absent from the risk sets.

Since the event is not rare, we constructed a generalized case-cohort (GCC) sample based on this cohort study to illustrate our proposed method. We consider the baseline forced expiratory volume in 1 second (FEV_1) as the expensive variable which is only available for the GCC sample. We are interested in studying the effect of FEV_1 on the recurrence of pulmonary exacerbations. We considered the sub-cohort selection proportion to be 0.2. Fifteen percent of all the individuals, who experienced at least one event but were not selected in the initial sub-cohort, were selected in the second stage to form the generalized case cohort sample. The total sample size of the GCC is 157 with 129 people in the sub-cohort. The following covariates were considered in the analysis: treatment group (rhDNase vs placebo) and FEV_1. Table 4.3 summarizes the distribution of the baseline variables in the sub-cohort and the full cohort.

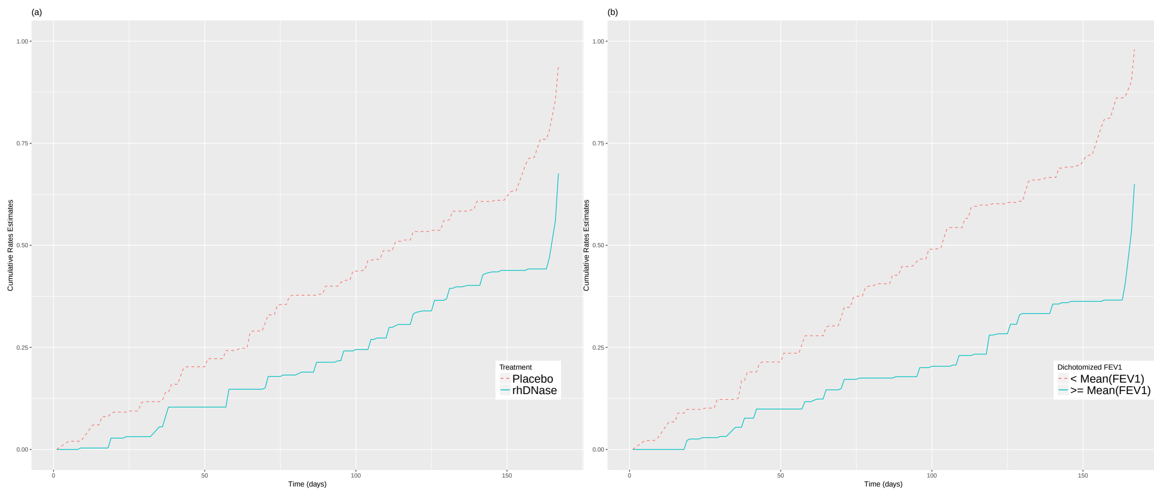

Figure 4.1: Plots of the cumulative rates function estimates vs time for the three variables

Table 4.3: Baseline Characteristics of the rhDNase study

Variables	Full Cohort (n = 645)	Subcohort ($\tilde{n}$ = 129)
Treatment (rhDNase vs. Placebo)	0.4977	0.4806
FEV_1	57.6 (38.4, 83.2)	60.8 (41.6, 83.5)

Table 4.3 shows that the distribution of the variables in the subcohort are quite similar to that in the full cohort. As an exploratory analysis, we have also looked at the cumulative rates for the two variables considered. The cumulative rate function was calculated non-parametrically for the different groups. It is defined as the ratio of the total number of events at time t and the total number of individuals at risk at that time. To get an idea about the cumulative rate function of FEV_1, we have dichotomized it. Further, we have centered and scaled FEV_1 to reduce the variability in the variable.

Table 4.4: Estimates and standard errors for the multiplicative rates model with data from GCC sample from the rhDNase study

Effects	Proposed method: GCC ($\tilde{\alpha} = 0.2$, $\tilde{q}$ =0.15)	
	Estimate	SE
Treatment (Reference: Placebo)	−0.00288	1.68×10^{-3}
FEV_1^*	−0.00204	7.26×10^{-4}

Figure 4.1 shows that, as time (measured in days) increases, the differences in the cumulative rates function estimates for two different levels of treatment as well as dichotomized FEV_1 increase approximately in a linear fashion. Therefore, it is reasonable to assume the additive effect of both the variables on the rates functions for the recurrence of pulmonary exacerbations. From the figure, one can visualize the effects to be different in the two groups. We fit the multivariable additive rates model to the data to test whether the differences are statistically significant. Table 4.4 provides results from the multivariable model. Adjusting for the treatment, forced expiratory volume in 1 second, FEV_1, is found to be statistically significant predictor of the recurrence of pulmonary aggravation. Hence, for an 1 sd increment in FEV_1, the risk of pulmonary exacerbations reduced by 0.2%. Further, the treatment variable is also found to be not significant.

In the above analysis, we defined a case as an individual who has experienced at least one event. One can use different definitions of a 'case' for recurrent events, based on the number of events. It is of interest to examine the performance of the estimates based on the different case definition for sampling. We conducted some simulations by sampling from the rhDNase study. Sub-cohort proportion was selected as 20% and the number of simulations considered is 1000. We have considered three definitions for 'case': subjects who experienced at least one event, at least two events, or at least three events. Table 4.5 shows the results from the simulations for the different case definitions. We note that the standard errors were quite similar for each of the case definitions, all of which are very close to 0.

Table 4.5: Estimates and standard errors for different definitions of 'case' for GCC sample from the rhDNase study

Effects	Case Definition					
	$\geq$ One Event		$\geq$ Two events		$\geq$ Three events	
	Estimate	SE	Estimate	SE	Estimate	SE
Treatment	-0.00104	0.00543	-0.000980	0.00536	-0.00107	0.00544
FEV_1^*	-0.00135	0.00012	-0.00135	0.00012	-0.00135	0.00012

4.6 Discussion

This article proposes methods of fitting marginal additive rates model for both the case-cohort and the generalized case-cohort designs with time-varying weights. The proposed estimators are natural generalizations of the full cohort estimators and possess intuitive interpretation. The proposed estimators are consistent and asymptotically normally distributed. They perform well in finite samples. We have applied the methods to the rhDNase data.

Throughout this article, we have assumed simple random sampling for both the sub-cohort and when sampling from the individuals who experienced the event but were not selected in the sub-cohort. Bernoulli sampling of the subcohort is another sampling scheme that can be considered. In the proposed method, the indicator variables are correlated with each other and the sub-cohort proportion is fixed, while in Bernoulli sampling scheme, the indicator variables are independently distributed leading to the sub-cohort proportion being the target value on the average. In our approach, we do not use all the covariate information that are available for the entire cohort. There have been studies (Jiang and Zhou 2007) that have examined the use of auxiliary covariates, available for all individuals, in the pseudo-score equation for estimating the parameters in additive rates model. It can be of interest, for future research, to extend the idea of using auxiliary variables to improve the efficiency of the estimators in recurrent events data.

CHAPTER 5: TWO-PHASE DESIGN SAMPLE SIZE AND POWER CALCULATION FOR TESTING INTERACTION BETWEEN TREATMENT AND EXPENSIVE BIOMARKER

5.1 Introduction

There is an increasing interest in discovering new biomarkers that can help predict the best treatment for a patient in cancer trials. However, finding new biomarkers is often time-consuming and costly, thus leading to the search for more efficient designs. In the new age of cancer research, there are often large quantities of clinical data and specimen available for further research after the randomized phase III trials are completed. Utilizing these already collected clinical data as the first phase data and the available biospecimen in the second phase of a two-phase design will shorten the discovery cycle. In the second phase, information from the existing biospecimen, which can be expensive, can be collected. The sampling of the second phase can utilize information that is already collected in the first phase, including the patients' disease status.

There have been extensive application of the two-phase design in epidemiologic cohort studies. Applications in clinical trials have focused on the study of predictive biomarkers. One of the most common schemes that is widely used is the case-cohort sampling design (Prentice 1986). In this design, a simple random sample (sub-cohort) is selected from the entire cohort along with all the subjects who experienced the event of interest (cases) during the study period. The case-cohort design have been extensively studied in time to event analysis (Prentice 1986, Lin and Ying 1993, Schouten et al. 1993, Barlow 1994, Liao et al. 1997, Barlow et al. 1999, Chen and Lo 1999, Savitz et al. 2000, Folsom et al. 2001, Chen 2001, Kulich and Lin 2004, Kang and Cai 2009b). Prentice (1986) proposed a pseudo-likelihood approach where the risk set at each failure consisted of only those subjects who were at risk in the sub-cohort. Self and Prentice (1988) slightly modified the pseudo-likelihood approach where the risk set at each failure includes the cases outside the sub-cohort along with the individuals in the sub-cohort. They also developed the asymptotic theory for the regression parameters. Lin and Ying (1993), Borgan et al. (2000), Barlow (1994), among others proposed different estimating equations and/or asymptotic variances for the parameters. There have been some studies examining the sample size and power calculation in case-cohort studies. Cai and Zeng (2004) provided an explicit procedure for calculating the sample size and power of their proposed test when the event is rare and the main exposure effect is of interest. They studied how much more power can be gained by using a case-cohort study instead of a simple random sample. Hu et al. (2014) derived such results for the stratified case-cohort design. They further studied the sample size and power for generalized case-cohort design (Cai and Zeng 2007). The aforementioned sample size and power formula are designed for testing main exposure effect. However, in studies of biomarker, the interaction between biomarker and treatment are often of interest. For example, in the pooled analysis of LACE (Lung Adjuvant Cisplatin Evaluation) and Cancer and Leukemia Group B (CALGB) 9633 databases (Shepherd et al. 2013), one is interested in developing a two-phase design to study how the effect of the KRAS-mutation on early stage resected non-small-cell lung cancer is different for individuals on adjuvant chemotherapy and the individuals who are on observation.

In this article, we propose methods for computing sample size and power for case-cohort studies when the event is rare and we are interested in assessing the interaction effect of an expensive biomarker and the treatment. We propose the test and bounds in Section 5.2 and investigate the finite sample properties by simulations in Section 5.3. In Section 5.4, we study the cost efficiency of the case-cohort design and apply our method to design a two-phased study based on information provided in LACE and CALGB 9633 databases. This example is considered to elaborate how much power one can achieve in testing an interaction in a case-cohort sampling scheme setting, for a particular sample size. Finally, in Section 5.5, we conclude based on our findings.

5.2 Methods

Assume that there are two biomarker groups ($j = 0, 1$) and two treatment groups ($k = 0, 1$) with n_{jk} individuals in each group, $\sum_{j,k} n_{jk} = n$. Let $T^*_{i,jk}$ be the potential failure time which can be censored by the potential censoring time, $C_{i,jk}$, for individual i in treatment k group and biomarker j group. $T_{ijk} = min(T^*_{i,jk}, C_{i,jk})$ is the observed time to event for individual i, ($i = 1, \ldots, n$) in treatment k group and biomarker j group. Let $\Delta_{ijk} = \mathbf{1}(T^*_{i,jk} \leq C_{i,jk})$ denote the failure indicator variable, X_i and A_i denote the biomarker and treatment covariates, respectively, for individual i. Let $\lambda_{j0}(t)$ ($j = 0, 1$) denote the hazard function for biomarker group, j, in treatment group, 0. We assume that the hazard rate for treatment group 1 satisfies: $\lambda_{j1}(t) = e^{\beta_j}\lambda_{j0}(t)$, where β_j is the unknown parameter for the effect of treatment in the j-th biomarker group. We are interested in testing whether the treatment effects are equal in the two biomarker groups, i.e., $H_0 : \beta_1 = \beta_0$.

For the case-cohort design, we select a random sample of the full cohort, called the sub-cohort, and all subjects who had the event. Let $\xi_{i,jk}$ be the indicator that subject i of treatment group k and biomarker group j is included in the sub-cohort. Define $\tilde{C}$ as the set of sub-cohort data, n_j and $\tilde{n}_j$ are the full cohort and sub-cohort sample sizes for the biomarker group j, respectively. The at-risk process is defined as $Y_{ijk}(t) = \mathbf{1}(T_{ijk} \geq t)$, $\pi_{jk}(x) = P(C_{i,jk} \geq x \mid X_i = j, A_i = k)$, $S_j(x) = P(T^*_{i,jk} > x \mid A_i = 0, X_i = j)$ for all i, $r_0 = 1 - r_1 = P(X_i = 0)$, $p_0 = P(A_i = 1 \mid X_i = 0)$ and $p_1 = P(A_i = 1 \mid X_i = 1)$. Let $\tilde{Y}_{jk}(t) = \sum_{i=1}^{\tilde{n}_j} Y_{ijk}(t)$ denote the total number of subjects at risk in the sub-cohort for biomarker j and treatment k and ψ_j as the sampling fraction for the subcohort of biomarker group j. We have $\hat{\psi}_j = \frac{\tilde{n}_j}{n_j}$, $\tilde{Y}_{jk}(t) = \hat{\psi}_j \bar{Y}_{jk}(t)$ where $\bar{Y}_{jk}(t) = \sum_{i=1}^{n_j} Y_{ijk}(t)$ is the risk set at time t for biomarker j and treatment k.

5.2.1 Proposed Tests

With the case-cohort design, the usual log-rank test statistic or the corresponding test based on the parameter estimate cannot be calculated since we do not have enough covariate information on all the individuals at the event times. The score function is given by:

$$\tilde{U}_j(\beta_j) = \sum_{i=1}^{n_{j1}} \frac{\tilde{Y}_{j0}(T_{ij1})\Delta_{ij1}}{\tilde{Y}_{j0}(T_{ij1}) + e^{\beta_j}\tilde{Y}_{j1}(T_{ij1})} - \sum_{i=1}^{n_{j0}} \frac{e^{\beta_j}\tilde{Y}_{j1}(T_{ij0})\Delta_{ij0}}{\tilde{Y}_{j0}(T_{ij0}) + e^{\beta_j}\tilde{Y}_{j1}(T_{ij0})} \quad \forall j = 0,1 \tag{5.1}$$

Denote the solution to $\tilde{U}_j(\beta_j) = 0$ by $\hat{\beta}_j$. The score function is same as that of the pseudo-partial likelihood function given by Self and Prentice (1988). Further, we have the asymptotic variance of the score function given by $\sigma_j^2 + \delta_j$, where $\sigma_j^2 = \int_0^1 \left(\frac{s^{(2)}(\beta_j,t)s^{(0)}(\beta_j,t) - s^{(1)}(\beta_j,t)^{\otimes 2}}{[s^{(0)}(\beta_j,t)]} \right) \lambda_j(t)$ $\times dt$ and $\delta_j = \int_0^1 \int_0^1 G(\beta_j,x,w)s^{(0)}(\beta_j,x)\, s^{(0)}(\beta_j,w)\lambda_j(x)\lambda_j(w)dxdw$, $q^{(l)}(\beta_j,x,w) = \sum_{k=0}^1 e^{2k\beta_j}$ $\times(1-p_j)^{1-k}p_j^k\pi_{jk}(x \vee w)S_j(x \vee w)^{e^{k\beta_j}}k^l$ $\forall l = 0,1$ and $0^0 = 1$, $s^{(l)}(\beta_j,x) = \sum_{k=0}^1 e^{k\beta_j}(1-p_j)^{1-k}p_j^k\pi_{jk}(x)\ S_j(x)^{e^{k\beta_j}}k^l$ $\forall l = 0,1,2$, and $h^{(k+l)}(\beta_j,x,w) = q^{(k+l)}(\beta_j,x,w) - s^{(k)}(\beta_j,w)$ $s^{(l)}(\beta_j,x)'$ $\forall l \le k = 0,1$. Since treatment is binary, $q^{(1)}(\beta_j,x,w) = q^{(2)}(\beta_j,x,w)$ and $s^{(1)}(\beta_j,x) = s^{(2)}(\beta_j,x)$. $G(\beta_j,x,w) = \frac{1-\psi_j}{\psi_j}\Big[\left\{s^{(0)}(\beta_j,x)s^{(0)}(\beta_j,w)\right\}^{-1} h^{(2)}(\beta_j,x,w) +$ $\left\{s^{(0)}(\beta_j,x)s^{(0)}(\beta_j,w)\right\}^{-2} s^{(1)}(\beta_j,x)s^{(1)}(\beta_j,w)\ \ h^{(0)}(\beta_j,x,w) - s^{(0)}(\beta_j,x)^{-1}s^{(0)}(\beta_j,w)^{-2}$ $s^{(1)}(\beta_j,w)h^{(1)}(\beta_j,w,x)s^{(0)}(\beta_j,w)^{-1}s^{(0)}(\beta_j,x)^{-2}\ \ -s^{(1)}(\beta_j,x)h^{(1)}(\beta_j,x,w)\Big]$. The first term of the variance, σ_j^2, corresponds to the variability in the cohort whereas δ_j is the variability due to sampling of the subcohort from the entire cohort. The asymptotic variance terms for each j can be estimated by $\hat{\sigma}_j^2 + \hat{\delta}_j$, where $\hat{\sigma}_j^2$ can be approximated by

$$\frac{1}{n_j}\left[\sum_{i=1}^{n_{j0}} \frac{e^{2\hat{\beta}_j}\tilde{Y}_{j1}(T_{ij0})^2\Delta_{ij0}}{\left(\tilde{Y}_{j0}(T_{ij0}) + e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ij0})\right)^2} + \sum_{i=1}^{n_{j1}} \frac{\tilde{Y}_{j0}(T_{ij1})^2\Delta_{ij1}}{\left(\tilde{Y}_{j0}(T_{ij1}) + e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ij1})\right)^2}\right]$$

and

$$\begin{aligned}\hat{\delta}_j = 2\frac{e^{2\hat{\beta}_j}(1-\hat{\psi}_j)}{n_j} \sum_{k,k'=0}^{1}\sum_{i=1}^{n_k}\sum_{i'=1}^{n_{k'}} \Bigg\{ & \frac{\Delta_{ijk}\Delta_{i'jk'}\mathbf{1}(T_{ijk'} \le T_{ijk})\tilde{Y}_{j1}(T_{ijk})\tilde{Y}_{j0}(T_{ijk})}{\left(\tilde{Y}_{j0}(T_{ijk}) + e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk})\right)^2\left(\tilde{Y}_{j0}(T_{ijk'}) + e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk'})\right)^2} \\ & \times\left(\tilde{Y}_{j1}(T_{ijk'}) + \tilde{Y}_{j0}(T_{ijk'})\right)\Bigg\} - \frac{e^{2\hat{\beta}_j}(1-\hat{\psi}_j)}{n_j}\sum_{k=0}^{1}\sum_{i=1}^{n_k}\Bigg\{\frac{\Delta_{ijk}\tilde{Y}_{j1}(T_{ijk})\tilde{Y}_{j0}(T_{ijk})}{\left(\tilde{Y}_{j0}(T_{ijk}) + e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk})\right)^4} \\ & \times\left(\tilde{Y}_{j1}(T_{ijk}) + \tilde{Y}_{j0}(T_{ijk})\right)\Bigg\}.\end{aligned}$$

From Self and Prentice (1988), we have $n^{1/2}\{\hat{\beta}_j - \beta_j\}$ converges to a Gaussian distribution with mean 0 and variance $r_j^{-1}\left(\sigma_j^{-2} + \sigma_j^{-4}\delta_j\right)$ for $j = 0, 1$ independently. Based on this asymptotic result, we propose the following test statistic:

$$TS_n = n^{1/2} \frac{\{\hat{\beta}_1 - \hat{\beta}_0\}}{\sqrt{\sum_j \frac{n}{n_j}\left(\hat{\sigma}_j^{-2} + \hat{\sigma}_j^{-4}\hat{\delta}_j\right)}}, \tag{5.2}$$

Assuming $\frac{n_j}{n} \to r_j \in (0,1)$ as $n \to \infty$, $TS_n \xrightarrow{D} N(0,1)$ under $H_0 : \beta_1 - \beta_0 = 0$. Hence, for level of significance, α, the rejection region of the test will be given by $TS_n > Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $100(1-\alpha)^{th}$ percentile of a standard normal distribution for $H_1 : \beta_1 > \beta_0$. One can note that under the alternative hypothesis $H_1 : \beta_1 - \beta_0 \gtrless 0$ the distribution of the statistic, $n^{1/2}\{(\hat{\beta}_1 - \hat{\beta}_0) - (\beta_1 - \beta_0)\}$ is $N\left(0, \sum_j r_j^{-1}(\sigma_j^{-2} + \sigma_j^{-4}\delta_j)\right)$.

5.2.2 Power Calculation

We have the following assumptions for sample size/power calculation:

(a) The censoring distributions are the same in the two treatment groups within each biomarker group.

(b) The proportion of failures are small in the full cohort.

(c) We assume the distribution of the underlying time to be continuous and hence, no ties of failures are observed.

We can approximate the ratio of the risk sets with the corresponding ratio of the probability of surviving at the end of the study. Under these assumptions and considering the alternative hypothesis $H_1 : \beta_1 - \beta_0 = \theta > 0$, $\theta = O(\tilde{n}^{-1/2})$ (where $\tilde{n} = \sum_j \tilde{n}_j$ and $\tilde{n}_j$ are of the same order), the power of the test statistic TS_n can be approximated by

$$\Phi\left[\sqrt{n}\,(\beta_1 - \beta_0)\left\{\sum_{j=0}^{1} r_j^{-1} \frac{1}{p_j(1-p_j)\left[e^{2\beta_j}p_j(1-p_D^{j1})^2 p_D^{j0} + (1-p_j)(1-p_D^{j0})^2 p_D^{j1}\right]}\right.\right.$$
$$\times\left(\left((1-p_j)(1-p_D^{j0}) + e^{\beta_j}p_j(1-p_D^{j1})\right)^2\right.$$

$$+\frac{\left[e^{2\beta_j}(1-\psi_j)(1-p_D^{j1})(1-p_D^{j0})\right]\left((1-p_j)p_D^{j0}+p_jp_D^{j1}\right)^2}{\psi_j\left(e^{2\beta_j}p_j(1-p_D^{j1})^2p_D^{j0}+(1-p_j)(1-p_D^{j0})^2p_D^{j1}\right)}\Bigg)\Bigg\}^{-1/2}-Z_{1-\alpha}\Bigg]. \tag{5.3}$$

where α is the level of significance, $p_D^{j1} = P(T^*_{ij1} \leq C_{ij1})$, $p_D^{j0} = P(T^*_{ij0} \leq C_{ij0})$, $j = 0, 1$. Define m_{*j0} as the risk set for biomarker j and treatment 0 at the last event time. If we assume that $\frac{p_D^{j1}}{p_D^{j0}} \approx 1$ and $\frac{m_{*j0}}{n_j} \approx (1-p_j)\left[p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right]$, the power of the test statistic can be written as

$$\Phi\Bigg[\sqrt{n}\,(\beta_1-\beta_0)\Bigg\{\sum_{j=0}^{1} r_j^{-1}\frac{1}{p_j(1-p_j)\left[e^{2\beta_j}p_jp_D^{j0}+(1-p_j)p_D^{j1}\right]}\times\Bigg(\left((1-p_j)+e^{\beta_j}p_j\right)^2$$
$$+\frac{e^{2\beta_j}(1-\psi_j)\left((1-p_j)p_D^{j0}+p_jp_D^{j1}\right)^2}{\psi_j\left(e^{2\beta_j}p_jp_D^{j0}+p_D^{j1}\right)\left(p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right)}\Bigg)\Bigg\}^{-1/2}-Z_{1-\alpha}\Bigg]. \tag{5.4}$$

Assuming that the censoring variable is degenerate at τ with probability $1-p_C$ and the approximation of the risk sets as $\frac{m_{*j0}}{n_j} \approx (1-p_C)(1-p_j)\left[p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right]$ and $\frac{m_{*j1}}{n_j} \approx (1-p_C)p_j\left[p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right]$, the power of the test statistic is given by

$$\Phi\Bigg[\sqrt{n}\,(\beta_1-\beta_0)\Bigg\{\sum_{j=0}^{1} r_j^{-1}\frac{1}{p_j(1-p_j)\left[e^{2\beta_j}p_jp_D^{j0}+(1-p_j)p_D^{j1}\right]}\times\Bigg(\left((1-p_j)+e^{\beta_j}p_j\right)^2$$
$$+\frac{e^{2\beta_j}(1-\psi_j)\left((1-p_j)p_D^{j0}+p_jp_D^{j1}\right)^2}{(1-p_C)\psi_j\left(e^{2\beta_j}p_jp_D^{j0}+p_D^{j1}\right)\left(p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right)}\Bigg)\Bigg\}^{-1/2}-Z_{1-\alpha}\Bigg]. \tag{5.5}$$

5.2.3 Sample Size formula

Based on the approximated power formula in the previous section, for a given power ϑ, significance level α, the entire cohort size n, and the denominator of any of the power formula ((5.3), (5.4) and (5.5)), denoted as σ_{den}, to detect the ratio of the hazard ratio, $exp(\beta_1-\beta_0)$, for the treatment effect between the two biomarker groups, the required total cohort size is

$$\frac{\left(Z_\vartheta+Z_{1-\alpha}\right)^2\sigma^2_{\text{den}}}{(\beta_1-\beta_0)^2}. \tag{5.6}$$

Specifically, based on power function, (5.3), the required sample size is

$$\frac{(Z_\vartheta + Z_{1-\alpha})^2}{(\beta_1-\beta_0)^2}\left\{\sum_{j=0}^{1} r_j^{-1}\frac{1}{p_j(1-p_j)\left[e^{2\beta_j}p_j(1-p_D^{j1})^2p_D^{j0}+(1-p_j)(1-p_D^{j0})^2p_D^{j1}\right]}(((1-p_j)\right.$$
$$\left.\times(1-p_D^{j0})+e^{\beta_j}p_j(1-p_D^{j1})\Big)^2+\frac{\left[e^{2\beta_j}(1-\psi_j)(1-p_D^{j1})(1-p_D^{j0})\right]\left((1-p_j)p_D^{j0}+p_jp_D^{j1}\right)^2}{\psi_j\left(e^{2\beta_j}p_j(1-p_D^{j1})^2p_D^{j0}+(1-p_j)(1-p_D^{j0})^2p_D^{j1}\right)}\Bigg)\right\} \tag{5.7}$$

Similarly, based on formula (5.4)

$$\frac{(Z_\vartheta + Z_{1-\alpha})^2}{(\beta_1-\beta_0)^2}\left\{\sum_{j=0}^{1} r_j^{-1}\frac{1}{p_j(1-p_j)\left[e^{2\beta_j}p_jp_D^{j0}+(1-p_j)p_D^{j1}\right]}\times\left(\left((1-p_j)+e^{\beta_j}p_j\right)^2\right.\right.$$
$$\left.\left.+\frac{e^{2\beta_j}(1-\psi_j)\left((1-p_j)p_D^{j0}+p_jp_D^{j1}\right)^2}{\psi_j\left(e^{2\beta_j}p_jp_D^{j0}+p_D^{j1}\right)\left(p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right)}\right)\right\} \tag{5.8}$$

and based on (5.5) we have

$$\frac{(Z_\vartheta + Z_{1-\alpha})^2}{(\beta_1-\beta_0)^2}\left\{\sum_{j=0}^{1} r_j^{-1}\frac{1}{p_j(1-p_j)\left[e^{2\beta_j}p_jp_D^{j0}+(1-p_j)p_D^{j1}\right]}\times\left(\left((1-p_j)+e^{\beta_j}p_j\right)^2\right.\right.$$
$$\left.\left.+\frac{e^{2\beta_j}(1-\psi_j)\left((1-p_j)p_D^{j0}+p_jp_D^{j1}\right)^2}{(1-p_C)\psi_j\left(e^{2\beta_j}p_jp_D^{j0}+p_D^{j1}\right)\left(p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right)}\right)\right\} \tag{5.9}$$

For a particular proportion of biomarker r_j and sub-cohort proportion ψ_j, the sample size required for the sub-cohort is

$$\tilde{n}_j = n \times r_j \times \psi_j \quad \forall j = 0, 1.$$

5.2.4 Bounds for the power formula

We consider the bounds for the power function. This also serves as a check for whether the power lies within the bounds. The bounds for power are defined as

$$\Phi\left[\frac{\sqrt{n}\,(\beta_1-\beta_0)}{\sqrt{\sum_{j=0}^{1} r_j^{-1}\left(\sigma_j^{-2}+\sigma_j^{-4}\delta_{j,ub}\right)}}-Z_{1-\alpha}\right]\le \text{Power}\le \Phi\left[\frac{\sqrt{n}\,(\beta_1-\beta_0)}{\sqrt{\sum_{j=0}^{1} r_j^{-1}\left(\sigma_j^{-2}+\sigma_j^{-4}\delta_{j,lb}\right)}}-Z_{1-\alpha}\right], \tag{5.10}$$

where $\sigma_j^2 \approx (1-p_j)p_j\frac{e^{2\beta_j}p_j(1-p_D^{j1})^2p_D^{j0}+(1-p_j)(1-p_D^{j0})^2p_D^{j1}}{\left((1-p_j)(1-p_D^{j0})+e^{\beta_j}p_j(1-p_D^{j1})\right)^2}$,

$$\hat{\delta}_{j,ub}\approx 2\frac{e^{2\hat{\beta}_j}(1-\psi_j)p_j(1-p_j)(1-p_D^{j1})(1-p_D^{j0})}{n_j\psi_j\left((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\right)^3}\left(p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right)$$
$$\times\sum_{l=1}^{m}\sum_{k=0}^{1}\frac{d_{lj}-1/2}{\left(\{n_j(1-p_j)-l_{j,0}+1\}(1+e^{\hat{\beta}_j}\frac{p_j(1-p_D^{j1})}{(1-p_j)(1-p_D^{j0})})\right)^{1-k}}\frac{1}{\left(\{n_jp_j-l_{j,1}+1\}(e^{\hat{\beta}_j}+\frac{(1-p_j)(1-p_D^{j0})}{p_j(1-p_D^{j1})})\right)^{k}},$$

$$\hat{\delta}_{j,lb}\approx 2\frac{e^{2\hat{\beta}_j}(1-\psi_j)p_j(1-p_j)(1-p_D^{j1})(1-p_D^{j0})\left(p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right)}{n_j\psi_j\left((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\right)^3}$$
$$\times\sum_{l=1}^{m}\sum_{k=0}^{1}\frac{d_{lj}-1/2}{\left(\{n_j(1-p_j)-n_{j0C}-D_0^j\}(1+e^{\hat{\beta}_j}\frac{p_j(1-p_D^{j1})}{(1-p_j)(1-p_D^{j0})})\right)^{1-k}\left(\{n_jp_j-n_{j1C}-D_1^j\}(e^{\hat{\beta}_j}+\frac{(1-p_j)(1-p_D^{j0})}{p_j(1-p_D^{j1})})\right)^{k}},$$

where d_{lj} is the risk set in the biomarker group j at the l-th index of the ordered failure time and m is the total number of failures, n_{jkC} and D_k^j are the total number of censored individuals in $[0,\tau)$ and number of deaths for biomarker j and treatment group k respectively.

5.3 Simulation Results

We conducted simulations to examine the finite sample properties of the formulae. The empirical results based on the test statistic from the full-cohort and sub-cohort are also reported. It should be noted that full cohort design is not possible when one considers the case-cohort sampling scheme as information is not collected on individuals who are not in the case-cohort. We have reported the simulation results to provide an upper bound. We examined different scenarios to ascertain the performance of the proposed formula. In all the situations, we assumed that the treatment is randomly assigned with probability 0.5. Different biomarker proportions (0.3, 0.5) are considered. We generated the censoring time from a mixture distribution, with probability p_C from uniform distribution in $[0, \tau]$ and probability $(1 - p_C)$ being degenerate at τ. For the simulations, we considered $p_C = 0.3, 0.2, 0.1$. The bound for the censoring distribution, τ, is calculated such that the censoring proportion for the (0, 0) group of the (Biomarker, Treatment) combination (denoted by $1 - p_D^{00}$) is 95%, 90% and 80%.

The effect of treatment is assumed to be multiplicative in all the cases. Table 5.1 show that the Type I error of the test based on data from the Weibull distribution with hazard given by $\lambda_{j1}(t) = 2\lambda_j t e^{\beta_j}$, $t \in (0, \infty)$, and $\lambda_{j0}(t) = 2\lambda_j t$, $j = 0, 1$ and $p_C = 0.2$. Both the biomarker proportions were considered for the simulations with $\lambda_0 = 1$ and $\lambda_1 = 0.75$, 1 and 1.25. The number of simulations was 20000. The Type I error is always very close to 0.05 (the nominal level) for each of the designs for all the situations considered which shows that the proposed test is reliable.

Table 5.1: Summary of Type I Error for Weibull (2) for $\beta_1 - \beta_0 = 0.25$ and $1 - p_C = 0.8$)

Distribution (l_0, l_1)	Event prop. (p_D^{00})	Biomarker prop. r_0	Full Cohort	Case-Cohort	Sub-cohort
(1, 0.75)	(0.05)	0.3	0.0517	0.0518	0.0422
		0.5	0.0484	0.0501	0.0468
	(0.1)	0.3	0.0492	0.0460	0.0501
		0.5	0.0521	0.0494	0.0504
	(0.2)	0.3	0.0505	0.0513	0.0507
		0.5	0.0508	0.0495	0.049
(1, 1)	(0.05)	0.3	0.0502	0.0501	0.0419
		0.5	0.0493	0.0487	0.0406
	(0.1)	0.3	0.0498	0.0477	0.0485
		0.5	0.0486	0.0505	0.0484
	(0.2)	0.3	0.0482	0.0486	0.0494
		0.5	0.0485	0.0509	0.0474
(1, 1.25)	(0.05)	0.3	0.0489	0.0473	0.036
		0.5	0.0479	0.0490	0.0231
	(0.1)	0.3	0.0493	0.0512	0.0502
		0.5	0.0509	0.0509	0.0503
	(0.2)	0.3	0.0486	0.0499	0.0508
		0.5	0.0492	0.0503	0.0493

After confirming the performance of the empirical Type I error, we compared the empirical and theoretical powers under different situations. Tables 5.5 - 5.10 show data from Weibull distribution with hazard, $\lambda_{j0}(t) = l\lambda_j t^{l-1}, \quad t \in (0, \infty)$ for treatment group 0 and $\lambda_{j1}(t) = l\lambda_j t^{l-1} e^{\beta_j}$, $t \in (0, \infty)$, $j = 0, 1, l = 2, 3$ for treatment group 1. The sample size considered was 4000. Tables 5.2 - 5.4 & 5.11 - 5.13 summarized the power for data from an exponential distribution with hazards in one biomarker group = 1 and considering $0.75, 1$ or 1.25 for the other group. For the power calculation, we considered 5000 simulations with 10% sub-cohort sampling proportion. We considered $\beta_0 = 0.5$ and $\beta_1 = 1$ for Tables 5.2 - 5.10 and $\beta_0 = 0.5$ and $\beta_1 = 0.75$ for Tables 5.11 - 5.13 and $P(\text{Treatment} = 1 \mid X = j) = 0.5$ for illustration purposes. We summarized the theoretical formula for the power by the following and reported in the tables.

When the event rate is low (≤ 0.05) and the percentage of individuals censored at the end of the study is $< 90\%$, the power formula (5.3) and (5.4) overestimates the attainable power whereas (5.5) under-estimates the empirical power slightly whereas when the percentage of individuals censored is $\geq 90\%$, (5.3) works quite well. For event rate $\in (0.05, 0.1]$, then for the situation with $p_C = 0.3$, we have (5.5), for $p_C = 0.2$, we have (5.3) and finally, for $p_C = 0.1$, (5.4) was found to be closest to the empirical power. When the event rate is $\in (0.1, 0.2]$, (5.4) works well for $p_C < 0.3$ whereas for $p_C = 0.3$, (5.3) best estimates the power. Based on these performances of the different theoretical power formulae with the empirical formulae, we summarized our recommendations below :

$$\begin{aligned}
&Eqn(5.4) \quad \text{if } 1 - p_C \geq 0.9, p_D^{00} \geq 0.2 \text{ or if } 0.8 \leq 1 - p_C < 0.9, p_D^{00} \geq 0.2,\\
&Eqn(5.3) \quad \text{if } 1 - p_C \geq 0.9, p_D^{00} < 0.2 \text{ or if } 0.8 \leq 1 - p_C < 0.9, p_D^{00} \geq 0.1 \text{ or if } 1 - p_C < 0.8, p_D^{00} \geq 0.2,\\
&Eqn(5.5) \quad \text{if } 0.8 \leq 1 - p_C < 0.9, p_D^{00} < 0.1 \text{ or if } 1 - p_C < 0.8, p_D^{00} < 0.2.
\end{aligned} \tag{5.11}$$

The combined power formula (5.11) is more or less conservative for all the distributions for all event rates. For Weibull distribution, when the hazard function is a cubic polynomial and censoring degenerate probability $(1 - p_C)$ is 0.8, the formula slightly over-estimates the empirical formula whereas for all the other scenarios, it is conservative. One can note that there is a gain in using case-cohort design rather than only a sample of the subjects. For rare event, the proposed formula works pretty well. Even though the event rate is quite low, we have calculated the empirical bounds for the power function based on (5.10). The empirical and theoretical powers for the case-cohort design almost always lie within the bounds.

Table 5.2: Summary of Power Calculation for Exponential Distribution with $\beta_1 - \beta_0 = 0.5$ and $1 - p_C = 0.7$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort				Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Final	Bounds	Empirical
(1, 0.75)	(0.05)	0.3	0.502	0.391	0.408/0.410/0.378	0.378	(0.367, 0.413)	0.129
		0.5	0.571	0.45	0.453/0.456/0.420	0.420	(0.423, 0.474)	0.124
	(0.1)	0.3	0.782	0.536	0.537/0.547/0.481	0.481	(0.467, 0.563)	0.211
		0.5	0.848	0.599	0.593/0.607/0.537	0.537	(0.524, 0.612)	0.223
	(0.2)	0.3	0.957	0.634	0.597/0.626/0.528	0.597	(0.492, 0.663)	0.324
		0.5	0.980	0.683	0.656/0.693/0.590	0.656	(0.559, 0.716)	0.339
(1, 1)	(0.05)	0.3	0.540	0.414	0.423/0.426/0.391	0.391	(0.38, 0.428)	0.142
		0.5	0.612	0.490	0.480/0.485/0.444	0.444	(0.448, 0.499)	0.169
	(0.1)	0.3	0.816	0.547	0.547/0.560/0.490	0.490	(0.465, 0.575)	0.29
		0.5	0.885	0.617	0.618/0.630/0.553	0.553	(0.535, 0.638)	0.233
	(0.2)	0.3	0.966	0.650	0.596/0.632/0.529	0.596	(0.483, 675)	0.348
		0.5	0.989	0.707	0.652/0.702/0.591	0.652	(0.544, 0.724)	0.384
(1, 1.25)	(0.05)	0.3	0.558	0.424	0.433/0.436/0.399	0.399	(0.391, 0.437)	0.152
		0.5	0.649	0.49	0.499/0.505/0.460	0.460	(0.462, 0.517)	0.159
	(0.1)	0.3	0.826	0.557	0.553/0.568/0.495	0.495	(0.472, 0.586)	0.228
		0.5	0.905	0.619	0.620/0.644/0.561	0.561	(0.533, 0.662)	0.259

	(0.2)	0.3	0.972	0.639	0.590/0.633/0.527	0.590	(0.470, 0.679)	0.356
		0.5	0.991	0.698	0.652/0.702/0.591	0.652	(0.521, 0.735)	0.392

Table 5.3: Summary of Power Calculation for Exponential Distribution with $\beta_1 - \beta_0 = 0.5$ and $1 - p_C = 0.8$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort				Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Final	Bounds	Empirical
(1, 0.75)	(0.05)	0.3	0.519	0.404	0.418/0.421/0.400	0.400	(0.388, 0.419)	0.134
		0.5	0.598	0.459	0.464/0.467/0.444	0.444	(0.446, 0.481)	0.131
	(0.1)	0.3	0.798	0.551	0.545/0.555/0.512	0.545	(0.501, 0.572)	0.223
		0.5	0.864	0.622	0.602/0.616/0.570	0.602	(0.565, 0.63)	0.242
	(0.2)	0.3	0.958	0.633	0.597/0.626/0.561	0.626	(0.539, 0.674)	0.336
		0.5	0.981	0.693	0.655/0.694/0.625	0.694	(0.604, 0.721)	0.353
(1, 1)	(0.05)	0.3	0.569	0.436	0.433/0.437/0.413	0.413	(0.404, 0.439)	0.148
		0.5	0.634	0.492	0.491/0.497/0.470	0.470	(0.472, 0.508)	0.169
	(0.1)	0.3	0.836	0.57	0.554/0.568/0.522	0.554	(0.507, 0.585)	0.23
		0.5	0.901	0.639	0.618/0.638/0.587	0.618	(0.571, 0.655)	0.239
	(0.2)	0.3	0.967	0.643	0.594/0.631/0.563	0.631	(0.535, 0.679)	0.323
		0.5	0.990	0.724	0.650/0.702/0.627	0.702	(0.593, 0.736)	0.371
(1, 1.25)	(0.05)	0.3	0.564	0.435	0.443/0.447/0.422	0.422	(0.411, 0.451)	0.157
		0.5	0.675	0.513	0.510/0.517/0.487	0.487	(0.488, 0.529)	0.180

	(0.1)	0.3	0.845	0.57	0.559/0.576/0.527	0.559	(0.509, 0.595)	0.243
		0.5	0.915	0.650	0.625/0.651/0.596	0.625	(0.577, 0.67)	0.277
	(0.2)	0.3	0.972	0.643	0.588/0.633/0.561	0.633	(0.526, 0.685)	0.343
		0.5	0.992	0.721	0.640/0.705/0.624	0.705	(0.573, 0.748)	0.422

Table 5.4: Summary of Power Calculation for Exponential Distribution with $\beta_1 - \beta_0 = 0.5$ and $1 - p_C = 0.9$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort				Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Final	Bounds	Empirical
(1, 0.75)	(0.05)	0.3	0.538	0.416	0.428/0.431/0.420	0.428	(0.412, 0.431)	0.145
		0.5	0.613	0.483	0.475/0.479/0.467	0.475	(0.469, 0.489)	0.135
	(0.1)	0.3	0.790	0.549	0.540/0.551/0.530	0.551	(0.518, 0.562)	0.217
		0.5	0.858	0.619	0.597/0.611/0.588	0.611	(0.576, 0.622)	0.224
	(0.2)	0.3	0.963	0.649	0.596/0.627/0.592	0.627	(0.579, 0.672)	0.327
		0.5	0.980	0.706	0.654/0.694/0.656	0.694	(0.630, 0.726)	0.359
(1, 1)	(0.05)	0.3	0.579	0.446	0.443/0.447/0.435	0.443	(0.429, 0.449)	0.148
		0.5	0.653	0.514	0.502/0.508/0.494	0.502	(0.497, 0.518)	0.168
	(0.1)	0.3	0.825	0.572	0.550/0.563/0.541	0.563	(0.527, 0.576)	0.231
		0.5	0.893	0.630	0.611/0.633/0.607	0.633	(0.592, 0.636)	0.249
	(0.2)	0.3	0.969	0.659	0.593/0.631/0.593	0.631	(0.577, 0.680)	0.348

Distribution (λ_0, λ_1)	Event prop. (p_D^{00})	Biom. prop. r_0	Full Cohort Empirical	Case-Cohort Empirical	Case-Cohort Theoretical	Case-Cohort Final	Case-Cohort Bounds	Sub-cohort Empirical
		0.5	0.993	0.720	0.649/0.702/0.659	0.702	(0.623, 0.739)	0.386
(1, 1.25)	(0.05)	0.3	0.589	0.449	0.453/0.457/0.445	0.453	(0.435, 0.456)	0.164
		0.5	0.692	0.521	0.520/0.528/0.512	0.520	(0.509, 0.536)	0.173
	(0.1)	0.3	0.841	0.572	0.555/0.571/0.547	0.571	(0.53, 0.585)	0.235
		0.5	0.908	0.648	0.622/0.647/0.618	0.647	(0.598, 0.654)	0.259
	(0.2)	0.3	0.974	0.656	0.587/0.633/0.592	0.633	(0.568, 0.686)	0.351
		0.5	0.994	0.730	0.637/0.704/0.656	0.704	(0.614, 0.746)	0.413

Table 5.5: Summary of Power Calculation for Weibull(2) Distribution with $\beta_1 - \beta_0 = 0.5$ and $1 - p_C = 0.7$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort				Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Final	Bounds	Empirical
(1, 0.75)	(0.05)	0.3	0.559	0.414	0.434/0.437/0.399	0.399	(0.388, 0.445)	0.149
		0.5	0.621	0.469	0.481/0.485/0.445	0.445	(0.433, 0.488)	0.141
	(0.1)	0.3	0.794	0.539	0.543/0.553/0.486	0.486	(0.475, 0.574)	0.221
		0.5	0.852	0.601	0.599/0.613/0.541	0.541	(0.515, 0.616)	0.221
	(0.2)	0.3	0.964	0.607	0.6/0.63/0.53	0.6	(0.477, 0.671)	0.329
		0.5	0.983	0.681	0.658/0.698/0.593	0.658	(0.533, 0.710)	0.370
(1, 1)	(0.05)	0.3	0.597	0.438	0.449/0.453/0.412	0.412	(0.405, 0.459)	0.159
		0.5	0.672	0.497	0.508/0.514/0.467	0.467	(0.446, 0.510)	0.171
	(0.1)	0.3	0.822	0.525	0.553/0.566/0.495	0.495	(0.479, 0.592)	0.227

		0.5	0.891	0.611	0.617/0.637/0.558	0.558	(0.533, 0.637)	0.251
	(0.2)	0.3	0.969	0.619	0.598/0.635/0.531	0.598	(0.470, 0.677)	0.353
		0.5	0.990	0.687	0.655/0.707/0.594	0.655	(0.520, 0.722)	0.392
(1, 1.25)	(0.05)	0.3	0.623	0.450	0.458/0.463/0.419	0.419	(0.406, 0.466)	0.165
		0.5	0.705	0.526	0.526/0.534/0.482	0.482	(0.456, 0.528)	0.183
	(0.1)	0.3	0.826	0.547	0.557/0.499	0.499	(0.482, 0.595)	0.227
		0.5	0.907	0.616	0.625/0.649/0.565	0.565	(0.530, 0.650)	0.270
	(0.2)	0.3	0.978	0.621	0.592/0.637/0.529	0.592	(0.456, 0.682)	0.345
		0.5	0.993	0.681	0.645/0.71/0.591	0.645	(0.496, 0.737)	0.409

Table 5.6: Summary of Power Calculation for Weibull(2) Distribution with $\beta_1 - \beta_0 = 0.5$ and $1 - p_C = 0.8$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort				Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Final	Bounds	Empirical
(1, 0.75)	(0.05)	0.3	0.560	0.422	0.427/0.43/0.408	0.408	(0.406, 0.438)	0.151
		0.5	0.608	0.47	0.474/0.478/0.453	0.453	(0.443, 0.475)	0.153
	(0.1)	0.3	0.818	0.559	0.55/0.561/0.517	0.55	(0.506, 0.577)	0.225
		0.5	0.865	0.625	0.606/0.622/0.574	0.606	(0.555, 0.630)	0.240
	(0.2)	0.3	0.967	0.623	0.598/0.629/0.563	0.629	(0.531, 0.674)	0.328
		0.5	0.984	0.691	0.656/0.697/0.627	0.697	(0.584, 0.719)	0.367
(1, 1)	(0.05)	0.3	0.577	0.436	0.442/0.446/0.422	0.422	(0.412, 0.449)	0.154

		0.5	0.65	0.490	0.501/0.507/0.479	0.479	(0.467, 0.500)	0.17
	(0.1)	0.3	0.830	0.553	0.559/0.573/0.526	0.559	(0.517, 0.589)	0.235
		0.5	0.899	0.615	0.623/0.644/0.592	0.623	(0.564, 0.646)	0.251
	(0.2)	0.3	0.975	0.647	0.596/0.634/0.565	0.634	(0.526, 0.679)	0.360
		0.5	0.989	0.702	0.651/0.705/0.63	0.705	(0.575, 0.732)	0.396
(1, 1.25)	(0.05)	0.3	0.605	0.454	0.452/0.456/0.43	0.43	(0.420, 0.454)	0.165
		0.5	0.693	0.523	0.519/0.527/0.496	0.496	(0.475, 0.517)	0.183
	(0.1)	0.3	0.849	0.557	0.562/0.58/0.531	0.562	(0.518, 0.597)	0.244
		0.5	0.923	0.640	0.629/0.656/0.6	0.629	(0.571, 0.66)	0.277
	(0.2)	0.3	0.979	0.634	0.589/0.635/0.563	0.635	(0.514, 0.687)	0.348
		0.5	0.995	0.701	0.641/0.707/0.626	(0.553, 0.736)	0.420	

Table 5.7: Summary of Power Calculation for Weibull(2) Distribution with $\beta_1 - \beta_0 = 0.5$ and $1 - p_C = 0.9$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort				Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Final	Bounds	Empirical
(1, 0.75)	(0.05)	0.3	0.586	0.449	0.441/0.444/0.432	0.441	(0.434, 0.457)	0.160
		0.5	0.640	0.490	0.489/0.493/0.48	0.489	(0.475, 0.497)	0.155
	(0.1)	0.3	0.802	0.562	0.545/0.555/0.534	0.555	(0.525, 0.573)	0.216
		0.5	0.862	0.626	0.601/0.615/0.592	0.615	(0.577, 0.623)	0.235
	(0.2)	0.3	0.972	0.658	0.597/0.63/0.594	0.63	(0.565, 0.669)	0.336

		0.5	0.989	0.703	0.654/0.698/0.658	0.698	(0.631, 0.728)	0.377
(1, 1)	(0.05)	0.3	0.600	0.457	0.456/0.46/0.447	0.456	(0.446, 0.467)	0.155
		0.5	0.681	0.515	0.516/0.522/0.507	0.516	(0.501, 0.524)	0.171
	(0.1)	0.3	0.828	0.564	0.554/0.568/0.545	0.568	(0.536, 0.586)	0.224
		0.5	0.900	0.634	0.618/0.638/0.612	0.638	(0.592, 0.645)	0.259
	(0.2)	0.3	0.982	0.656	0.593/0.634/0.595	0.634	(0.563, 0.674)	0.368
		0.5	0.991	0.719	0.647/0.705/0.66	0.705	(0.616, 0.736)	0.408
(1, 1.25)	(0.05)	0.3	0.635	0.476	0.465/0.47/0.456	0.465	(0.451, 0.478)	0.158
		0.5	0.720	0.543	0.533/0.541/0.525	0.533	(0.51, 0.537)	0.186
	(0.1)	0.3	0.836	0.568	0.558/0.575/0.55	0.575	(0.540, 0.595)	0.242
		0.5	0.915	0.648	0.625/0.651/0.621	0.651	(0.601, 0.655)	0.270
	(0.2)	0.3	0.983	0.662	0.585/0.634/0.59	0.634	(0.553, 0.676)	0.373
		0.5	0.997	0.722	0.634/0.705/0.656	0.705	(0.605, 0.745)	0.427

Table 5.8: Summary of Power Calculation for Weibull(3) Distribution with $\beta_1 - \beta_0 = 0.5$ and $1 - p_C = 0.7$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort				Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Final	Bounds	Empirical
(1, 0.75)	(0.05)	0.3	0.549	0.406	0.431/0.434/0.397	0.397	(0.391, 0.446)	0.145
		0.5	0.620	0.465	0.479/0.483/0.442	0.442	(0.431, 0.485)	0.138
	(0.1)	0.3	0.780	0.527	0.538/0.548/0.482	0.482	(0.472, 0.567)	0.217

		0.5	0.842	0.589	0.594/0.607/0.537	0.537	(0.513, 0.608)	0.215
	(0.2)	0.3	0.964	0.589	0.6/0.63/0.531	0.6	(0.48, 0.665)	0.328
		0.5	0.982	0.667	0.659/0.698/0.594	0.659	(0.536, 0.709)	0.364
(1, 1)	(0.05)	0.3	0.587	0.434	0.45/0.454/0.413	0.413	(0.397, 0.456)	0.159
		0.5	0.664	0.488	0.506/0.512/0.465	0.465	(0.447, 0.509)	0.173
	(0.1)	0.3	0.808	0.516	0.552/0.566/0.494	0.494	(0.477, 0.582)	0.217
		0.5	0.882	0.592	0.613/0.631/0.554	0.554	(0.528, 0.629)	0.249
	(0.2)	0.3	0.97	0.603	0.599/0.505/0.532	0.599	(0.469, 0.676)	0.348
		0.5	0.989	0.675	0.657/0.708/0.597	0.657	(0.516, 0.719)	0.391
(1, 1.25)	(0.05)	0.3	0.612	0.444	0.456/0.46/0.418	0.418	(0.402, 0.464)	0.166
		0.5	0.701	0.520	0.524/0.531/0.48	0.48	(0.455, 0.528)	0.186
	(0.1)	0.3	0.813	0.532	0.553/0.568/0.496	0.496	(0.481, 0.589)	0.223
		0.5	0.898	0.601	0.621/0.645/0.563	0.563	(0.522, 0.644)	0.259
	(0.2)	0.3	0.977	0.605	0.594/0.638/0.531	0.594	(0.454, 0.680)	0.341
		0.5	0.993	0.666	0.649/0.712/0.594	0.649	(0.491, 0.729)	0.405

Table 5.9: Summary of Power Calculation for Weibull(3) Distribution with $\beta_1 - \beta_0 = 0.5$ and $1 - p_C = 0.8$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort				Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Final	Bounds	Empirical
(1, 0.75)	(0.05)	0.3	0.574	0.430	0.435/0.438/0.415	0.415	(0.406, 0.447)	0.157

		0.5	0.626	0.485	0.483/0.487/0.461	0.461	(0.453, 0.488)	0.152
	(0.1)	0.3	0.801	0.54	0.543/0.553/0.511	0.543	(0.492, 0.564)	0.215
		0.5	0.852	0.611	0.599/0.613/0.567	0.599	(0.546, 0.615)	0.232
	(0.2)	0.3	0.959	0.613	0.598/0.627/0.563	0.627	(0.528, 0.667)	0.318
		0.5	0.98	0.676	0.656/0.694/0.626	0.694	(0.584, 0.709)	0.347
(1, 1)	(0.05)	0.3	0.596	0.441	0.45/0.454/0.429	0.429	(0.414, 0.457)	0.152
		0.5	0.666	0.499	0.51/0.516/0.486	0.486	(0.47, 0.510)	0.172
	(0.1)	0.3	0.819	0.539	0.552/0.566/0.52	0.552	(0.505, 0.577)	0.228
		0.5	0.887	0.603	0.617/0.636/0.586	0.617	(0.559, 0.635)	0.245
	(0.2)	0.3	0.969	0.634	0.597/0.632/0.565	0.632	(0.526, 0.677)	0.345
		0.5	0.986	0.692	0.654/0.704/0.63	0.704	(0.577, 0.721)	0.378
(1, 1.25)	(0.05)	0.3	0.626	0.462	0.459/0.464/0.437	0.437	(0.427, 0.466)	0.169
		0.5	0.711	0.532	0.528/0.536/0.503	0.503	(0.481, 0.529)	0.192
	(0.1)	0.3	0.831	0.550	0.557/0.573/0.525	0.557	(0.509, 0.580)	0.229
		0.5	0.912	0.630	0.624/0.649/0.595	0.624	(0.564, 0.652)	0.270
	(0.2)	0.3	0.972	0.622	0.591/0.634/0.564	0.634	(0.514, 0.679)	0.333
		0.5	0.994	0.693	0.645/0.707/0.629	0.707	(0.559, 0.730)	0.408

Table 5.10: Summary of Power Calculation for Weibull(3) Distribution with $\beta_1 - \beta_0 = 0.5$ and $1 - p_C = 0.9$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort	Sub-cohort

(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Final	Bounds	Empirical
(1, 0.75)	(0.05)	0.3	0.561	0.436	0.43/0.433/0.422	0.43	(0.424, 0.445)	0.157
		0.5	0.622	0.478	0.477/0.481/0.469	0.477	(0.463, 0.482)	0.148
	(0.1)	0.3	0.808	0.559	0.546/0.557/0.535	0.557	(0.525, 0.574)	0.219
		0.5	0.862	0.624	0.602/0.617/0.593	0.582	(0.582, 0.627)	0.233
	(0.2)	0.3	0.972	0.653	0.598/0.63/0.594	0.63	(0.563, 0.666)	0.333
		0.5	0.988	0.698	0.655/0.697/0.658	0.697	(0.628, 0.730)	0.375
(1, 1)	(0.05)	0.3	0.582	0.447	0.447/0.451/0.438	0.447	(0.437, 0.460)	0.145
		0.5	0.657	0.502	0.504/0.51/0.496	0.504	(0.488, 0.509)	0.161
	(0.1)	0.3	0.832	0.562	0.549/0.561/0.539	0.561	(0.535, 0.584)	0.224
		0.5	0.899	0.631	0.619/0.64/0.597/0.613	0.64	(0.590, 0.646)	0.256
	(0.2)	0.3	0.980	0.652	0.594/0.634/0.595	0.634	(0.559, 0.672)	0.366
		0.5	0.991	0.715	0.648/0.705/0.661	0.705	(0.614, 0.737)	0.407
(1, 1.25)	(0.05)	0.3	0.615	0.463	0.454/0.459/0.446	0.454	(0.447, 0.467)	0.156
		0.5	0.698	0.535	0.522/0.53/0.514	0.522	(0.502, 0.530)	0.185
	(0.1)	0.3	0.838	0.567	0.559/0.576/0.551	0.576	(0.538, 0.595)	0.241
		0.5	0.916	0.646	0.626/0.652/0.622	0.652	(0.601, 0.658)	0.278
	(0.2)	0.3	0.982	0.660	0.586/0.635/0.593	0.635	(0.550, 0.673)	0.371
		0.5	0.996	0.717	0.636/0.706/0.657	0.706	(0.604, 0.745)	0.424

Table 5.11: Summary of Power Calculation for Weibull(2) Distribution with $\beta_1 - \beta_0 = 0.25$ and $1 - p_C = 0.7$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort				Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Final	Bounds	Empirical
(1, 0.75)	(0.05)	0.3	0.231	0.171	0.181/0.182/0.169	0.169	(0.166, 0.198)	0.086
		0.5	0.251	0.188	0.196/0.197/0.184	0.184	(0.165, 0.200)	0.078
	(0.1)	0.3	0.340	0.223	0.22/0.223/0.199	0.199	(0.181, 0.248)	0.111
		0.5	0.388	0.246	0.242/0.245/0.219	0.219	(0.189, 0.257)	0.107
	(0.2)	0.3	0.530	0.248	0.245/0.254/0.216	0.245	(0.168, 0.299)	0.149
		0.5	0.592	0.278	0.271/0.283/0.239	0.271	(0.186, 0.319)	0.164
(1, 1)	(0.05)	0.3	0.240	0.180	0.187/0.188/0.174	0.174	(0.168, 0.206)	0.091
		0.5	0.270	0.199	0.207/0.208/0.192	0.192	(0.172, 0.215)	0.095
	(0.1)	0.3	0.362	0.219	0.225/0.228/0.203	0.203	(0.179, 0.258)	0.12
		0.5	0.402	0.243	0.251/0.255/0.225	0.225	(0.191, 0.267)	0.119
	(0.2)	0.3	0.568	0.251	0.245/0.256/0.216	0.245	(0.164, 0.305)	0.154
		0.5	0.618	0.279	0.272/0.287/0.24	0.272	(0.174, 0.324)	0.176
(1, 1.25)	(0.05)	0.3	0.249	0.192	0.19/0.191/0.177	0.177	(0.171, 0.21)	0.093
		0.5	0.290	0.222	0.214/0.215/0.198	0.198	(0.181, 0.222)	0.100
	(0.1)	0.3	0.361	0.214	0.227/0.231/0.204	0.204	(0.181, 0.263)	0.111
		0.5	0.441	0.257	0.255/0.261/0.228	0.228	(0.188, 0.284)	0.127

	(0.2)	0.3	0.572	0.262	0.244/0.256/0.215	0.244	(0.154, 0.308)	0.156
		0.5	0.657	0.289	0.27/0.288/0.239	0.27	(0.166, 0.328)	0.175

Table 5.12: Summary of Power Calculation for Weibull(2) Distribution with $\beta_1 - \beta_0 = 0.25$ and $1 - p_C = 0.8$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort				Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Final	Bounds	Empirical
(1, 0.75)	(0.05)	0.3	0.249	0.189	0.186/0.187/0.179	0.179	(0.181, 0.205)	0.095
		0.5	0.261	0.195	0.202/0.203/0.194	0.194	(0.173, 0.198)	0.09
	(0.1)	0.3	0.366	0.23	0.225/0.228/0.211	0.225	(0.197, 0.249)	0.113
		0.5	0.410	0.255	0.247/0.251/0.232	0.247	(0.206, 0.256)	0.113
	(0.2)	0.3	0.556	0.261	0.245/0.255/0.229	0.255	(0.185, 0.295)	0.153
		0.5	0.615	0.284	0.272/0.284/0.254	0.284	(0.21, 0.321)	0.159
(1, 1)	(0.05)	0.3	0.257	0.193	0.192/0.193/0.184	0.184	(0.185, 0.211)	0.095
		0.5	0.284	0.207	0.212/0.214/0.203	0.203	(0.187, 0.216)	0.1
	(0.1)	0.3	0.380	0.221	0.229/0.233/0.215	0.229	(0.2, 0.257)	0.126
		0.5	0.423	0.244	0.255/0.26/0.239	0.255	(0.207, 0.271)	0.122
	(0.2)	0.3	0.594	0.272	0.245/0.256/0.229	0.256	(0.182, 0.295)	0.165
		0.5	0.644	0.292	0.271/0.287/0.255	0.287	(0.200, 0.331)	0.180
(1, 1.25)	(0.05)	0.3	0.265	0.195	0.195/0.196/0.187	0.187	(0.185, 0.214)	0.1
		0.5	0.302	0.227	0.219/0.221/0.209	0.209	(0.189, 0.222)	0.109

	(0.1)	0.3	0.388	0.230	0.231/0.235/0.217	0.231	(0.196, 0.260)	0.121
		0.5	0.455	0.272	0.259/0.265/0.243	(0.213, 0.279)	0.132	
	(0.2)	0.3	0.597	0.269	0.243/0.256/0.228	0.256	(0.174, 0.300)	0.149
		0.5	0.686	0.291	0.268/0.287/0.253	0.287	(0.187, 0.334)	0.182

Table 5.13: Summary of Power Calculation for Weibull(2) Distribution with $\beta_1 - \beta_0 = 0.25$ and $1 - p_C = 0.9$

Distribution	Event prop.	Biom. prop.	Full Cohort	Case-Cohort				Sub-cohort
(λ_0, λ_1)	(p_D^{00})	r_0	Empirical	Empirical	Theoretical	Final	Bounds	Empirical
(1, 0.75)	(0.05)	0.3	0.258	0.202	0.191/0.192/0.187	0.191	(0.188, 0.202)	0.100
		0.5	0.270	0.204	0.207/0.209/0.204	0.207	(0.191, 0.208)	0.093
	(0.1)	0.3	0.378	0.231	0.228/0.232/0.223	0.232	(0.216, 0.255)	0.120
		0.5	0.435	0.261	0.251/0.255/0.245	0.255	(0.225, 0.265)	0.116
	(0.2)	0.3	0.574	0.269	0.245/0.255/0.241	0.255	(0.212, 0.305)	0.158
		0.5	0.653	0.300	0.271/0.285/0.268	0.285	(0.234, 0.334)	0.171
(1, 1)	(0.05)	0.3	0.268	0.200	0.196/0.197/0.193	0.196	(0.19, 0.207)	0.091
		0.5	0.296	0.218	0.218/0.219/0.214	0.218	(0.201, 0.221)	0.094
	(0.1)	0.3	0.411	0.228	0.232/0.236/0.227	0.236	(0.218, 0.262)	0.120
		0.5	0.450	0.26	0.258/0.264/0.253	0.264	(0.236, 0.283)	0.123
	(0.2)	0.3	0.622	0.279	0.244/0.256/0.241	0.256	(0.211, 0.309)	0.165

		0.5	0.668	0.302	0.269/0.287/0.268	0.287	(0.226, 0.338)	0.178
(1, 1.25)	(0.05)	0.3	0.280	0.201	0.2/0.201/0.196	0.2	(0.195, 0.213)	0.096
		0.5	0.311	0.236	0.224/0.227/0.22	0.224	(0.204, 0.228)	0.109
	(0.1)	0.3	0.407	0.237	0.234/0.239/0.229	0.239	(0.223, 0.268)	0.130
		0.5	0.480	0.278	0.261/0.269/0.257	0.269	(0.234, 0.293)	0.137
	(0.2)	0.3	0.618	0.268	0.241/0.256/0.239	0.256	(0.204, 0.315)	0.167
		0.5	0.712	0.310	0.265/0.286/0.266	0.286	(0.214, 0.346)	0.190

In order to check how the formulae performs, we calculated the sample size formula based on (5.6) for 80% power and $p_C = 0.15$. We have considered $\beta_1 = 1$ and $\beta_0 = 0.5$ and $P(A = 1 \mid X = j) = 0.5$. Using the computed sample size, simulations were performed to get the empirical powers. The number of simulations considered were 2500. Table (5.14) summarizes the cohort size and empirical power that can be attained based on a theoretical power of 80% under various scenarios. We considered both exponential and Weibull(2) distributions to illustrate which formula (5.3, 5.4 and 5.5) results in a smaller cohort size for different sampling fractions without loss in power. It was found that for different scenarios, different sample size performed well. When the event and biomarker proportions are really low (5%), (5.5) is the only one that works well. When either one of the proportions improve, the other formulas (5.3 and 5.4) start producing better results. If one wants to be conservative, (5.9) is recommended.

Table 5.14: Summary of Sample Size and Empirical Power for Theoretical Power = 80%

Event	Biomarker	Subcohort	Exponential		Weibull(2)	
prop.(p_D^{00})	prop.	prop.	n	Empirical Power	n	Empirical Power
0.05	0.3	0.1	11940/11778/12617	0.7968/0.7828/0.8	11939/11776/12616	0.7932/0.7932/0.8192
		0.2	9356/9284/9657	0.7888/0.7944/0.8088	9357/9284/9657	0.7908/0.7916/0.8048
	0.4	0.1	10375/10207/10949	0.8108/0.7736/0.8244	10374/10205/10947	0.804/0.794/0.808
		0.2	8101/8026/8356	0.798/0.8008/0.804	8102/8027/8356	0.8036/0.7876/0.808
	0.5	0.1	9891/9704/10423	0.8016/0.7828/0.8224	9889/9702/10421	0.802/0.7908/0.816
		0.2	7695/7612/7932	0.7956/0.7984/0.814	7696/7612/7932	0.8056/0.8008/0.7988
0.1	0.3	0.1	8797/8440/9385	0.818/0.8032/0.8344	8791/8435/9378	0.8032/0.7884/0.82
		0.2	5951/5792/6212	0.816/0.8164/0.8292	5950/5791/6210	0.8064/0.8/0.8192
	0.4	0.1	7703/7332/8175	0.8244/0.8032/0.8376	7696/7326/8167	0.8148/0.7816/0.832
		0.2	5184/5019/5393	0.7916/0.8168/0.8188	5182/5017/5391	0.8016/0.784/0.8244
	0.5	0.1	7399/6987/7812	0.814/0.806/0.8416	7392/6981/7803	0.8192/0.7964/0.8396
		0.2	4953/4770/5137	0.8144/0.796/0.8168	4951/4768/5134	0.7976/0.8032/0.8268
0.2	0.3	0.1	8151/7290/8502	0.8504/0.8228/0.8708	8126/7272/8476	0.842/0.8144/0.8476
		0.2	4660/4277/4816	0.8424/0.8236/0.8548	4650/4271/4806	0.8376/0.8252/0.8412
	0.4	0.1	7236/6339/7438	0.8536/0.8156/0.8772	7209/6321/7412	0.8436/0.8108/0.852
		0.2	4111/3713/4201	0.8456/0.8112/0.8412	4101/3706/4190	0.8436/0.808/0.838

0.5	0.1	7046/6048/7140	0.8556/0.816/0.864	7016/6028/7110	0.864/0.79/0.8628
	0.2	3979/3536/4021	0.8496/0.8004/0.8508	3967/3528/4009	0.8308/0.796/0.8444

5.4 Practical Application

5.4.1 Cost Efficiency of Case-Cohort Design

In practice, one might be interested in the comparison of the cost efficiency of the test based on the case-cohort sampling scheme and a random sample from the entire cohort, assuming that the biomarker proportions (r_0, r_1), treatment proportion in each biomarker group (p_0, p_1), the failure proportion in the four groups $(p_D^{j0}, p_D^{j1},\ j = 0, 1)$ and the log-hazard of interest $(abs\{\beta_1 - \beta_0\})$ are constant. The cost efficiency is measured as the ratio of the sample sizes required in the two sampling schemes to attain the same power ϑ. Note that from (5.2), the corresponding test for a simple random sample design is $\tilde{TS}_n = \frac{\sqrt{n^*}\{\hat{\beta}_1 - \hat{\beta}_0\}}{\sqrt{\sum_j \frac{1}{r_j}\hat{\sigma}_j}}$ where $\hat{\sigma}_j^2 = \frac{1}{n_j^*}\left[\sum_{i=1}^{n_{j0}^*} \frac{e^{2\hat{\beta}_j} W(T_{i,j0})\tilde{Y}_{j1}(T_{i,j0})^2\Delta_{i,j0}}{\left(\tilde{Y}_{j0}(T_{i,j0}) + e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{i,j0})\right)^2} + \sum_{i=1}^{n_{j1}^*} \frac{W(T_{i,j1})\tilde{Y}_{j0}(T_{i,j1})^2\Delta_{i,j1}}{\left(\tilde{Y}_{j0}(T_{i,j1}) + e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{i,j1})\right)^2}\right]$, $\frac{n_j^*}{n^*} \to r_j$. Using approximations similar to those for the case-cohort power calculation, the following is the power function for simple random sample.

$$\Phi\left[\sqrt{n^*}\left(\beta_1 - \beta_0\right)\left\{\sum_{j=0}^{1} r_j^{-1}\frac{p_j(1-p_j)\left[e^{2\beta_j}p_j(1-p_D^{j1})^2 p_D^{j0} + (1-p_j)(1-p_D^{j0})^2 p_D^{j1}\right]}{\left((1-p_j)(1-p_D^{j0}) + e^{\beta_j}p_j(1-p_D^{j1})\right)^2}\right\}^{1/2} - Z_{1-\alpha}\right]. \tag{5.12}$$

Under the assumption that $\frac{p_D^{j1}}{p_D^{j0}} \approx 1$ and $\frac{m_{*j0}}{n_j} \approx (1-p_j)\left[p_j(1-p_D^{j1}) + (1-p_j)(1-p_D^{j0})\right]$, the power function for the simple random sample is

$$\Phi\left[\sqrt{n^*}\left(\beta_1 - \beta_0\right)\left\{\sum_{j=0}^{1} r_j^{-1}\frac{p_j(1-p_j)\left[e^{2\beta_j}p_j p_D^{j0} + (1-p_j)p_D^{j1}\right]}{\left((1-p_j) + e^{\beta_j}p_j\right)^2}\right\}^{1/2} - Z_{1-\alpha}\right]. \tag{5.13}$$

Note that since (5.5) was based on an approximation of $\hat{\delta}_j$ (see Appendix), it is not reflected in the power formula based on only the sub-cohort. Denoting the denominator of $\tilde{TS}_n$ for SRS by $\sigma_{\text{den}_{SRS}}$, we have the SRS sample size as $n_{SRS}^* = \frac{(Z_\vartheta + Z_{1-\alpha})^2 \times \sigma^2_{\text{den}_{SRS}}}{(\beta_1 - \beta_0)^2}$. The sample size for the case-cohort design is n_{CC}^* (5.6) Hence, the ratio of the two is given by $R = \frac{\sigma^2_{\text{den}_{SRS}}}{\psi\sigma^2_{\text{den}}\sum_{j=0}^{1}\left[r_j\left\{1 + (\frac{1-\psi}{\psi})\left(p_j p_D^{j1} + (1-p_j)p_D^{j0}\right)\right\}\right]}$ for fixed total cohort size n and assuming that the

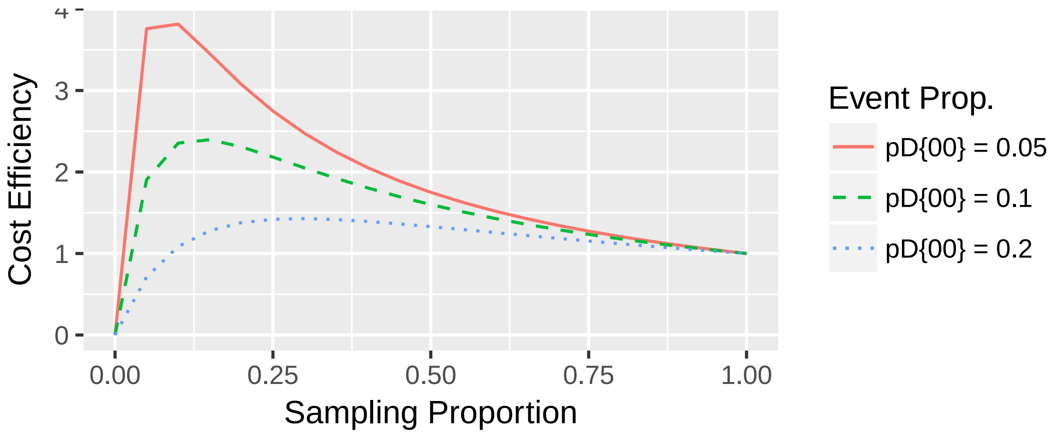

Figure 5.1: The Cost-Efficiency Curve of the Case-Cohort Design

sub-cohort proportion , ψ_j is the same in the two biomarker groups. When $R > 1$, then the case-cohort design is more cost-effective than the SRS. We used (5.9) for the sample size required for the case-cohort design because it corresponds to the most conservative power formula for the case-cohort design and the (5.13) for the simple random sample design which is the most liberal formula for the simple random sample design. We plotted the corresponding cost efficiency curve of the case-cohort design for different sub-cohort proportions and different event proportions. We considered $\beta_0 = 0.5$, $\beta_1 = 1$, $p_j = r_j = 0.5$, $p_D^{00} = p_D^{10}$, $p_D^{11} - p_D^{10} = 0.5$ and $p_D^{01} - p_D^{00} = 0.25$. From figure 5.1, it can be noted that the cost efficiency of the case-cohort study is always greater than the simple random sample design. For $p_D^{00} = 0.05$, the cost efficiency curve reaches its maximum of 3.85 at around 10% sub-cohort proportion which is higher than that achieved when $p_D^{00} = 0.1$ or 0.2. The cost efficiency curve increases, reaches a maximum and then tapers down to 1 as the sub-cohort proportion reaches 1 for all the event proportions considered.

5.4.2 Real Data Analysis

For illustration purposes, we used information from Lung Adjuvant Cisplatin Evaluation (LACE) and Cancer and Leukemia Group B (CALGB) 9633 databases (Shepherd et al. 2013) to design a two-stage study. The goal of the study is to examine whether the effect of adjuvant chemotherapy (ACT) with cisplatin/vinorelbine vs. Observation on early stage resected non-small-cell lung cancer is different for those who have KRAS-wild-type and those who have KRAS-mutated gene. There are 1543 individuals in the study, of which 780 are on adjuvant chemotherapy (ACT) with cisplatin/vinorelbine and 763 are on observation (OBS) alone. Data on the interaction effect of KRAS-mutation and treatment was available on 1422 patients as they had second primary cancer 1146 (80.6%) patients had KRAS-wild-type biomarker whereas 276 (19.4%) patients had KRAS-mutated biomarker. In the KRAS-wild-type biomarker group, 581 (50.7%) patient were in the ACT arm and 565 (49.3%) patients were in the OBS arm. Similarly, 143 (51.8%) patients are in the ACT arm and 133 (48.2%) are in the OBS arm. Further, the proportion of events (second primaries) in the KRAS-wild-type biomarker group and OBS arm, $p_D^{00} = 0.044$, that in the KRAS-wild-type group and ACT arm, $p_D^{01} = 0.065$, KRAS-mutated group and OBS arm, $p_D^{10} = 0.098$ and KRAS-mutated group and ACT arm, $p_D^{11} = 0.035$. Based on the power formula (5.11) and the corresponding sample size formula (5.6), with a significance level of 0.05 and power of 80%, we were not able to calculate a subcohort sample size to detect a hazard ratio of 0.25, which is based on the interaction term of KRAS and treatment variable in Shepherd et al. (2013). We also considered $exp(\hat{\beta}_1) = 0.32$ from the paper. For the above parameters, for power 60%, 70% and 75%, the subcohort sample sizes are 143, 347 and 837 respectively, using the most conservative power formula (5.5). However, to attain 80% power for sub-cohort proportions 0.1 and 0.2, the required cohort sizes are 2441 and 1956 respectively.

5.5 Discussion

We have considered a log-rank type test statistic for testing the interaction between the expensive biomarker and treatment for data from a two-stage study design. Explicit formulas are obtained for the calculation of power and sample size and their performances are examined for data from different distributions. When the incidence of disease is low, the simulation studies show that the two-stage design produces fairly high power, compared to that based on the full cohort. We have also demonstrated that the two-stage design is more cost efficient than the simple random sampling scheme when the sub-cohort proportion is low and it tapers to being equal when the sub-cohort proportion tends to one. These formula can be used for designing case-cohort studies as well.

The three power formulas that were derived based on different approximations of the power formula are shown to work well for different situations. The sample size/power formula are developed in the context of clinical trials but the methods can be adopted to any two-phase study including case-cohort studies. In this paper, we have considered a binary biomarker variable. A natural extension is to consider an expensive discrete biomarker covariate (number of values > 2). Further development to extend to a continuous biomarker variable and to stratified case-cohort design are worthwhile.

CHAPTER 6: FUTURE RESEARCH

In this dissertation, we studied two perspectives of case-cohort study. Chapters 3 and 4 developed estimation procedures under the case-cohort design and the generalized case-cohort design for recurrent events. Chapter 3 considered the marginal multiplicative rates model whereas Chapter 4 investigated the marginal additive rates model. Both estimation procedures can incorporate time-varying weights. They were shown to perform well in finite samples through simulations. The proposed estimators are consistent and asymptotically normally distributed.

We have assumed simple random sampling when selecting the sub-cohort and when sampling from the individuals who experienced the event but were not selected in the sub-cohort. Bernoulli sampling of the subcohort is another sampling scheme that can be considered. In our approach, we do not use all the covariate information that is available for the entire cohort. There have been studies (Jiang and Zhou 2007) that have examined the use of auxiliary covariates, available for all individuals, in the pseudo-score equation for estimating the parameters in additive rates model. It can be of interest, for future research, to extend the idea of using auxiliary variables to improve the efficiency of the estimators in recurrent events data.

The marginal models considered in this dissertation do not explicitly model the intra-subject correlation. When one is primarily interested in the effect of the risk factors on the event times, rather than the correlation among the events for each subject, the marginal model approach is the recommended method. However, in some situations, the interest may lie in a more subject-level inference including the strength and nature of dependencies within each individual's recurrent event times. In such situations, the frailty models are advocated. In our analyses, we modeled a single type of recurrent event. However, the data may comprise of different types of recurrent events. One such example is the time to hospitalization. It can be further decomposed into hospitalizations due to different reasons. Modeling the hospitalization times as originating from the same underlying process may not depict the entire picture. In such cases, it is of interest to extend the approaches outlined in Chapters 3 and 4 to incorporate multiple types of recurrent events. Kang and Cai (2009a) and Kang et al. (2013) developed methods for single time-to-event data, which can be extended to recurrent events. Further, the diagnostics under any of the two modeling approaches have not been addressed. We have developed the two methodologies for recurrent events. Section 4.5 briefly discusses using exploratory methods to examine the use of additive rates model. Nonetheless, rigorous tests of the error processes to study the fit of the data require further consideration.

Chapter 5 considered the design aspect of the two-stage design for a single time to event when the interest is in the interaction between two dichotomized variables. A log-rank type test statistic was considered for testing the interaction between the expensive biomarker and treatment for the analysis. Explicit formulas were developed for power and sample size calculation and their performances are examined through simulation studies. Further, the cost efficiency of the case-cohort design is shown as compared with a simple random sampling scheme.

The formulae we developed considered dichotomized biomarker. A natural extension is to consider an expensive discrete biomarker covariate (number of levels > 2). The test statistics to examine whether the effects are equal in all of the groups need to be developed. Furthermore, tests need to be developed for continuous biomarker variables. Even though it is not common, interest may lie in testing the equality of multiple treatment effects in different biomarker groups. In such a case, we will extend the proposed tests to incorporate multiple treatment levels. Another extension is to develop tests for power/sample size calculation for the interaction term of biomarker and treatment for stratified two-stage design. Hu et al. (2014) studied the tests for only the treatment effect. Implementation of those tests to interaction effect of treatment and expensive biomarker needs further study.

APPENDIX A: TECHNICAL DETAILS FOR CHAPTER 3

Here is the Appendix containing the proofs of the theorems 1 and 2 in Chapter 3.

A.1 Regularity Conditions

(i) $(T_i^*, C_i, Z_i(t))$ $\forall i = 1, 2, \ldots, n$ are independent and identically distributed.

(ii) $P(Y(\tau) > 0) > 0$ and $N_i(\tau)$ $(\forall i = 1, 2, \ldots, n)$ are bounded by a constant

(iii) $\mid Z_i(0) \mid + \int_0^\tau \mid dZ_i(u) \mid < C_z < \infty$ almost surely for some constant C_z

(iv) The matrix $A(\beta_0) = \int_0^\tau \vartheta(\beta_0, t) s^{(0)}(\beta_0, t) d\mu_0(t)$ is positive definite.

(v) (Finite Interval) $\int_0^\tau d\mu_0(t) < \infty$

(vi) (Asymptotic Stability) There exists a neighborhood $\mathcal{B}$ of β_0 that satisfies the following:

(a) There exists functions $s^{(0)}(\beta, t), s^{(1)}(\beta, t)$ and $s^{(2)}(\beta, t)$ defined on $\mathcal{B} \times [0, \tau]$ such that

$$\sup_{t \in [0,\tau], \beta \in \mathcal{B}} \|S^{(d)}(\beta, t) - s^{(d)}(\beta, t)\| \xrightarrow{P} 0$$

(b) There exists a matrix $Q(\beta)$ such that

$$\frac{1}{n} \sum_{i=1}^{n} Var(M_{\tilde{Z},i}(\beta_0)) \xrightarrow{P} Q(\beta_0)$$

(vii) (Asymptotic Regularity) For all $\beta \in \mathcal{B}, t \in [0, \tau]$, $s^{(1)}(\beta, t) = \frac{\delta}{\delta\beta} s^{(0)}(\beta, t)$, $s^{(2)}(\beta, t) = \frac{\delta}{\delta\beta} s^{(1)}(\beta, t) = \frac{\delta^2}{\delta\beta\delta\beta'} s^{(0)}(\beta, t)$ where $s^{(d)}(\beta, t)$ are continuous functions of $\beta \in \mathcal{B}$, uniformly in $t \in [0, \tau]$ and bounded on $\mathcal{B} \times [0, \tau]$ and $s^{(0)}(\beta, t)$ is bounded away from zero on $\mathcal{B} \times [0, \tau]$ The following conditions are pertaining to the asymptotic convergences of case-cohort sampling design.

(viii) (Non-trivial subcohort and cases) $\tilde{\alpha} = \frac{\tilde{n}}{n} \xrightarrow{P} \alpha \in (0, 1)$, $\tilde{q} \to q$, $\frac{n_c}{n} \xrightarrow{P} p \in (0, 1)$ as $n \to \infty$ where n_c is the number of individuals in the cohort who experienced at least one event.

(ix) (Asymptotic Normality of subcohort averages at β_0) For $\epsilon > 0$

$$\sup_t \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\{e^{\beta Z_i(t)}\}^2 \mathbf{1}(n^{-1/2}Y_i(t)e^{\beta Z_i(t)} > \epsilon) \xrightarrow{P} 0$$

$$\sup_t \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\{|Z_i(t)\|\}^2\{e^{\beta Z_i(t)}\}^2 \mathbf{1}(n^{-1/2}Y_i(t)|Z_i(t)\|e^{\beta Z_i(t)} > \epsilon) \xrightarrow{P} 0$$

(x) (Asymptotic Normality of samples) As $n \to \infty$

$$\frac{1}{n}\sup_{i,t\in[0,\tau]} exp\left(2\beta' Z_i(t)\right) \xrightarrow{P} 0, \quad \frac{1}{n}\sup_{i,t\in[0,\tau]} \|Z_i(t)\|^2 exp\left(2\beta' Z_i(t)\right) \xrightarrow{P} 0$$

(xi) (Asymptotic stability)As $n \to \infty$, we have the following

(a) There exists a positive definite matrix, $V^I(\beta_0)$, such that

$$var\left[n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \left(R_i(\beta_0,t) - \frac{Y_i(t)E\{1-\Delta_i\}R_i(\beta_0,t)}{E\{(1-\Delta_1)Y_1(t)\}}\right)d\mu_0(t)\right] \xrightarrow{P} V^I(\beta_0),$$

where $R_i(\beta_0,t) = Y_i(t)\tilde{Z}_i(t)e^{\beta Z_i(t)}$.

(b) There is a positive definite matrix, $V^{II}(\beta_0)$ such that

$$var\left[n^{-1/2}\sum_{i=1}^{n}\left(M_{\tilde{Z},i}(\beta_0) - \int_0^\tau \frac{Y_i(t)E\left\{dM_{\tilde{Z},1}(\beta_0,t) \mid \Delta_1 = 1, \xi_1 = 0\right\}}{E\{Y_1(t) \mid \Delta_1 = 1\}}\right)\right.$$
$$\left. \mid \Delta_i = 1, \xi_i = 0\right] \xrightarrow{P} V^{II}(\beta_0),$$

where $M_{\tilde{Z},i}(\beta_0) = \int_0^\tau \tilde{Z}_i(t)dM_i(t)$, $dM_{\tilde{Z},i}(\beta_0,t) = \tilde{Z}_i(t)dM_i(t)$ and $\tilde{Z}_i(t) = Z_i(t) - e(\beta_0,t)$.

Lemma A1 :

Let $\varphi = (\varphi_1, \varphi_2, \ldots, \varphi_n)$ be a random vector containing n^* ones and $n - n^*$ zeroes, with each permutation equally likely. Let $B_i(t)$ be independent and identically distributed real-valued processes on [0, τ] with E$(B_i(t)) = \mu_B(t)$, $var(B_i(0)) < \infty$ and $var(B_i(t)) < \infty$. Let

$B(t) = (B_1(t), B_2(t), \ldots, B_n(t))$ and φ are independent and let us suppose that all paths of $B_i(t)$ have finite variation. Then $n^{-1/2}\sum_{i=1}^{n}\varphi_i\left[B_i(t) - \mu_B(t)\right]$ converges weakly in $l^{\infty}[0,\tau]$ to a zero-mean Gaussian process and $n^{-1}\sum_{i=1}^{n}\varphi_i\left[B_i(t) - \mu_B(t)\right] \xrightarrow{P} 0$ uniformly in t. This lemma was stated in (Kang and Cai 2009a).

Lemma A2 :

Let $W_n(t)$ and $G_n(t)$ be two sequences of bounded processes. For some constant, τ let us assume that the following conditions hold.

(i) $\sup_{0\le t\le\tau}\|W_n(t) - W(t)\| \xrightarrow{P} 0$ for some bounded process, $W(t)$,

(ii) $W_n(t)$ is monotone on $[0,\tau]$ and

(iii) $G_n(t)$ converges to a zero-mean process with continuous sample paths. Then

$$\sup_{0\le s\le\tau}\|\int_0^s (W_n(t) - W(t))dG_n(t)\| \xrightarrow{P} 0, \quad \sup_{0\le s\le\tau}\|\int_0^s (G_n(t) - G(t))dW_n(t)\| \xrightarrow{P} 0.$$

This was stated in Kang and Cai (2009b).

First we will look at the asymptotic properties of the time-varying sampling weights. More specifically, $\hat{\alpha}(t) = \frac{\sum_{i=1}^{n}(1-\Delta_i)\xi_i Y_i(t)}{\sum_{i=1}^{n}(1-\Delta_i)Y_i(t)}$ and $\hat{q}(t) = \frac{\sum_{i=1}^{n}\Delta_i(1-\xi_i)\eta_i Y_i(t)}{\sum_{i=1}^{n}\Delta_i(1-\xi_i)Y_i(t)}$. Looking at the Taylor series expansion of $\hat{\alpha}(t)^{-1}$ around $\tilde{\alpha}^{-1}$, we have

$$\hat{\alpha}(t)^{-1} - \tilde{\alpha}^{-1} = -\frac{1}{\alpha^*(t)^2}(\hat{\alpha}(t) - \tilde{\alpha}) = \frac{\tilde{\alpha}}{\alpha^*(t)^2}\frac{1}{\sum_{i=1}^{n}(1-\Delta_i)Y_i(t)}\left[\sum_{i=1}^{n}(1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})Y_i(t)\right],$$

where $\alpha^*(t)$ is on the line between $\hat{\alpha}(t)$ and $\tilde{\alpha}$. Now by the Glivenko-Cantelli theorem, one can show that $\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)Y_i(t)$ converges to $E\left((1-\Delta_i)Y_i(t)\right)$. $(1-\Delta_i)Y_i(t)$ is bounded and monotone functions of t. They are also independent of ξ_i. Hence, by Lemma A2, $n^{-1/2}\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})\left[(1-\Delta_i)Y_i(t)\right]$ converges weakly to a zero-mean Gaussian process (Kang and Cai 2009b). This implies that $\frac{1}{n}\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})\left[(1-\Delta_i)Y_i(t)\right] \xrightarrow{P} 0$ uniformly in t. Further, $\hat{\alpha}(t)$, $\alpha^*(t)$ and $\tilde{\alpha}$ converges to the same limit. Using Slutsky's theorem, we have

$$n^{1/2}(\hat{\alpha}(t)^{-1} - \tilde{\alpha}^{-1})$$

$$= n^{1/2}\frac{\tilde{\alpha}}{\alpha^*(t)^2}\frac{1}{\sum_{i=1}^n(1-\Delta_i)Y_i(t)}\left[\sum_{i=1}^n(1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})Y_i(t)\right]$$

$$= \frac{1}{\tilde{\alpha}E((1-\Delta_i)Y_i(t))}\left[n^{-1/2}\sum_{i=1}^n(1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})Y_i(t)\right]$$

$$+ \tilde{\alpha}\left(\frac{1}{\alpha^*(t)^2}\frac{1}{\sum_{i=1}^n(1-\Delta_i)Y_i(t)} - \frac{1}{\tilde{\alpha}^2E((1-\Delta_i)Y_i(t))}\right)\left[n^{-1/2}\sum_{i=1}^n(1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})Y_i(t)\right]$$

$$= \frac{1}{\tilde{\alpha}E((1-\Delta_i)Y_i(t))}\left[n^{-1/2}\sum_{i=1}^n(1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})Y_i(t)\right] + o_P(1). \tag{6.1}$$

Similarly,

$$\hat{q}(t)^{-1} - \tilde{q}^{-1} = -\frac{1}{q^*(t)^2}(\hat{q}(t)-\tilde{q}) = \frac{\tilde{q}}{q^*(t)^2}\frac{1}{\sum_{i=1}^n\Delta_i(1-\xi_i)Y_i(t)}\left[\sum_{i=1}^n\Delta_i(1-\xi_i)(1-\frac{\eta_i}{\tilde{q}})Y_i(t)\right],$$

where $q^*(t)$ is on the line between $\hat{q}(t)$ and $\tilde{q}$. Proceeding as before, we can see that $\hat{q}(t)$, $q^*(t)$ and $\tilde{q}$ converges to the same limit. Using Slutsky's theorem, we have

$$n^{1/2}(\hat{q}(t)^{-1} - \tilde{q}^{-1}) = \frac{1}{\tilde{q}E(\Delta_i(1-\xi_i)Y_i(t))}\left[n^{-1/2}\sum_{i=1}^n\Delta_i(1-\xi_i)(1-\frac{\eta_i}{\tilde{q}})Y_i(t)\right] + o_P(1). \tag{6.2}$$

A.2 Proof of Theorem 1

Let us define $U_n(\beta) = \frac{1}{n}\hat{U}^{II}(\beta)$. Based on similar arguments, as in Foutz(1977), the consistency of $\hat{\beta}^{II}$ can be shown by proving the following :

- $\frac{\delta}{\delta\beta'}U_n(\beta)$ exists and is continuous in an open neighborhood of β_0 in $\mathcal{B}$
- $\frac{\delta}{\delta\beta_0'}U_n(\beta_0)$ is negative definite w.p. $\to 1$ as $n\to\infty$
- - $\frac{\delta}{\delta\beta'}U_n(\beta) \xrightarrow{P} A(\beta_0)$ uniformly for β in a neighborhood of β_0
- $U_n(\beta) \xrightarrow{P} 0$

Taking derivative of the expression, we get

$$\frac{\delta}{\delta\beta'}U_n(\beta) = -\frac{1}{n}\sum_{i=1}^n\int_0^\tau w_i^{II}(t)\frac{\hat{S}^{(2)}(\beta,t)\hat{S}^{(0)}(\beta,t) - \hat{S}^{(1)}(\beta,t)^{\otimes 2}}{\hat{S}^{(0)}(\beta,t)^2}dN_i(t)$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau w_i^{II}(t)\hat{V}(\beta,t)dN_i(t) = -\int_0^\tau \hat{V}(\beta,t)d\bar{N}_w(t), \tag{6.3}$$

where $d\bar{N}_w(t) = \frac{1}{n}\sum_{i=1}^n w_i^{II}(t)N_i(t)$. Define $d\bar{M}_w(t) = \frac{1}{n}\sum_{i=1}^n w_i^{II}(t)M_i(t)$, $d\bar{N}(t) = \frac{1}{n}\sum_{i=1}^n dN_i(t)$, $d\bar{M}(t) = \frac{1}{n}\sum_{i=1}^n dM_i(t)$. Note that,

$$dM_i(t) = dN_i(t) - Y_i(t)e^{\beta Z_i(t)}d\mu_0(t) \Rightarrow \sum_{i=1}^n w_i^{II}(t)dM_i(t) = \sum_{i=1}^n w_i^{II}(t)dN_i(t) - n\hat{S}^{(0)}(\beta,t)d\mu_0(t)$$

Now, we need to show that this goes to $A(\beta)$ in probability. Using the above formula, we have

$$\begin{aligned}
&\|-\frac{\delta}{\delta\beta'}U_n(\beta) - A(\beta)\| \\
&= \|\int_0^\tau \hat{V}(\beta,t)d\bar{N}_w(t) - \int_0^\tau \vartheta(\beta,t)s^{(0)}(\beta,t)d\mu_0(t)\| \\
&= \|\int_0^\tau \left(\hat{V}(\beta,t) - \vartheta(\beta,t)\right)d\bar{N}_w(t) + \int_0^\tau \vartheta(\beta,t)d\bar{M}_w(t) \\
&+ \int_0^\tau \vartheta(\beta,t)\left\{\hat{S}^{(0)}(\beta,t) - s^{(0)}(\beta,t)\right\}d\mu_0(t)\| \\
&\le \|\int_0^\tau \left(\hat{V}(\beta,t) - \vartheta(\beta,t)\right)d\bar{N}_w(t)\| + \|\int_0^\tau \vartheta(\beta,t)d\bar{M}_w(t)\| \\
&+ \|\int_0^\tau \vartheta(\beta,t)\left\{\hat{S}^{(0)}(\beta,t) - s^{(0)}(\beta,t)\right\}d\mu_0(t)\| \\
&\le \|\int_0^\tau \left(\hat{V}(\beta,t) - \vartheta(\beta,t)\right)d\bar{N}(t)\| + \|\frac{1}{n}\sum_{i=1}^n\int_0^\tau \left(\hat{V}(\beta,t) - \vartheta(\beta,t)\right)(w_i^{II}(t) - 1)dN_i(t)\| \\
&+ \|\int_0^\tau \vartheta(\beta,t)d\bar{M}(t)\| + \frac{1}{n}\|\sum_{i=1}^n\int_0^\tau \vartheta(\beta,t)(w_i^{II}(t) - 1)dM_i(t)\| \\
&+ \|\int_0^\tau \vartheta(\beta,t)\left\{\hat{S}^{(0)}(\beta,t) - s^{(0)}(\beta,t)\right\}d\mu_0(t)\|. \qquad (6.4)
\end{aligned}$$

We need to show that each of these terms converges to zero as $n \to \infty$. The first term can be rewritten as

$$\begin{aligned}
&\|\int_0^\tau \left(\hat{V}(\beta,t) - \vartheta(\beta,t)\right)d\bar{N}(t)\| \\
&\le \|\int_0^\tau \left(\frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} - \frac{s^{(2)}(\beta,t)}{s^{(0)}(\beta,t)}\right)d\bar{N}(t)\| + \|\int_0^\tau \left([\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}]^{\otimes 2} - [\frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}]^{\otimes 2}\right)d\bar{N}(t)\|
\end{aligned}$$

$$= \| \int_0^\tau \left\{ \left(\frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} - \frac{S^{(2)}(\beta,t)}{S^{(0)}(\beta,t)} \right) + \left(\frac{S^{(2)}(\beta,t)}{S^{(0)}(\beta,t)} - \frac{s^{(2)}(\beta,t)}{s^{(0)}(\beta,t)} \right) \right\} d\bar{N}(t) \|$$
$$+ \| \int_0^\tau \left\{ \left([\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}]^{\otimes 2} - [\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]^{\otimes 2} \right) + \left([\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]^{\otimes 2} - [\frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}]^{\otimes 2} \right) \right\} d\bar{N}(t) \|.$$

By the regularity condition (vi), we have noted the asymptotic properties of $S^{(d)}(\beta,t)$ uniformly in t,

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}} \|S^{(d)}(\beta,t) - s^{(d)}(\beta,t)\| \xrightarrow{P} 0.$$

In order to show that the first term of Equation (6.4) converges to 0 as $n \to \infty$, it is sufficient to show that

$$\sup_{\beta\in\mathcal{B},t\in[0,\tau]} \|\hat{S}^{(d)}(\beta,t) - S^{(d)}(\beta,t)\| \xrightarrow{P} 0 \text{ for all d = 0, 1, 2.}$$

We can write

$$\begin{aligned}
\hat{S}^{(d)}(\beta,t) - S^{(d)}(\beta,t) &= \frac{1}{n}\sum_{i=1}^n \frac{(1-\Delta_i)\xi_i}{\hat{\alpha}(t)} Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)} + \frac{1}{n}\sum_{i=1}^n \frac{\Delta_i(1-\xi_i)\eta_i}{\hat{q}(t)} Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)} \\
&+ \frac{1}{n}\sum_{i=1}^n \Delta_i \xi_i Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)} - \frac{1}{n}\sum_{i=1}^n Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)} \\
&= \frac{1}{n}\sum_{i=1}^n (1-\Delta_i)(\frac{\xi_i}{\hat{\alpha}(t)} - 1) Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)} \\
&+ \frac{1}{n}\sum_{i=1}^n \Delta_i(1-\xi_i)(\frac{\eta_i}{\hat{q}(t)} - 1) Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)} \\
&= \frac{1}{n}\sum_{i=1}^n (1-\Delta_i)(\frac{\xi_i}{\tilde{\alpha}} - 1) Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)} \\
&+ \frac{1}{n}\sum_{i=1}^n (1-\Delta_i)\xi_i(\frac{1}{\hat{\alpha}(t)} - \frac{1}{\tilde{\alpha}}) Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)} \\
&+ \frac{1}{n}\sum_{i=1}^n \Delta_i(1-\xi_i)(\frac{\eta_i}{\tilde{q}(t)} - 1) Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)} \\
&+ \frac{1}{n}\sum_{i=1}^n \Delta_i(1-\xi_i)\eta_i(\frac{1}{\hat{q}(t)} - \frac{1}{\tilde{q}}) Y_i(t) Z_i(t)^{\otimes d} e^{\beta Z_i(t)}.
\end{aligned}$$

Therefore,

$$\left\| \hat{S}^{(d)}(\beta,t) - S^{(d)}(\beta,t) \right\|$$

$$\leq \frac{1}{n}\left\|\sum_{i=1}^{n}(1-\Delta_i)(\frac{\xi_i}{\tilde{\alpha}}-1)Y_i(t)Z_i(t)^{\otimes d}e^{\beta Z_i(t)}\right\| + \frac{1}{n} \mid (\frac{1}{\hat{\alpha}(t)}-\frac{1}{\tilde{\alpha}}) \mid \times$$

$$\left\|\sum_{i=1}^{n}(1-\Delta_i)\xi_i Y_i(t)Z_i(t)^{\otimes d}e^{\beta Z_i(t)}\right\| + \frac{1}{n}\left\|\sum_{i=1}^{n}\Delta_i(1-\xi_i)(\frac{\eta_i}{\tilde{q}(t)}-1)Y_i(t)Z_i(t)^{\otimes d}e^{\beta Z_i(t)}\right\|$$

$$+ \mid \frac{1}{\hat{q}(t)}-\frac{1}{\tilde{q}} \mid \frac{1}{n}\left\|\sum_{i=1}^{n}\Delta_i(1-\xi_i)\eta_i Y_i(t)Z_i(t)^{\otimes d}e^{\beta Z_i(t)}\right\|$$

$$\leq \left\|\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)(\frac{\xi_i}{\tilde{\alpha}}-1)Z_i(t)^{\otimes d}Y_i(t)e^{\beta Z_i(t)}\right\|$$

$$+ \frac{1}{n} \mid (\frac{1}{\hat{\alpha}(t)}-\frac{1}{\tilde{\alpha}}) \mid \sum_{i=1}^{n}(1-\Delta_i)\xi_i Y_i(t)\left\|Z_i(t)^{\otimes d}\right\| e^{\beta Z_i(t)}$$

$$+ \left\|\frac{1}{n}\sum_{i=1}^{n}\Delta_i(1-\xi_i)(\frac{\eta_i}{\tilde{q}}-1)Z_i(t)^{\otimes d}Y_i(t)e^{\beta Z_i(t)}\right\|$$

$$+ \frac{1}{n} \mid (\frac{1}{\hat{q}(t)}-\frac{1}{\tilde{q}}) \mid \sum_{i=1}^{n}\Delta_i(1-\xi_i)\eta_i Y_i(t)\left\|Z_i(t)^{\otimes d}\right\| e^{\beta Z_i(t)}. \tag{6.5}$$

Note that $(1-\Delta_i)Y_i(t)Z_i(t)^{\otimes d}e^{\beta Z_i(t)}$ is bounded with total variation being finite in $[0,\tau]$ by the regularity condition (iii). Using Lemma A1 and the fact that $E\left((1-\Delta_i)(\frac{\xi_i}{\tilde{\alpha}}-1)Y_i(t)Z_i(t)^{\otimes d} \times e^{\beta Z_i(t)}\right)=0$, $E\left(\Delta_i(1-\xi_i)(\frac{\eta_i}{\tilde{q}}-1)Y_i(t)Z_i(t)^{\otimes d}e^{\beta Z_i(t)}\right)=0$, we have the first term and third terms of Equation (6.5) converge to zero in probability uniformly in t. For the second term, we can note the asymptotic distribution of $\hat{\alpha}(t)$ proved earlier in equation (6.1) and the fact that $\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\xi_i Y_i(t)\left|Z_i(t)^{\otimes d}\right| e^{\beta Z_i(t)}$ converges to a finite quantity in probability uniformly in t and β by Lemma A1. Similarly, based on equation (6.2) and Lemma A1, we have the fourth term to converge to zero in probability uniformly in t and β. Combining these results, $\hat{S}^{(d)}(\beta,t)$ and $S^{(d)}(\beta,t)$ is shown to converge to the same limit. Therefore, we have

$$\sup_{t\in[0,\tau],\beta\in\mathcal{B}}\left\|\hat{V}(\beta,t)-\vartheta(\beta,t)\right\| \xrightarrow{P} 0 \text{ as n } \to\infty$$

by using Slutsky's theorem, the fact that $\sup_{t\in[0,\tau],\beta\in\mathcal{B}}\left\|\hat{S}^{(d)}(\beta,t)-s^{(d)}(\beta,t)\right\| \xrightarrow{P} 0$ and since $s^{(0)}(\beta,t)$ is bounded away from zero on $\mathcal{B}\times[0,\tau]$ by condition (vii).

From the regularity condition (i), we can see that $P(\sup_{i,t\in[0,\tau]} N_i(t)<c)=1$ and so, it can be easily seen that $\bar{N}(t)$ has bounded variation. Thus, the first term of (6.4) converges to

zero in probability. Next, we look at $n^{-1/2}\bar{M}(t)$, which is a sum of independent and identically distributed zero-mean random variables, for fixed t. By CLT, one can show that $n^{1/2}\bar{M}(t)$ converges to a zero-mean Gaussian process , say $\mathcal{W}_M$ (Lin et al. 2000). Further, $M_i(t)$ can be expressed as the difference of two monotone functions in t. As monotone functions have pseudo-dimension 1 (Pollard 1990, Bilias et al. 1997, Lin et al. 2000), the processes $\{M_i(t) \mid i = 1, 2, \ldots, n\}$ are manageable. From the functional Central Limit Theorem (Pollard 1990), we can conclude that $n^{-1/2}\bar{M}(t)$ is tight and converges weakly to $\mathcal{W}_M$ (Van Der Vaart and Wellner 1996). Furthermore, we can show that $E(\mathcal{W}_M(t) - \mathcal{W}_M(s))^4 \leq K(\mu_0(t) - \mu_0(s))^2$ for some constant K. Specifically, $E(\mathcal{W}_M(t) - \mathcal{W}_M(s))^4 = 3\left(E(\mathcal{W}_M(t) - \mathcal{W}_M(s))^2\right)^2$ since, $\mathcal{W}_M(t)$ is a zero-mean normal random variable, for fixed t. $(E(\mathcal{W}_M(t) - \mathcal{W}_M(s))^2 = E(\mathcal{W}_M(t))^2 + E(\mathcal{W}_M(s))^2 - 2E(\mathcal{W}_M(t)\mathcal{W}_M(s))$ which in turn can be written $= E(\mathcal{W}_M(t))^2 - E(\mathcal{W}_M(s))^2$ for $s \leq t$. Next, note that $E(\mathcal{W}_M(t))^2 = E(M_i(t))^2 = E(\int_0^t Y_i(u) exp\{\beta_0' Z_i(u)\} d\mu_0(u))$ and so, $E(\mathcal{W}_M(t))^2 - E(\mathcal{W}_M(s))^2 \leq \sqrt{K}(\mu_0(t) - \mu_0(s)) \Rightarrow E(\mathcal{W}_M(t) - \mathcal{W}_M(s))^4 \leq K(\mu_0(t) - \mu_0(s))^2$. Then it follows from the Kolmogorov- Centsov Theorem (Karatzas and Shereve 1988) that $\mathcal{W}_M$ has continuous sample paths. Using Lemma A2, we get the third term of (6.4) converging to 0 since $\vartheta(\beta, t)$ is a monotone function of t. We can conclude that the last term of (6.4) converges to zero using the uniform convergence of $\hat{S}^{(d)}(\beta, t)$ to $s^{(d)}(\beta, t)$. For the second term of equation (6.4) we can write

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(w_i^{II}(t) - 1)dN_i(t) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{\Delta_i\xi_i + \frac{(1-\Delta_i)\xi_i}{\hat{\alpha}(t)} + \frac{\Delta_i(1-\xi_i)\eta_i}{\hat{q}(t)} - 1\right\} dN_i(t) \\
&= \frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\left\{\frac{\xi_i}{\hat{\alpha}(t)} - 1\right\} dN_i(t) + \frac{1}{n}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\left\{\frac{\eta_i}{\hat{q}(t)} - 1\right\} dN_i(t) \\
&= \frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\left\{\frac{\xi_i}{\tilde{\alpha}} - 1\right\} dN_i(t) + \frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\xi_i\left\{\frac{1}{\hat{\alpha}(t)} - \frac{1}{\tilde{\alpha}}\right\} dN_i(t) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\left\{\frac{\eta_i}{\tilde{q}} - 1\right\} dN_i(t) + \frac{1}{n}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\eta_i\left\{\frac{1}{\hat{q}(t)} - \frac{1}{\tilde{q}}\right\} dN_i(t) \\
&= \frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\left\{\frac{\xi_i}{\tilde{\alpha}} - 1\right\} dN_i(t)
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\xi_i\left\{\frac{1}{\tilde{\alpha}E((1-\Delta_1)Y_1(t)}\left[\frac{1}{n}\sum_{j=1}^{n}(1-\Delta_j)(1-\frac{\xi_j}{\tilde{\alpha}})Y_j(t)\right]\right\}dN_i(t)\\
&+\frac{1}{n}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\left\{\frac{\eta_i}{\tilde{q}}-1\right\}dN_i(t)\\
&+\frac{1}{n}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\eta_i\left\{\frac{1}{\tilde{q}E(\Delta_1(1-\xi_1)Y_1(t))}\left[\frac{1}{n}\sum_{j}\Delta_j(1-\xi_j)(1-\frac{\eta_j}{\tilde{q}})Y_j(t)\right]\right\}dN_i(t)+o_P(1)\\
&=\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\left\{\frac{\xi_i}{\tilde{\alpha}}-1\right\}dN_i(t)\\
&+\frac{1}{n}\sum_{j=1}^{n}\frac{1}{E((1-\Delta_1)Y_1(t)}(1-\Delta_j)(1-\frac{\xi_j}{\tilde{\alpha}})Y_j(t)\left[\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\xi_i}{\tilde{\alpha}}dN_i(t)\right]\\
&+\frac{1}{n}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\left\{\frac{\eta_i}{\tilde{q}}-1\right\}dN_i(t)\\
&+\frac{1}{n}\sum_{j=1}^{n}\frac{1}{E(\Delta_1(1-\xi_1)Y_1(t))}\Delta_j(1-\xi_j)(1-\frac{\eta_j}{\tilde{q}})Y_j(t)\left[\frac{1}{n}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\frac{\eta_i}{\tilde{q}}dN_i(t)\right]+o_P(1)
\end{aligned}
\tag{6.6}
$$

Since, $N_i(t)$ is bounded for all t, we have the first and third term of equation(6.6) to be also bounded. $\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i}{\tilde{\alpha}}\left[(1-\Delta_i)dN_i(t)-E((1-\Delta_1)dN_1(t))\right]$ and $\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i}{\tilde{q}}(1-\xi_i)$ $[\Delta_i dN_i(t)-E(\Delta_1 dN_1(t))]$ converges to zero in probability, uniformly in t respectively, by Lemma A1. Using Lemma from Kulich and Lin (2004) we can note that $E((1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})dN_i(t)) = E((1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})dY_i(t)) = E(\Delta_i(1-\xi_i)(1-\frac{\eta_i}{\tilde{q}})dN_i(t)) = E(\Delta_i(1-\xi_i)(1-\frac{\eta_i}{\tilde{q}})dY_i(t)) = 0$. Hence, we have all the terms of (6.6) converging to zero in probability uniformly in t. Hence, the second term of (6.4) converges to 0 in probability. Proceeding as earlier, we can show that $n^{-1/2}\sum_{i=1}^{n}\int\{w_i^{II}(t)-1\}dM_i(t)$ converges to a zero-mean Guassian process, since, $E\left((1-\Delta_i)(\frac{\xi_i}{\tilde{\alpha}}-1)dM_i(t)\right)=0$ and $E\left(\Delta_i(1-\xi_i)(\frac{\eta_i}{\tilde{q}}-1)dM_i(t)\right)=0$. This implies that the fourth terms of (6.4) converges to zero in probability. Hence, we have

$$
-\frac{\delta U_n(\beta)}{\delta\beta'}\xrightarrow{P}A(\beta)\text{ as n }\to\infty\text{ uniformly in }\beta\in\mathcal{B}.
$$

Next, we want to look at $n^{1/2}U_n(\beta)$.

$$
\begin{aligned}
n^{1/2}U_n(\beta) &= n^{-1/2}\sum_{i=1}^{n}\int_0^\tau w_i^{II}(t)\left[Z_i(t)-\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right]dN_i(t)\\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^\tau w_i^{II}(t)\left[Z_i(t)-\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right]dN_i(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^\tau w_i^{II}(t)\left[Z_i(t)-\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right]Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^\tau w_i^{II}(t)\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^\tau w_i^{II}(t)\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t}-\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right]dM_i(t)\\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \{w_i^{II}(t)-1\}\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t}-\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right]dM_i(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \{w_i^{II}(t)-1\}\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t}-\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right]dM_i(t). \qquad (6.7)
\end{aligned}
$$

The first term converges to a zero-mean Gaussian process based on Lin et al. (2000) paper. Further, we can replace it with $n^{-1/2}\sum_{i=1}^{n}M_{\tilde{Z},i}(\beta)=n^{-1/2}\sum_{i=1}^{n}\int_0^\tau(Z_i(t)-e(\beta,t))dM_i(t)$ which is asymptotically equivalent to it (since, $\tilde{Z}_i(t)=Z_i(t)-e(\beta,t); e(\beta,t)=\frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}$). One can note that for each t, $n^{1/2}\bar{M}(t)$ is the sum of zero-mean i.i.d. random variables and can be written as the difference of two monotone functions. Since, each of the functions are bounded, $n^{1/2}\bar{M}(t)$ is also a bounded process. Moreover, from example 2.11.16 of Van Der Vaart and Wellner (1996) we have that $n^{-1/2}\bar{M}(t)$ converges weakly to a zero-mean Gaussian process, $\mathcal{W}_M$ with $E\left(\mathcal{W}_M(t)-\mathcal{W}_M(s)\right)^4 < C\left[\mu_0(t)-\mu_0(s)\right]^2$ for some positive constant, C. By the Kolmogov-Centsov Theorem (Karatzas and Shereve 1988), $\mathcal{W}_M(t)$ has continuous sample paths. Also, from regularity condition (vi), we have that $S^{(1)}(\beta,t)$ and $S^{(0)}(\beta,t)$ are of bounded variations and $S^{(0)}(\beta,t)$ is bounded away from zero such that $\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}$

also has bounded variation and can be written as the difference of two monotone functions, specifically, $\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)} = G_1(t)-G_2(t)$ where $G_i(t)$ are non-negative, monotone in t and bounded. Similarly, we have $\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}$ has bounded variation and is equal to $H_1(t) - H_2(t)$ where $H_i(t)$ are non-negative, monotone in $\mathbf{t}$ and bounded. One can also note that since $\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}$ converges uniformly to $\frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}$ and regularity condition (vi) implies that $\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}$ and $\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}$ have the same limit uniformly. Based on Lemma A2, we can rewrite the third quantity on the RHS of (6.7) as

$$\int_0^\tau \left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right] n^{-1/2}\sum_{i=1}^n dM_i(t)$$
$$= \int_0^\tau \left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t} - \frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}\right] n^{-1/2}\sum_{i=1}^n dM_i(t) - \int_0^\tau \left[\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t} - \frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}\right] n^{-1/2}\sum_{i=1}^n dM_i(t)$$
$$\xrightarrow{P} 0.$$

Hence, the third term of equation (6.7) converges to zero in probability. Next, we look at the fourth term of the equation.

$$n^{-1/2}\sum_{i=1}^n \int_0^\tau \{w_i^{II}(t)-1\}\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right] dM_i(t)$$
$$= n^{-1/2}\sum_{i=1}^n \int_0^\tau \{\xi_i\Delta_i + (1-\Delta_i)\frac{\xi_i}{\hat{\alpha}(t)} + \Delta_i(1-\xi_i)\frac{\eta_i}{\hat{q}(t)} - 1\}\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right] dM_i(t)$$
$$= n^{-1/2}\sum_{i=1}^n \int_0^\tau (1-\Delta_i)\{\frac{\xi_i}{\hat{\alpha}(t)} - 1\}\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right] dM_i(t)$$
$$+ n^{-1/2}\sum_{i=1}^n \int_0^\tau \Delta_i(1-\xi_i)\{\frac{\eta_i}{\hat{q}(t)} - 1\}\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right] dM_i(t)$$
$$= n^{-1/2}\sum_{i=1}^n \int_0^\tau (1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}} - 1\}\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right] dM_i(t)$$
$$+ n^{-1/2}\sum_{i=1}^n \int_0^\tau (1-\Delta_i)\xi_i\{\frac{1}{\hat{\alpha}(t)} - \frac{1}{\tilde{\alpha}}\}\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right] dM_i(t)$$
$$+ n^{-1/2}\sum_{i=1}^n \int_0^\tau \Delta_i(1-\xi_i)\{\frac{\eta_i}{\hat{q}(t)} - 1\}\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right] dM_i(t)$$
$$+ n^{-1/2}\sum_{i=1}^n \int_0^\tau \Delta_i(1-\xi_i)\eta_i\{\frac{1}{\hat{q}(t)} - \frac{1}{\tilde{q}}\}\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right] dM_i(t). \tag{6.8}$$

$n^{-1/2}\sum_{i=1}^{n}(1-\Delta_i)M_i(t)$ is the sum of independent and identically distributed zero-mean random variables, for fixed t. Using Lemma A1 and noting that for finite samples, $\mu_B(t) = \frac{1}{n}\sum_{i=1}^{n} B_i(t)$, we have $n^{-1/2}\sum_{i=1}^{n}(1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}}-1\}M_i(t)$ converges to a zero-mean Gaussian process. Using Kolmogorv-Centsov Theorem, we can show that the limiting process has continuous sample paths. Further, we have by Slutsky's theorem, $\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} \xrightarrow{P} 0$, uniformly in t. Hence, the first term of (6.8) converges to 0 in probability by using Lemma A2. Similarly, the third term of (6.8) can be shown to converge to 0 in probability. Considering the second term of the above equation, we have

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\int_0^\tau (1-\Delta_i)\xi_i\{\frac{1}{\hat{\alpha}(t)}-\frac{1}{\tilde{\alpha}}\}\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t}-\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right]dM_i(t) \\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^\tau (1-\Delta_i)\xi_i\sum_j\left\{\frac{1}{\tilde{\alpha}E\left[(1-\Delta_1)Y_1(t)\right]}\frac{1}{n}(1-\frac{\xi_j}{\tilde{\alpha}})(1-\Delta_j)Y_j(t)\right\} \\
&\times\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t}-\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right]dM_i(t)+o_p(1) \\
&= n^{-1/2}\sum_{j=1}^{n}\int_0^\tau \frac{1}{E\left[(1-\Delta_1)Y_1(t)\right]}(1-\frac{\xi_j}{\tilde{\alpha}})(1-\Delta_j)Y_j(t)\left\{\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\xi_i}{\tilde{\alpha}}dM_i(t)\right\} \\
&\times\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t}-\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right]+o_p(1)
\end{aligned}
$$

From Lemma A1, we have $\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\xi_i}{\tilde{\alpha}}dM_i(t)$ converges to zero in probability uniformly in t and we have shown earlier that $S^{(d)}(\beta,t)$ and $\hat{S}^{(d)}(\beta,t)$ have the same limit in probability uniformly in t. $S^{(0)}(\beta,t)$ and $\hat{S}^{(0)}(\beta,t)$ are bounded away from zero and hence, $\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}$ converges to zero in probability by the Slutsky's theorem. Hence, the second term of equation (6.8) converges to zero in probability. Similarly, we have the fourth term of (6.8) as

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \Delta_i(1-\xi_i)\eta_i\{\frac{1}{\hat{\alpha}(t)}-\frac{1}{\tilde{\alpha}}\}\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t}-\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right]dM_i(t) \\
&= n^{-1/2}\sum_{j=1}^{n}\int_0^\tau \frac{1}{E\left[\Delta_1(1-\xi_1)Y_1(t)\right]}(1-\frac{\eta_j}{\tilde{q}})\Delta_j(1-\xi_j)Y_j(t)\left\{\frac{1}{n}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\frac{\eta_i}{\tilde{q}}dM_i(t)\right\} \\
&\left[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t}-\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right]+o_p(1)
\end{aligned}
$$

Proceeding in the same way, it can be easily shown that the fourth term of (6.8) converges to 0 in probability, implying that the fourth term of equation (6.7) converges to zero. Next, looking at the second term of the equation :

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{w_i^{II}(t)-1\}\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)\\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{\xi_i\Delta_i+(1-\Delta_i)\frac{\xi_i}{\hat{\alpha}(t)}+\Delta_i(1-\xi_i)\frac{\eta_i}{\hat{q}(t)}-1\}\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)\\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}(1-\Delta_i)\{\frac{\xi_i}{\hat{\alpha}(t)}-1\}\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\Delta_i(1-\xi_i)\{\frac{\eta_i}{\hat{q}(t)}-1\}\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)\\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}(1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}}-1\}\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}(1-\Delta_i)\xi_i\{\frac{1}{\hat{\alpha}(t)}-\frac{1}{\tilde{\alpha}}\}\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}}-1\}\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\Delta_i(1-\xi_i)\eta_i\{\frac{1}{\hat{q}(t)}-\frac{1}{\tilde{q}}\}\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)\\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}(1-\Delta_i)\{1-\frac{\xi_i}{\tilde{\alpha}}\}\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\\
&- n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}(1-\Delta_i)\xi_i\sum_j\left\{\frac{1}{\tilde{\alpha}E\left[(1-\Delta_1)Y_1(t)\right]}\frac{1}{n}(1-\frac{\xi_j}{\tilde{\alpha}})(1-\Delta_j)Y_j(t)\right\}\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]\\
&\times Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)+n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}}-1\}\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)\\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\Delta_i(1-\xi_i)\eta_i\sum_j\left\{\frac{1}{\tilde{q}E\left[\Delta_1(1-\xi_1)Y_1(t)\right]}\frac{1}{n}(1-\frac{\eta_j}{\tilde{q}})\Delta_j(1-\xi_j)Y_j(t)\right\}\\
&\times\left[Z_i(t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)+o_P(1)\\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}(1-\Delta_i)\{1-\frac{\xi_i}{\tilde{\alpha}}\}\left[Z_i(t)-e(\beta,t)\right]Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\\
&- n^{-1/2}\sum_{j=1}^{n}\int_0^{\tau}\frac{1}{\tilde{\alpha}E\left[(1-\Delta_1)Y_1(t)\right]}(1-\frac{\xi_j}{\tilde{\alpha}})(1-\Delta_j)Y_j(t)\frac{1}{n}\sum_{i=1}^{n}\left\{(1-\Delta_i)\frac{\xi_i}{\tilde{\alpha}}\left[Z_i(t)-e(\beta,t)\right]\right.
\end{aligned}
$$

$$\times Y_i(t)e^{\beta Z_i(t)}\Big\}\, d\mu_0(t) + n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}} - 1\}\,[Z_i(t) - e(\beta,t)]\, dM_i(t)$$

$$+ n^{-1/2}\sum_{j=1}^{n}\int_0^\tau \frac{1}{E\left[\Delta_1(1-\xi_1)Y_1(t)\right]}(1-\frac{\eta_j}{\tilde{q}})\Delta_j(1-\xi_j)Y_j(t)\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i(1-\xi_i)\frac{\eta_i}{\tilde{q}}\,(Z_i(t)\right.\right.$$

$$\left.\left. -e(\beta,t))\right]dM_i(t)\right\} + o_P(1). \tag{6.9}$$

We get the last equality because $\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} - e(\beta,t) \xrightarrow{P} 0$ uniformly in t. Further, by Lemma A1, $\frac{1}{n}\sum_{i=1}^{n}\left\{(1-\Delta_i)\frac{\xi_i}{\tilde{\alpha}}\left[Z_i(t) - e(\beta,t)\right]Y_i(t)e^{\beta Z_i(t)}\right\}$ and $\frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i(1-\xi_i)\frac{\eta_i}{\tilde{q}}\left(Z_i(t) - e(\beta,t)\right)\right]$ $\times dM_i(t)$ are asymptotically equivalent to $E\left\{(1-\Delta_1)\left[Z_1(t) - e(\beta,t)\right]Y_1(t)e^{\beta' Z_1(t)}\right\}$ and $E\left[\Delta_i(1-\xi_i)\left(Z_i(t) - e(\beta,t)\right)dM_i(t)\right]$, respectively. Therefore, the above can be simplified into

$$n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \{w_i^{II}(t) - 1\}\left[Z_i(t) - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}\right]dM_i(t)$$

$$= n^{-1/2}\sum_{i=1}^{n}\int_0^\tau (1-\Delta_i)\{1 - \frac{\xi_i}{\tilde{\alpha}}\}\left[\left(Z_i(t) - e(\beta,t)\right)Y_i(t)e^{\beta Z_i(t)}\right.$$

$$\left. - \frac{Y_i(t)E\left[(1-\Delta_1)\left(Z_1(t) - e(\beta,t)\right)Y_1(t)e^{\beta' Z_1(t)}\right]}{E\left[(1-\Delta_1)Y_1(t)\right]}\right]d\mu_0(t)$$

$$+ n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}} - 1\}\left[\left(Z_i(t) - e(\beta,t)\right)dM_i(t) - \right.$$

$$\left.\frac{Y_i(t)E\left[\left(Z_1(t) - e(\beta,t)\right)dM_1(t) \mid \Delta_1 = 1, \xi_1 = 0\right]}{E\left[Y_1(t) \mid \Delta_1 = 1\right]}\right] + o_P(1)$$

$$= n^{-1/2}\sum_{i=1}^{n}\int_0^\tau (1-\Delta_i)\{1 - \frac{\xi_i}{\tilde{\alpha}}\}\left[R_i(\beta,t) - \frac{Y_i(t)E\left[(1-\Delta_1)R_1(\beta,t)\right]}{E\left[(1-\Delta_1)Y_1(t)\right]}\right]d\mu_0(t)$$

$$+ n^{-1/2}\sum_{i=1}^{n}\int_0^\tau \Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}} - 1\}\left[dM_{\tilde{Z},i}(\beta,t) - \frac{Y_i(t)E\left[dM_{\tilde{Z},1}(\beta,t) \mid \Delta_1 = 1, \xi_1 = 0\right]}{E\left[Y_1(t) \mid \Delta_1 = 1\right]}\right]$$

$$+ o_P(1).$$

where $R_i(\beta,t) = Y_i(t)\tilde{Z}_i(t)e^{\beta Z_i(t)} = Y_i(t)\left(Z_i(t) - e(\beta,t)\right)e^{\beta Z_i(t)}$, $M_{\tilde{Z},i}(\beta) = \int_0^\tau \left(Z_i(t) - e(\beta,t)\right)$ $\times dM_i(t)$ and $dM_{\tilde{Z},i}(\beta,t) = \left(Z_i(t) - e(\beta,t)\right)dM_i(t)$. Therefore, we can state that $n^{1/2}U_n(\beta)$

is asymptotically equivalent to

$$n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau} dM_{\tilde{Z},i}(\beta)+n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{1-\frac{\xi_i}{\tilde{\alpha}}\}(1-\Delta_i)\left(R_i(\beta,t)-\frac{Y_i(t)E\left[(1-\Delta_i)R_i(\beta,t)\right]}{E\left[(1-\Delta_i)Y_i(t)\right]}\right)$$
$$d\mu_0(t)+n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}}-1\}\left[dM_{\tilde{Z},i}(\beta,t)-\frac{Y_i(t)E\left[dM_{\tilde{Z},1}(\beta,t)\mid\Delta_1=1,\xi_1=0\right]}{E\left[Y_1(t)\mid\Delta_1=1\right]}\right]. \tag{6.10}$$

Under the regularity conditions, from Lin et al. (2000), the first term has covariance, $E(M_{\tilde{Z},i}(\beta_0))^{\otimes 2}$. The second and third terms can be shown to be asymptotically zero-mean normal with covariance matrix, $V^{I}(\beta_0)$ and $V^{II}(\beta_0)$ respectively based on Hájek's (1960) central limit theorem for finite population sampling. First we need to check the following conditions in order to use Hájek's Theorem.

(a) $\tilde{\alpha}$ converges to $\alpha\in(0,1)$.

(b) $\frac{1}{n}\left|a'\left(\int_0^{\tau}R_i^*(\beta,t)d\mu_0(t)-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}R_i^*(\beta,t)d\mu_0(t)\right)\right|$ converges to 0 in probability.

(c) $\frac{1}{n-1}\sum_{i=1}^{n}\left(a'\int_0^{\tau}R_i^*(\beta,t)d\mu_0(t)-\frac{1}{n}\sum_{i=1}^{n}a'\int_0^{\tau}R_i^*(\beta,t)d\mu_0(t)\right)^2$ converges to some quantity, $\sigma^2>0$.

For this, let us take $\mathbf{a}=(a_1,a_2,\dots,a_p)'$ be a p $\times$ 1 real-valued vector and let us define $R_i^*(\beta,t)=R_i(\beta,t)-\frac{Y_i(t)E[(1-\Delta_i)R_i(\beta,t)]}{E[(1-\Delta_i)Y_i(t)]}$ Then we have

$$\begin{aligned}&\left|a'\left(\int_0^{\tau}R_i^*(\beta,t)d\mu_0(t)-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}R_i^*(\beta,t)d\mu_0(t)\right)\right|\\&\le\left|a'\left(\int_0^{\tau}R_i(\beta,t)d\mu_0(t)-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}R_i(\beta,t)d\mu_0(t)\right)\right|\\&+\left|a'\left(\int_0^{\tau}e^*(\beta,t)Y_i(t)d\mu_0(t)-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\int_0^{\tau}e^*(\beta,t)Y_i(t)d\mu_0(t)\right)\right|.\end{aligned} \tag{6.11}$$

where $e^*(\beta,t)=\frac{E[(1-\Delta_i)R_i(\beta,t)]}{E[(1-\Delta_i)Y_i(t)]}$. $\left|a'\left(\int_0^{\tau}e^*(\beta,t)Y_i(t)d\mu_0(t)-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\int_0^{\tau}e^*(\beta,t)Y_i(t)d\mu_0(t)\right)\right|=\left|a'\int_0^{\tau}e^*(\beta,t)\left(Y_i(t)-\bar{Y}(t)\right)d\mu_0(t)\right|$, $\bar{Y}(t)=\frac{\sum_{i=1}^{n}Y_i(t)}{n}$ and $-1\le\left(Y_i(t)-\bar{Y}(t)\right)\le 1$. $\Rightarrow$ $\left|a'\int_0^{\tau}e^*(\beta,t)\left(Y_i(t)-\bar{Y}(t)\right)d\mu_0(t)\right|\le\left|a'\int_0^{\tau}e^*(\beta,t)d\mu_0(t)\right|$. For the first term on the right hand side of (6.11),

$$
\begin{aligned}
&\left|a'\left(\int_0^\tau R_i(\beta,t)d\mu_0(t)-\frac{1}{n}\sum_{i=1}^n\int_0^\tau R_i(\beta,t)d\mu_0(t)\right)\right| \\
&=\left|\left(a'\int_0^\tau \tilde{Z}_i(t)Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\right)-\frac{1}{n}\sum_{i=1}^n\left(a'\int_0^\tau \tilde{Z}_i(t)Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\right)\right| \\
&\le\left|\left(a'\int_0^\tau \tilde{Z}_i(t)Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\right)\right|+\left|\frac{1}{n}\sum_{i=1}^n\left(a'\int_0^\tau \tilde{Z}_i(t)Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\right)\right| \\
&\le\left|\left(a'\int_0^\tau Z_i(t)Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\right)\right|+\left|\left(a'\int_0^\tau e(\beta,t)Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\right)\right| \\
&+\left|\frac{1}{n}\sum_{i=1}^n\left(a'\int_0^\tau Z_i(t)Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\right)\right|+\left|\frac{1}{n}\sum_{i=1}^n\left(a'\int_0^\tau e(\beta,t)Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\right)\right|.
\end{aligned}
$$

Next, we see that

$$
\begin{aligned}
&\max_i\left(\left|a'\left(\int_0^\tau R_i^*(\beta,t)d\mu_0(t)-\frac{1}{n}\sum_{i=1}^n\int_0^\tau R_i^*(\beta,t)d\mu_0(t)\right)\right|\right)^2 \\
&\le\max_i\left(\left|a'\left(\int_0^\tau R_i(\beta,t)d\mu_0(t)-\frac{1}{n}\sum_{i=1}^n\int_0^\tau R_i(\beta,t)d\mu_0(t)\right)\right|\right. \\
&\left.+\left|a'\int_0^\tau e^*(\beta,t)\left(Y_i(t)-\bar{Y}(t)\right)d\mu_0(t)\right|\right)^2 \\
&\le\max_i\left(\left|a'\left(\int_0^\tau R_i(\beta,t)d\mu_0(t)-\frac{1}{n}\sum_{i=1}^n\int_0^\tau R_i(\beta,t)d\mu_0(t)\right)\right|\right)^2 \\
&+2\max_i\left|a'\left(\int_0^\tau R_i(\beta,t)d\mu_0(t)-\frac{1}{n}\sum_{i=1}^n\int_0^\tau R_i(\beta,t)d\mu_0(t)\right)\right|\left|a'\int_0^\tau e^*(\beta,t)\left(Y_i(t)-\bar{Y}(t)\right)d\mu_0(t)\right| \\
&+\max_i\left(\left|a'\int_0^\tau e^*(\beta,t)\left(Y_i(t)-\bar{Y}(t)\right)d\mu_0(t)\right|\right)^2. \qquad (6.12)
\end{aligned}
$$

The first term on the RHS of (6.12) can be rewritten as

$$
\begin{aligned}
&\max_i\left(\left|a'\left(\int_0^\tau R_i^*(\beta,t)d\mu_0(t)-\frac{1}{n}\sum_{i=1}^n\int_0^\tau R_i^*(\beta,t)d\mu_0(t)\right)\right|\right)^2 \\
&\le\max_i\left(\left|\left(a'\int_0^\tau Z_i(t)Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\right)\right|+\left|\left(a'\int_0^\tau e(\beta,t)Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\right)\right|\right. \\
&\left.+\left|\frac{1}{n}\sum_{i=1}^n\left(a'\int_0^\tau Z_i(t)Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\right)\right|+\left|\frac{1}{n}\sum_{i=1}^n\left(a'\int_0^\tau e(\beta,t)Y_i(t)e^{\beta Z_i(t)}d\mu_0(t)\right)\right|\right)^2
\end{aligned}
$$

$$\leq \left(\max_i \left| \left(a' \int_0^\tau Z_i(t) Y_i(t) e^{\beta Z_i(t)} d\mu_0(t) \right) \right| + \max_i \left| \left(a' \int_0^\tau e(\beta, t) Y_i(t) e^{\beta Z_i(t)} d\mu_0(t) \right) \right| \right.$$
$$\left. + \max_i \left| \frac{1}{n} \sum_{i=1}^n \left(a' \int_0^\tau Z_i(t) Y_i(t) e^{\beta Z_i(t)} d\mu_0(t) \right) \right| + \max_i \left| \frac{1}{n} \sum_{i=1}^n \left(a' \int_0^\tau e(\beta, t) Y_i(t) e^{\beta Z_i(t)} d\mu_0(t) \right) \right| \right)^2 .$$

We can note that,

$$\max_i \left(\left| \int_0^\tau a' Z_i(t) Y_i(t) e^{\beta Z_i(t)} d\mu_0(t) \right| \right) \leq \sup_{i,t} \left(\left| a' Z_i(t) \right| Y_i(t) e^{\beta Z_i(t)} \mu_0(t) \right)$$
$$\max_i \left(\left| \int_0^\tau a' e(\beta, t) Y_i(t) e^{\beta Z_i(t)} d\mu_0(t) \right| \right) \leq \sup_{i,t} \left(\left| a' e(\beta, t) \right| Y_i(t) e^{\beta Z_i(t)} \mu_0(t) \right)$$
$$\max_i \left(\left| \frac{1}{n} \sum_{i=1}^n \int_0^\tau a' Z_i(t) Y_i(t) e^{\beta Z_i(t)} d\mu_0(t) \right| \right) \leq \sup_{i,t} \left(\left| a' Z_i(t) \right| Y_i(t) e^{\beta Z_i(t)} \mu_0(t) \right)$$
$$\max_i \left(\left| \frac{1}{n} \sum_{i=1}^n \int_0^\tau a' e(\beta, t) Y_i(t) e^{\beta Z_i(t)} d\mu_0(t) \right| \right) \leq \sup_{i,t} \left(\left| a' e(\beta, t) \right| Y_i(t) e^{\beta Z_i(t)} \mu_0(t) \right)$$
$$\max_i \left(\left| a' \int_0^\tau e^*(\beta, t) \left(Y_i(t) - \bar{Y}(t) \right) d\mu_0(t) \right| \right) \leq \sup_t \left(|a' e^*(\beta, t)| \mu_0(t) \right)$$

Note that, we can easily show that $\frac{1}{n} \sup_{i,t} \left(|a' Z_i(t)| \, Y_i(t) e^{\beta Z_i(t)} \mu_0(t) \right) \xrightarrow{P} 0$ and $\frac{1}{n} \sup_{i,t} (|a' e(\beta, t)| \; Y_i(t) e^{\beta Z_i(t)} \mu_0(t)) \xrightarrow{P} 0$ as $n \to \infty$ as the functions are bounded from the regularity conditions, and hence, we see that (b) holds. As for (c), note that, $\frac{1}{n} \sum_{i=1}^n \left(a' \int_0^\tau R_i^*(\beta, t) d\mu_0(t) - \frac{1}{n} \sum_{i=1}^n a' \int_0^\tau R_i^*(\beta, t) d\mu_0(t) \right)^2 \leq \max_i \left(a' \int_0^\tau R_i^*(\beta, t) d\mu_0(t) - \frac{1}{n} \sum_{i=1}^n a' \int_0^\tau R_i^*(\beta, t) d\mu_0(t) \right)^2$ which is bounded and $\frac{n}{n-1} \to 1$ as $n \to \infty$. Hence, we have $n^{-1/2} \sum_{i=1}^n \int_0^\tau a' \left(\{1 - \frac{\xi_i}{\tilde{\alpha}}\}(1 - \Delta_i) R_i^*(\beta, t) \right) d\mu_0(t)$ converges to a normal distribution. Therefore, by Cramer- Wold device, we have $n^{-1/2} \sum_{i=1}^n \int_0^\tau \left(\{1 - \frac{\xi_i}{\tilde{\alpha}}\}(1 - \Delta_i) R_i^*(\beta, t) \right) d\mu_0(t)$ converges to a zero-mean Normal random variable with variance,

$$\frac{1-\alpha}{\alpha} V^I(\beta) = E \left(\int_0^\tau (1 - \Delta_i) R_i^*(\beta, t) d\mu_0(t) \right)^{\otimes 2}$$
$$= E \left(\int_0^\tau (1 - \Delta_i) \left\{ R_i(\beta, t) - \frac{Y_i(t) E\left[(1 - \Delta_i) R_i(\beta, t)\right]}{E\left[(1 - \Delta_i) Y_i(t)\right]} \right\} d\mu_0(t) \right)^{\otimes 2} .$$

Similarly, we can show that $n^{-1/2} \sum_{i=1}^n \{1 - \frac{\eta_i}{\tilde{q}}\} \Delta_i (1 - \xi_i) \int_0^\tau \Big[dM_{\tilde{Z},i}(\beta, t) - Y_i(t)$

$\times \frac{E\left(dM_{\tilde{Z},1}(\beta,t)|\Delta_1=1,\xi_1=0\right)}{E(Y_1(t)|\Delta_1=1)}\Big]$ converges to a zero-mean Normal random variable with variance,

$$
\begin{aligned}
&(1-\alpha)\frac{1-q}{q}P(\Delta_1=1)V^{II}(\beta) \\
&= var\left(\int_0^\tau \left[dM_{\tilde{Z},i}(\beta,t) - \frac{Y_i(t)E\left(dM_{\tilde{Z},1}(\beta,t) \mid \Delta_1=1,\xi_1=0\right)}{E\left(Y_1(t) \mid \Delta_1=1\right)}\right]\right).
\end{aligned}
$$

Next let us consider the covariance between $n^{-1/2}\sum_{i=1}^n M_{\tilde{Z},i}(\beta,t)$ and $n^{-1/2}\sum_{i=1}^n \int_0^\tau (1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)R_i^*(\beta,t)d\mu_0(t)$. Defining the marginal filtration as , $\mathcal{F}_i(t)=\sigma\{N_i(t),Y_i(t),Z_i(t)\}, t\in[0,\tau]$ for the i-th individual at time t, we have

$$
\begin{aligned}
&cov(n^{-1/2}\sum_{i=1}^n M_{\tilde{Z},i}(\beta,t), n^{-1/2}\sum_{i=1}^n (1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)\int_0^\tau R_i^*(\beta,t)d\mu_0(t)) \\
&= E\left(n^{-1/2}\sum_{i=1}^n M_{\tilde{Z},i}(\beta,t) n^{-1/2}\sum_{i=1}^n (1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)\int_0^\tau R_i^*(\beta,t)d\mu_0(t)\right) \\
&= E\left[E\left(n^{-1/2}\sum_{i=1}^n M_{\tilde{Z},i}(\beta,t) n^{-1/2}\sum_{i=1}^n (1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)\int_0^\tau R_i^*(\beta,t)d\mu_0(t) \,\Big|\, \mathcal{F}(\tau)\right)\right] \\
&= E\left[n^{-1/2}\sum_{i=1}^n M_{\tilde{Z},i}(\beta,t) n^{-1/2}\sum_{i=1}^n E\left(1-\frac{\xi_i}{\tilde{\alpha}} \,\Big|\, \mathcal{F}(\tau)\right)(1-\Delta_i)\int_0^\tau R_i^*(\beta,t)d\mu_0(t)\right] = 0.
\end{aligned}
$$

Similarly, the covariance between $n^{-1/2}\sum_{i=1}^n M_{\tilde{Z},i}(\beta,t)$ and $n^{-1/2}\sum_{i=1}^n (\frac{\eta_i}{\tilde{q}}-1)\Delta_i(1-\xi_i)$ $M^*_{\tilde{Z},i}(\beta,t)$ where $M^*_{\tilde{Z},i}(\beta,t)=\int_0^\tau \left[dM_{\tilde{Z},i}(\beta,t) - \frac{Y_i(t)E(dM_{\tilde{Z},1}(\beta,t)|\Delta_1=1,\xi_1=0)}{E(Y_1(t)|\Delta_1=1,\xi_1=0)}\right]$ is given by

$$
\begin{aligned}
&cov\left(n^{-1/2}\sum_{i=1}^n M_{\tilde{Z},i}(\beta,t), n^{-1/2}\sum_{i=1}^n (\frac{\eta_i}{\tilde{q}}-1)\Delta_i(1-\xi_i)M^*_{\tilde{Z},i}(\beta,t)\right) \\
&= E\left(n^{-1}\sum_{i=1}^n M_{\tilde{Z},i}(\beta,t)\sum_{i=1}^n (\frac{\eta_i}{\tilde{q}}-1)\Delta_i(1-\xi_i)M^*_{\tilde{Z},i}(\beta,t)\right) \\
&= E\left[E\left(n^{-1}\sum_{i=1}^n M_{\tilde{Z},i}(\beta,t)\sum_{i=1}^n (\frac{\eta_i}{\tilde{q}}-1)\Delta_i(1-\xi_i)M^*_{\tilde{Z},i}(\beta,t) \,\Big|\, \mathcal{F}(\tau)\right)\right] \\
&= E\left[n^{-1}\sum_{i=1}^n M_{\tilde{Z},i}(\beta,t)\sum_{i=1}^n E\left(1-\xi_i)(1-\frac{\eta_i}{\tilde{q}}) \,\Big|\, \mathcal{F}(\tau)\right)\Delta_i M^*_{\tilde{Z},i}(\beta,t)\right] = 0.
\end{aligned}
$$

Finally, we can show that $cov\Big(n^{-1/2}\sum_{i=1}^n \int_0^\tau R_i^*(\beta,t)d\mu_0(t), n^{-1/2}\sum_{i=1}^n (\frac{\eta_i}{\tilde{q}}-1)\Delta_i(1-\xi_i)$ $M^*_{\tilde{Z},i}(\beta,t)\Big)=0$. Therefore, the variance of $n^{1/2}U_n(\beta)$ is given by $E(M_{\tilde{Z},i}(\beta))^{\otimes 2}+\frac{1-\alpha}{\alpha}V^I(\beta)+$

$(1-\alpha)\frac{1-q}{q}P(\Delta_1 = 1)V^{II}(\beta)$. This implies that $U_n(\beta) \xrightarrow{P} 0$ as $n \to \infty$. Further, we can use Taylor Series expansion of $n^{1/2}U_n(\hat{\beta}^{II})$ around β_0 to get

$$n^{1/2}U_n(\hat{\beta}^{II}) \approx n^{1/2}U_n(\beta) + n^{1/2}(\hat{\beta}^{II} - \beta_0)'\frac{\delta}{\delta\beta}U_n^{II}(\beta^*)$$

where β^* belongs to the line segment between β_0 and $\hat{\beta}^{II}$. Hence, we can say that $n^{1/2}(\hat{\beta}^{II} - \beta_0)$ converges to a normal distribution with mean 0 and variance

$$A(\beta_0)^{-1}\left(E(M_{\tilde{Z},i}(\beta))^{\otimes 2} + \frac{1-\alpha}{\alpha}V^I(\beta) + (1-\alpha)\frac{1-q}{q}P(\Delta_1 = 1)V^{II}(\beta)\right)A(\beta_0)^{-1}.$$

A.3 Proof of Theorem 2

Next, we look at the distribution of the estimate of the rate function, $\hat{\mu}_0(\hat{\beta}^{II}, t)$.

$$
\begin{aligned}
&n^{1/2}\{\hat{\mu}_0(\hat{\beta}^{II}, t) - \mu_0(t)\} \\
&= n^{1/2}\{\hat{\mu}_0(\hat{\beta}^{II}, t) - \int_0^t \frac{\sum_{i=1}^n w_i^{II}(t)dN_i(u)}{n\hat{S}^{(0)}(\beta_0, u)}\} + n^{1/2}\{\int_0^t \frac{\sum_{i=1}^n w_i^{II}(t)dN_i(u)}{n\hat{S}^{(0)}(\beta_0, u)} - \mu_0(t)\} \\
&= n^{1/2}\int_0^t \left\{\frac{1}{n\hat{S}^{(0)}(\hat{\beta}^{II}, u)} - \frac{1}{n\hat{S}^{(0)}(\beta_0, u)}\right\}\sum_{i=1}^n w_i^{II}(u)dM_i(u) + n^{1/2}\int_0^t \left\{\frac{1}{\hat{S}^{(0)}(\hat{\beta}^{II}, u)}\right. \\
&\left. - \frac{1}{\hat{S}^{(0)}(\beta_0, u)}\right\}\hat{S}^{(0)}(\beta_0, u)d\mu_0(u) + \int_0^t \frac{1}{\hat{S}^{(0)}(\beta_0, u)}\left\{n^{-1/2}\sum_{i=1}^n w_i^{II}(u)dM_i(u)\right\} \\
&= n^{1/2}\int_0^t \left\{\frac{1}{n\hat{S}^{(0)}(\hat{\beta}^{II}, u)} - \frac{1}{n\hat{S}^{(0)}(\beta_0, u)}\right\}\sum_{i=1}^n dM_i(u) \\
&+ n^{1/2}\int_0^t \left\{\frac{1}{n\hat{S}^{(0)}(\hat{\beta}^{II}, u)} - \frac{1}{n\hat{S}^{(0)}(\beta_0, u)}\right\}\sum_{i=1}^n (w_i^{II}(u) - 1)dM_i(u) \\
&+ n^{1/2}\int_0^t \left\{\frac{1}{\hat{S}^{(0)}(\hat{\beta}^{II}, u)} - \frac{1}{\hat{S}^{(0)}(\beta_0, u)}\right\}\hat{S}^{(0)}(\beta_0, u)d\mu_0(u) + \int_0^t \frac{1}{\hat{S}^{(0)}(\beta_0, u)}\left\{n^{-1/2}\sum_{i=1}^n dM_i(u)\right\} \\
&+ \int_0^t \frac{1}{\hat{S}^{(0)}(\beta_0, u)}\left\{n^{-1/2}\sum_{i=1}^n (w_i^{II}(u) - 1)dM_i(u)\right\}. \qquad (6.13)
\end{aligned}
$$

By the Taylor series expansion of $\hat{S}^{(0)}(\hat{\beta}^{II}, u)^{-1}$ around β_0 we have,

$$
\hat{S}^{(0)}(\hat{\beta}^{II}, u)^{-1} \approx \hat{S}^{(0)}(\beta_0, u)^{-1} - (\hat{\beta}^{II} - \beta_0)'\frac{\hat{S}^{(1)}(\beta^*, u)}{\hat{S}^{(0)}(\beta^*, u)}
$$

where β^* belongs to the line segment β_0 and $\hat{\beta}^{II}$. Therefore the first term of the equation (6.13) can be written as

$$
\begin{aligned}
&n^{1/2}\int_0^t \left\{\frac{1}{n\hat{S}^{(0)}(\hat{\beta}^{II}, u)} - \frac{1}{n\hat{S}^{(0)}(\beta_0, u)}\right\}\sum_{i=1}^n dM_i(u) \\
&= n^{-1/2}\int_0^t \left\{-(\hat{\beta}^{II} - \beta_0)'\frac{\hat{S}^{(1)}(\beta^*, u)}{\hat{S}^{(0)}(\beta^*, u)}\right\}\sum_{i=1}^n dM_i(u). \qquad (6.14)
\end{aligned}
$$

Note that, $\hat{S}^{(1)}(\beta^*, u)$ and $\hat{S}^{(0)}(\beta^*, u)$ have bounded variations and $\hat{S}^{(0)}(\beta^*, u)$ is bounded away from zero; hence, $\frac{\hat{S}^{(1)}(\beta^*,u)}{\hat{S}^{(0)}(\beta^*,u)}$ also has bounded variation and can be expressed as the difference of two monotone functions in t. Therefore, noting that $\hat{\beta}^{II} \xrightarrow{P} \beta_0$ and $n^{-1/2}\sum_{i=1}^{n} dM_i(u)$ converges to a zero-mean Gaussian process , having continuous sample paths, we can show that converges to 0 as n $\to \infty$ by using Lemma A1 from Lin et al. (2000). Next, the second term can be written as

$$n^{1/2}\int_0^t \left\{ \frac{1}{n\hat{S}^{(0)}(\hat{\beta}^{II},u)} - \frac{1}{n\hat{S}^{(0)}(\beta_0,u)} \right\} \sum_{i=1}^{n} (w_i^{II}(u)-1)dM_i(u)$$
$$= -n^{-1/2}\int_0^t \left\{ (\hat{\beta}^{II}-\beta_0)' \frac{\hat{S}^{(1)}(\beta^*,u)}{\hat{S}^{(0)}(\beta^*,u)} \right\} \sum_{i=1}^{n} \left(\Delta_i\xi_i + \frac{(1-\Delta_i)\xi_i}{\tilde{\alpha}(u)} + \frac{\Delta_i(1-\xi_i)\eta_i}{\hat{q}(u)} - 1 \right) dM_i(u)$$
$$+ o_P(1)$$
$$= -n^{-1/2}\int_0^t \left\{ (\hat{\beta}^{II}-\beta_0)' \frac{\hat{S}^{(1)}(\beta^*,u)}{\hat{S}^{(0)}(\beta^*,u)} \right\} \sum_{i=1}^{n} \left((1-\Delta_i)(\frac{\xi_i}{\tilde{\alpha}(u)}-1) + \Delta_i(1-\xi_i)(\frac{\eta_i}{\hat{q}(u)}-1) \right)$$
$$\times dM_i(u) + o_P(1)$$
$$= -n^{-1/2}\int_0^t \left\{ (\hat{\beta}^{II}-\beta_0)' \frac{\hat{S}^{(1)}(\beta^*,u)}{\hat{S}^{(0)}(\beta^*,u)} \right\} \sum_{i=1}^{n} (1-\Delta_i)\left(\frac{\xi_i}{\tilde{\alpha}}-1\right) dM_i(u)$$
$$- n^{-1/2}\int_0^t \left\{ (\hat{\beta}^{II}-\beta_0)' \frac{\hat{S}^{(1)}(\beta^*,u)}{\hat{S}^{(0)}(\beta^*,u)} \right\} \sum_{i=1}^{n} (1-\Delta_i)\xi_i \left(\frac{1}{\tilde{\alpha}E((1-\Delta_1)Y_1(t))} \frac{1}{n}\sum_j (1-\Delta_j) \right.$$
$$\left. (1-\frac{\xi_j}{\tilde{\alpha}})Y_j(t) \right) dM_i(u) - n^{-1/2}\int_0^t \left\{ (\hat{\beta}^{II}-\beta_0)' \frac{\hat{S}^{(1)}(\beta^*,u)}{\hat{S}^{(0)}(\beta^*,u)} \right\} \sum_{i=1}^{n} \Delta_i(1-\xi_i)\left(\frac{\eta_i}{\tilde{q}}-1\right) dM_i(u)$$
$$- n^{-1/2}\int_0^t \left\{ (\hat{\beta}^{II}-\beta_0)' \frac{\hat{S}^{(1)}(\beta^*,u)}{\hat{S}^{(0)}(\beta^*,u)} \right\} \sum_{i=1}^{n} \Delta_i(1-\xi_i)\eta_i \left(\frac{1}{\tilde{q}E(\Delta_1(1-\xi_1)Y_1(t))} \frac{1}{n}\sum_j \Delta_j(1-\xi_j) \right.$$
$$\left. \times(1-\frac{\eta_j}{\tilde{q}})Y_j(t) \right) dM_i(u) + o_P(1) \tag{6.15}$$

Note that $\hat{\beta}^{II} - \beta_0 \xrightarrow{P} 0$ as $n \to \infty$, $\frac{\hat{S}^{(1)}(\beta^*,t)}{\hat{S}^{(0)}(\beta^*,t)}$ is bounded away from zero. Further, we have already shown that $n^{1/2}(\hat{\beta}^{II}-\beta_0)'\frac{\hat{S}^{(1)}(\beta^*,t)}{\hat{S}^{(0)}(\beta^*,t)}$ converges to a zero-mean Gaussian process. Similar, we have also shown that $n^{-1/2}\sum_{i=1}^{n}(1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}}-1\}M_i(t)$ converges to a zero-mean Gaussian process (Lin et al. (2000), Van Der Vaart and Wellner (1996)(example 2.11.16 pg 215)). Using Kolmogorv-Centsov Theorem, we can show that the limiting process has continuous sample paths. Hence, the first term of (6.15) converges to 0 in probability by using Lemma A2. Similarly, the third term of (6.15) can be shown to converge to 0 in probability. Considering the second term of the above equation, we have

$$n^{-1/2}\sum_{i=1}^{n}\int_0^t (1-\Delta_i)\xi_i \sum_j \left\{\frac{1}{\tilde{\alpha}E\left[(1-\Delta_1)Y_1(u)\right]}\frac{1}{n}(1-\frac{\xi_j}{\tilde{\alpha}})(1-\Delta_j)Y_j(u)\right\}\left\{(\hat{\beta}^{II}-\beta_0)'\right.$$
$$\left.\times\frac{\hat{S}^{(1)}(\beta^*,u)}{\hat{S}^{(0)}(\beta^*,u)}\right\}dM_i(t)$$
$$=(\hat{\beta}^{II}-\beta_0)'n^{-1/2}\sum_{j=1}^{n}\int_0^t \frac{1}{E\left[(1-\Delta_1)Y_1(u)\right]}(1-\frac{\xi_j}{\tilde{\alpha}})(1-\Delta_j)Y_j(u)\left\{\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\xi_i}{\tilde{\alpha}}dM_i(u)\right\}$$
$$\times\left\{\frac{\hat{S}^{(1)}(\beta^*,u)}{\hat{S}^{(0)}(\beta^*,u)}\right\}$$

Note that $n^{-1/2}\sum_{j=1}^{n}\int_0^t(1-\frac{\xi_j}{\tilde{\alpha}})(1-\Delta_j)Y_j(u)$ is a bounded process. By Lemma A1, $\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\xi_i}{\tilde{\alpha}}dM_i(u)$ is asymptotically equivalent to $E\left((1-\Delta_i)dM_i(u)\right)$ in probability uniformly in u and $\frac{\hat{S}^{(1)}(\beta^*,u)}{\hat{S}^{(0)}(\beta^*,u)}\xrightarrow{P}e(\beta^*,u)$, uniformly in u. Hence, the above converges to zero in probability. Similarly, we have the fourth term of (6.15) converging to zero in probability. Let us look at the third term of (6.13).

$$n^{1/2}\int_0^t\left\{\frac{1}{\hat{S}^{(0)}(\hat{\beta}^{II},u)}-\frac{1}{\hat{S}^{(0)}(\beta_0,u)}\right\}S^{(0)}(\beta_0,u)d\mu_0(u)$$
$$=n^{1/2}\int_0^t\left\{-(\hat{\beta}^{II}-\beta_0)'\frac{\hat{S}^{(1)}(\beta^*,u)}{\hat{S}^{(0)}(\beta^*,u)}\right\}\hat{S}^{(0)}(\beta_0,u)d\mu_0(u).$$

Proceeding as earlier, by the consistency of $\hat{\beta}^{II}$, uniform consistency of $S^{(d)}(\beta,t)$ and $\hat{S}^{(d)}(\beta,t)$ and the boundedness of $\mu_0(t)$, we can further write

$$n^{1/2}\int_0^t\left\{-(\hat{\beta}^{II}-\beta_0)'\frac{\hat{S}^{(1)}(\beta^*,u)}{\hat{S}^{(0)}(\beta^*,u)}\right\}\hat{S}^{(0)}(\beta_0,u)d\mu_0(u)=n^{1/2}(\hat{\beta}^{II}-\beta_0)'r(\beta_0,t)+o_P(1),$$

where $r(\beta,t)=-\int_0^t s^{(1)}(\beta,u)d\mu_0(u)$. Now, for the fourth term of the equation (6.13), since $\hat{S}^{(0)}(\beta,t)$ converges to $s^{(0)}(\beta,t)$ uniformly in t and is bounded away from zero. $n^{-1/2}\sum_{i=1}^{n}M_i(t)$ converges to a zero-mean Gaussian process with continuous sample paths. Hence, $\int_0^t\frac{1}{\hat{S}^{(0)}(\beta_0,u)}\left\{n^{-1/2}\sum_{i=1}^{n}dM_i(u)\right\}$ can be written as $\int_0^t\frac{1}{s^{(0)}(\beta_0,u)}\left\{n^{-1/2}\sum_{i=1}^{n}dM_i(u)\right\}$. Now, for the last term in (6.13), we can note similarly that $S^{(0)}(\beta,t)$ converges uniformly to $s^{(0)}(\beta,t)$

and is bounded away from zero. Hence, one can write the fifth term of (6.13) as

$$\int_0^t \frac{1}{\hat{S}^{(0)}(\beta_0,u)}\left\{n^{-1/2}\sum_{i=1}^{n}(w_i^{II}(u)-1)dM_i(u)\right\}$$
$$= n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}\left\{(1-\Delta_i)(\frac{\xi_i}{\hat{\alpha}(t)}-1)Y_i(u)e^{\beta_0' Z_i(u)}d\mu_0(u)\right\}$$
$$+ n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}\left\{\Delta_i(1-\xi_i)(\frac{\eta_i}{\hat{q}(t)}-1)dM_i(u)\right\}$$
$$= n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}\left\{(1-\Delta_i)(\frac{\xi_i}{\tilde{\alpha}}-1)Y_i(u)e^{\beta_0' Z_i(u)}d\mu_0(u)\right\}$$
$$+ n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}(1-\Delta_i)\xi_i\left\{\frac{1}{\hat{\alpha}(t)}-\frac{1}{\tilde{\alpha}}\right\}Y_i(u)e^{\beta_0' Z_i(u)}d\mu_0(u)$$
$$+ n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}\left\{\Delta_i(1-\xi_i)(\frac{\eta_i}{\tilde{q}}-1)dM_i(u)\right\}$$
$$+ n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}\Delta_i(1-\xi_i)\eta_i\left\{\frac{1}{\hat{q}(t)}-\frac{1}{\tilde{q}}\right\}dM_i(u). \tag{6.16}$$

Using (6.1), (6.2) and Lemma A1, the second term of (6.16) $= n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}(1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})\frac{Y_i(u)E\left\{(1-\Delta_i)Y_i(u)e^{\beta_0' Z_i(u)}d\mu_0(u)\right\}}{E[(1-\Delta_i)Y_i(t)]}+o_P(1)$ and the third term $= n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}\Delta_i(1-\xi_i)(1-\frac{\eta_i}{\tilde{q}})Y_i(t)\frac{E\{\Delta_i(1-\xi_i)dM_i(u)\}}{E[\Delta_i(1-\xi_i)Y_i(t)]}+o_P(1)$. Therefore, the above equation can be rewritten as

$$\int_0^t \frac{1}{\hat{S}^{(0)}(\beta_0,u)}\left\{n^{-1/2}\sum_{i=1}^{n}(w_i^{II}(u)-1)dM_i(u)\right\}$$
$$= n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}(1-\Delta_i)(\frac{\xi_i}{\tilde{\alpha}}-1)\left\{Y_i(u)e^{\beta_0' Z_i(u)}-\frac{Y_i(u)E\left\{(1-\Delta_1)Y_1(u)e^{\beta_0' Z_1(u)}\right\}}{E\left[(1-\Delta_1)Y_1(u)\right]}\right\}$$
$$\times d\mu_0(t) + n^{-1/2}\sum_{i=1}^{n}\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}\Delta_i(1-\xi_i)(\frac{\eta_i}{\tilde{q}}-1)\left\{dM_i(u)-\frac{Y_i(u)E\left\{\Delta_1(1-\xi_1)dM_1(u)\right\}}{E\left[\Delta_1(1-\xi_1)Y_1(u)\right]}\right\}$$
$$+ o_P(1). \tag{6.17}$$

Combining all of the above, we have (6.13) as

$$n^{1/2}\{\hat{\mu}_0(\hat{\beta}^{II},t)-\mu_0(t)\}$$

$$\approx n^{1/2}(\hat{\beta}^{II}-\beta_0)'r(\beta_0,t)+\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}\left\{n^{-1/2}\sum_{i=1}^n dM_i(u)\right\}$$
$$+\,n^{-1/2}\sum_{i=1}^n\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}(1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})\left\{Y_i(u)e^{\beta_0' Z_i(u)}-\frac{Y_i(u)E\left\{(1-\Delta_1)Y_1(u)e^{\beta_0' Z_1(u)}\right\}}{E\left[(1-\Delta_1)Y_1(u)\right]}\right\}$$
$$\times\, d\mu_0(t)+n^{-1/2}\sum_{i=1}^n\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}\Delta_i(1-\xi_i)(\frac{\eta_i}{\tilde{q}}-1)\left\{dM_i(u)-\frac{Y_i(u)E\left\{\Delta_1(1-\xi_1)dM_1(u)\right\}}{E\left[\Delta_1(1-\xi_1)Y_1(u)\right]}\right\}. \tag{6.18}$$

We know from the asymptotic expansion of $n^{1/2}(\hat{\beta}^{II}-\beta_0)$ that

$$\begin{aligned}
&n^{1/2}(\hat{\beta}^{II}-\beta_0)\\
&=A(\beta_0)^{-1}n^{1/2}U_n(\beta_0)+o_P(1)\\
&=A(\beta_0)^{-1}\left(n^{-1/2}\sum_{i=1}^n M_{\tilde{Z},i}(\beta_0,t)+n^{-1/2}\sum_{i=1}^n\int_0^\tau\{1-\frac{\xi_i}{\tilde{\alpha}}\}(1-\Delta_i)\left(R_i(\beta_0,t)-Y_i(t)\right.\right.\\
&\left.\times\frac{E\left[(1-\Delta_i)R_i(\beta_0,t)\right]}{E\left[(1-\Delta_i)Y_i(t)\right]}\right)d\mu_0(t)-n^{-1/2}\sum_{i=1}^n\int_0^\tau\{1-\frac{\eta_i}{\tilde{q}}\}\Delta_i(1-\xi_i)\left(dM_{\tilde{Z},i}(\beta_0,t)-Y_i(t)\right.\\
&\left.\left.\frac{E\left[dM_{\tilde{Z},i}(\beta_0,t)\mid\Delta_i=1,\xi_i=0\right]}{E\left[Y_i(t)\mid\Delta_i=1,\xi_i=0\right]}\right)\right)+o_P(1).
\end{aligned}\tag{6.19}$$

Using equations (6.18) and (6.19), one can write

$$\begin{aligned}
&n^{1/2}\{\hat{\mu}_0(\hat{\beta}^{II},t)-\mu_0(t)\}\\
&=r(\beta_0,t)'A(\beta_0)^{-1}n^{1/2}U_n(\beta_0)+n^{1/2}(\hat{\beta}^{II}-\beta_0)'r(\beta_0,t)+\int_0^t\frac{1}{s^{(0)}(\beta_0,u)}\left\{n^{-1/2}\sum_{i=1}^n dM_i(u)\right\}\\
&+n^{-1/2}\sum_{i=1}^n\int_0^t\frac{1}{s^{(0)}(\beta_0,u)}(1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})\left\{Y_i(u)e^{\beta_0' Z_i(u)}-\frac{Y_i(u)E\left\{(1-\Delta_1)Y_1(u)e^{\beta_0' Z_1(u)}\right\}}{E\left[(1-\Delta_1)Y_1(u)\right]}\right\}d\mu_0(t)\\
&+n^{-1/2}\sum_{i=1}^n\int_0^t\frac{1}{s^{(0)}(\beta_0,u)}\Delta_i(1-\xi_i)(\frac{\eta_i}{\tilde{q}}-1)\left\{dM_i(u)-\frac{Y_i(u)E\left\{\Delta_1(1-\xi_1)dM_1(u)\right\}}{E\left[\Delta_1(1-\xi_1)Y_1(u)\right]}\right\}+o_P(1)\\
&=r(\beta_0,t)'A(\beta_0)^{-1}\left(n^{-1/2}\sum_{i=1}^n M_{\tilde{Z},i}(\beta_0)+n^{-1/2}\sum_{i=1}^n\int_0^\tau\{1-\frac{\xi_i}{\tilde{\alpha}}\}(1-\Delta_i)\left(R_i(\beta_0,t)-\right.\right.\\
&\left.\frac{Y_i(t)E\left[(1-\Delta_i)R_i(\beta,t)\right]}{E\left[(1-\Delta_i)Y_i(t)\right]}\right)d\mu_0(t)-n^{-1/2}\sum_{i=1}^n\int_0^\tau\{1-\frac{\eta_i}{\tilde{q}}\}\Delta_i(1-\xi_i)\left(dM_{\tilde{Z},i}(\beta_0,t)-\right.
\end{aligned}$$

$$\left.\left.\frac{Y_i(t)E\left[dM_{\tilde{Z},i}(\beta_0,t)\mid \Delta_i=1,\xi_i=0\right]}{E\left[Y_i(t)\mid \Delta_i=1,\xi_i=0\right]}\right)\right)+\int_0^t \frac{n^{-1/2}\sum_{i=1}^n dM_i(u)}{s^{(0)}(\beta_0,u)}$$

$$+n^{-1/2}\sum_{i=1}^n\int_0^t \frac{\{1-\frac{\xi_i}{\tilde{\alpha}}\}(1-\Delta_i)Y_i(u)}{s^{(0)}(\beta_0,u)}\left(e^{\beta_0 Z_i(u)}-\frac{E\left[(1-\Delta_i)Y_i(u)e^{\beta_0 Z_i(u)}\right]}{E\left[(1-\Delta_i)Y_i(u)\right]}\right)d\mu_0(u)$$

$$+n^{-1/2}\sum_{i=1}^n\int_0^t \frac{1}{s^{(0)}(\beta_0,u)}\Delta_i(1-\xi_i)(\frac{\eta_i}{\tilde{q}}-1)\left\{dM_i(u)-\frac{Y_i(u)E\left\{\Delta_1(1-\xi_1)dM_1(u)\right\}}{E\left[\Delta_1(1-\xi_1)Y_1(u)\right]}\right\}+o_P(1)$$

$$=n^{-1/2}\sum_{i=1}^n\left(r(\beta_0,t)'A(\beta_0)^{-1}\left(M_{\tilde{Z},i}(\beta_0)\right)+\int_0^t\frac{dM_i(u)}{s^{(0)}(\beta_0,u)}\right)$$

$$+n^{-1/2}\sum_{i=1}^n\{1-\frac{\xi_i}{\tilde{\alpha}}\}(1-\Delta_i)\left[r(\beta,t)'A(\beta)^{-1}\int_0^\tau\left(R_i(\beta_0,t)-\frac{Y_i(t)E\left[(1-\Delta_i)R_i(\beta_0,t)\right]}{E\left[(1-\Delta_i)Y_i(t)\right]}\right)d\mu_0(t)\right.$$

$$\left.+\int_0^t\frac{Y_i(u)}{s^{(0)}(\beta_0,u)}\left(e^{\beta_0 Z_i(u)}-\frac{E\left[(1-\Delta_i)Y_i(u)e^{\beta_0 Z_i(u)}\right]}{E\left[(1-\Delta_i)Y_i(u)\right]}\right)d\mu_0(u)\right]$$

$$+n^{-1/2}\sum_{i=1}^n\{\frac{\eta_i}{\tilde{q}}-1\}\Delta_i(1-\xi_i)\left[r(\beta,t)'A(\beta)^{-1}\int_0^\tau\left(dM_{\tilde{Z},i}(\beta_0,t)\right.\right.$$

$$\left.-\frac{Y_i(t)E\left[dM_{\tilde{Z},1}(\beta_0,t)(\beta_0,t)\mid \Delta_1=1,\xi_1=0\right]}{E\left[Y_1(t)\mid \Delta_1=1,\xi_1=0\right]}\right)$$

$$\left.+\int_0^t\frac{1}{s^{(0)}(\beta,u)}\left(dM_i(u)-\frac{Y_i(t)E\left[dM_1(u)(\beta_0,t)\mid \Delta_1=1,\xi_1=0\right]}{E\left[Y_1(t)\mid \Delta_1=1,\xi_1=0\right]}\right)d\mu_0(u)\right]+o_P(1)$$

$$=n^{-1/2}\sum_{i=1}^n\nu_i(\beta_0,t)+n^{-1/2}\sum_{i=1}^n\{1-\frac{\xi_i}{\tilde{\alpha}}\}(1-\Delta_i)\psi_i(\beta_0,t)-n^{-1/2}\sum_{i=1}^n\{1-\frac{\eta_i}{\tilde{q}}\}\Delta_i(1-\xi_i)\zeta_i(\beta_0,t)$$

$$+o_p(1),$$

where $\nu_i(\beta_0,t)=\left(r(\beta_0,t)'A(\beta_0)^{-1}\left(M_{\tilde{Z},i}(\beta_0)\right)+\int_0^t\frac{dM_i(u)}{s^{(0)}(\beta_0,u)}\right)$, $\psi_i(\beta_0,t)=\Big[r(\beta,t)'A(\beta)^{-1}\int_0^\tau \left(R_i(\beta,t)-\frac{Y_i(t)E[(1-\Delta_i)R_i(\beta,t)]}{E[(1-\Delta_i)Y_i(t)]}\right)d\mu_0(t)+\int_0^t\frac{Y_i(u)}{s^{(0)}(\beta,u)}\left(e^{\beta Z_i(u)}-\frac{E\left[(1-\Delta_i)Y_i(u)e^{\beta Z_i(u)}\right]}{E[(1-\Delta_i)Y_i(u)]}\right)d\mu_0(u)\Big]$, $\zeta(\beta_0,t)=[r(\beta,t)'A(\beta)^{-1}\int_0^\tau\left(dM_{\tilde{Z},i}(\beta_0,t)-\frac{Y_i(t)E\left[dM_{\tilde{Z},1}(\beta_0,t)(\beta_0,t)\mid\Delta_1=1,\xi_1=0\right]}{E[Y_1(t)\mid\Delta_1=1,\xi_1=0]}\right)+\int_0^t\frac{1}{s^{(0)}(\beta,u)}\left(dM_i(u)-\frac{Y_i(t)E[dM_1(u)(\beta_0,t)\mid\Delta_1=1,\xi_1=0]}{E[Y_1(t)\mid\Delta_1=1,\xi_1=0]}\right)]$.

Note that, since $r(\beta_0,t)$, $A(\beta_0)$, $E\left[(1-\Delta_i)R_i(\beta,t)\right]$, $E\left[(1-\Delta_i)Y_i(u)e^{\beta Z_i(u)}\right]$, $E\left[dM_{\tilde{Z},i}(\beta,t)\mid \Delta_i=1,\xi_i=0\right]$ and $E\left[dM_i(t)\mid \Delta_i=1,\xi_i=0\right]$ are bounded and $s^{(0)}(\beta_0,u)$, $E\left[(1-\Delta_i)Y_i(u)\right]$ $E\left[Y_i(u)\mid \Delta_i=1,\xi_i=0\right]$ are bounded away from zero along with the asymptotic normality of the rest of the terms, we can say that $\nu_i(\beta_0,t)$, $\psi_i(\beta_0,t)$ and $\zeta_i(\beta_0,t)$ are bounded processes. For any finite number of time-points, the joint distribution of $n^{1/2}W_n(t)=$

$n^{1/2}\{\hat{\mu}_0(\hat{\beta}^{II},t)-\mu_0(t)\}$ for various t would be asymptotically equivalent to a zero-mean Gaussian process. Finally, let us look at the covariance between $n^{-1/2}\sum_{i=1}^n \nu_i(\beta_0,t)$, $n^{-1/2}\sum_{i=1}^n\{1-\frac{\xi_i}{\tilde{\alpha}}\}(1-\Delta_i)\psi_i(\beta_0,t)$ and $n^{-1/2}\sum_{i=1}^n\{1-\frac{\eta_i}{\tilde{q}}\}\Delta_i(1-\xi_i)\zeta_i(\beta_0,t)$. The covariance between the first two quantities is

$$\begin{aligned}
&cov(n^{-1/2}\sum_{i=1}^n \nu_i(\beta_0,t), n^{-1/2}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)\psi_i(\beta_0,t))\\
&=E\left(n^{-1/2}\sum_{i=1}^n \nu_i(\beta_0,t)n^{-1/2}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)\psi_i(\beta_0,t)\right)\\
&=E\left[E\left(n^{-1/2}\sum_{i=1}^n \nu_i(\beta_0,t)n^{-1/2}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)\psi_i(\beta_0,t)\;\Big|\;\mathcal{F}(\tau)\right)\right]\\
&=E\left[n^{-1/2}\sum_{i=1}^n \nu_i(\beta_0,t)n^{-1/2}\sum_{i=1}^n E\left(1-\frac{\xi_i}{\tilde{\alpha}}\;\Big|\;\mathcal{F}(\tau)\right)(1-\Delta_i)\psi_i(\beta_0,t)\right]=0
\end{aligned}$$

Similarly, the covariance between the first and third quantities is given by

$$\begin{aligned}
&cov(n^{-1/2}\sum_{i=1}^n \nu_i(\beta_0,t), n^{-1/2}\sum_{i=1}^n(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)\zeta_i(\beta_0,t))\\
&=E\left(n^{-1/2}\sum_{i=1}^n \nu_i(\beta_0,t)n^{-1/2}\sum_{i=1}^n(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)\zeta_i(\beta_0,t)\right)\\
&=E\left[E\left(n^{-1/2}\sum_{i=1}^n \nu_i(\beta_0,t)n^{-1/2}\sum_{i=1}^n(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)\zeta_i(\beta_0,t)\;\Big|\;\mathcal{F}(\tau)\right)\right]\\
&=E\left[n^{-1/2}\sum_{i=1}^n \nu_i(\beta_0,t)n^{-1/2}\sum_{i=1}^n E\left((1-\xi_i)(1-\frac{\eta_i}{\tilde{q}})\;\Big|\;\mathcal{F}(\tau)\right)\Delta_i\zeta_i(\beta_0,t)\right]=0
\end{aligned}$$

Similarly the covariance between $n^{-1/2}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)\psi_i(\beta_0,t)$ and $n^{-1/2}\sum_{i=1}^n(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)\zeta_i(\beta_0,t)=0$. Hence, the covariance of $n^{1/2}W_n(t)$ and $n^{1/2}W_n(s)$ is given by $E\left(\nu_i(\beta_0,t))\nu_i(\beta_0,s)\right)+\frac{1-\alpha}{\alpha}E\left((1-\Delta_i)\psi_i(\beta_0,t))\psi_i(\beta_0,s)\right)$
$+(1-\alpha)\frac{1-q}{q}P(\Delta_i=1)E\left(\psi_i(\beta_0,t))\psi_i(\beta_0,s)\mid\Delta_i=1,\xi_i=0\right)$.

APPENDIX B: TECHNICAL DETAILS FOR CHAPTER 4

Here is the Appendix containing the proofs of the theorems in Chapter 4.

B.1 Regularity Conditions

(i) $(T_i^*, C_i, Z_i(t))\ \forall i = 1, 2, \dots, n$ are independent and identically distributed.

(ii) $P(Y(\tau) > 0) > 0$ and $N_i(\tau)$ $(\forall i = 1, 2, \dots, n)$ are bounded by a constant.

(iii) $\mid Z_i(0) \mid + \int_0^\tau \mid dZ_i(u) \mid < C_z < \infty$ almost surely for some constant C_z.

(iv) The matrix $A = E\left[\int_0^\tau Y_i(t)\left\{Z_i(t) - e(t)\right\}^{\otimes 2} dt\right]$ (where $e(t) = \frac{E(Y_i(t)Z_i(t))}{E((Y_i(t))}$) is positive definite.

(v) (Finite Interval) $\int_0^\tau d\mu_0(t) < \infty$.

(vi) As $n \to \infty$, $\frac{\tilde{n}}{n} = \tilde{\alpha} \to \alpha \in (0,1)$, $\tilde{q} \to q$.

(vii) (Asymptotic stability)As $n \to \infty$, we have the following:

(a) There exists a positive definite matrix, $V^I(\beta_0)$, such that

$$var\left[n^{-1/2}\sum_{i=1}^{n}\int_0^\tau\left(dR_i(\beta_0,t) - \frac{Y_i(t)E\left\{1-\Delta_1\right\}dR_1(\beta_0,t)}{E\left\{(1-\Delta_1)Y_1(t)\right\}}\right)\right] \xrightarrow{P} V^I(\beta_0).$$

(b) There is a positive definite matrix, $V^{II}(\beta_0)$ such that

$$var\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(M_{\tilde{Z},i}(\beta_0) - \int_0^\tau \frac{Y_i(t)E\left\{dM_{\tilde{Z},1}(\beta_0,t) \mid \Delta_1 = 1, \xi_1 = 0\right\}}{E\left\{Y_1(t) \mid \Delta_1 = 1\right\}}\right)\right.$$

$$\left. \mid \Delta_i = 1, \xi_i = 0\right] \xrightarrow{P} V^{II}(\beta_0).$$

(i) $\sup_{0\le t\le\tau}\|W_n(t) - W(t)\| \xrightarrow{P} 0$ for some bounded process, $W(t)$,

(ii) $W_n(t)$ is monotone on $[0,\tau]$ and

(iii) $G_n(t)$ converges to a zero-mean process with continuous sample paths. Then

$$\sup_{0\le s\le\tau}\|\int_0^s (W_n(t)-W(t))dG_n(t)\| \xrightarrow{P} 0, \quad \sup_{0\le s\le\tau}\|\int_0^s (G_n(t)-G(t))dW_n(t)\| \xrightarrow{P} 0.$$

This was stated in Kang and Cai (2009b).

Lemma A3 :

Let $B_i(t), i = 1, 2, \ldots, n$ be i.i.d. real-valued random processes on $[0,\tau]$ with $E(B_i(t)) = \mu_B(t)$, $var(B_i(0)) < \infty$, $var(B_i(\tau)) < \infty$ and suppose that almost all paths of $B_i(t)$ have finite variation. Then, $n^{-1/2}\sum_{i=1}^n \{B_i(t)-\mu_B(t)\}$ converges weakly to a zero-mean Gaussian process in $l^\infty[0,\tau]$ and therefore, $n^{-1}\sum_{i=1}^n \{B_i(t)-\mu_B(t)\}$ converges in probability to zero, uniformly in t. This lemma has been stated in Kulich and Lin (2004) as a proposition.

First we will look at the asymptotic properties of the time-varying sampling weights. More specifically, $\hat{\alpha}(t) = \frac{\sum_{i=1}^n (1-\Delta_i)\xi_i Y_i(t)}{\sum_{i=1}^n (1-\Delta_i)Y_i(t)}$ and $\hat{q}(t) = \frac{\sum_{i=1}^n \Delta_i(1-\xi_i)\eta_i Y_i(t)}{\sum_{i=1}^n \Delta_i(1-\xi_i)Y_i(t)}$. Looking at the Taylor series expansion of $\hat{\alpha}(t)^{-1}$ around $\tilde{\alpha}^{-1}$, we have

$$\hat{\alpha}(t)^{-1}-\tilde{\alpha}^{-1} = -\frac{1}{\alpha^*(t)^2}(\hat{\alpha}(t)-\tilde{\alpha}) = \frac{\tilde{\alpha}}{\alpha^*(t)^2}\frac{1}{\sum_{i=1}^n (1-\Delta_i)Y_i(t)}\left[\sum_{i=1}^n (1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})Y_i(t)\right],$$

where $\alpha^*(t)$ is on the line between $\hat{\alpha}(t)$ and $\tilde{\alpha}$. Now by the Glivenko-Cantelli theorem, one can show that $\frac{1}{n}\sum_{i=1}^n (1-\Delta_i)Y_i(t)$ converges to $E\left((1-\Delta_i)Y_i(t)\right)$. $(1-\Delta_i)Y_i(t)$ is bounded and monotone functions of t. They are also independent of ξ_i. Hence, by Lemma A1 and noting that when sampling from a finite population $\mu_B(t) = \sum_{i=1}^n B_i(t)$ where $B_i(t) = (1-\Delta_i)Y_i(t)$, $n^{-1/2}\sum_{i=1}^n (1-\frac{\xi_i}{\tilde{\alpha}})\left[(1-\Delta_i)Y_i(t)\right]$ converges weakly to a zero-mean Gaussian process(Kang and Cai 2009b). This implies that $\frac{1}{n}\sum_{i=1}^n (1-\frac{\xi_i}{\tilde{\alpha}})\left[(1-\Delta_i)Y_i(t)\right] \xrightarrow{P} 0$ uniformly in t. Further, $\hat{\alpha}(t)$, $\alpha^*(t)$ and $\tilde{\alpha}$ converges to the same limit. Using Slutsky's theorem, we have

$$\begin{aligned}
&n^{1/2}(\hat{\alpha}(t)^{-1}-\tilde{\alpha}^{-1})\\
&= n^{1/2}\frac{\tilde{\alpha}}{\alpha^*(t)^2}\frac{1}{\sum_{i=1}^n (1-\Delta_i)Y_i(t)}\left[\sum_{i=1}^n (1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})Y_i(t)\right]\\
&= \frac{1}{\tilde{\alpha}E((1-\Delta_i)Y_i(t))}\left[n^{-1/2}\sum_{i=1}^n (1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})Y_i(t)\right]
\end{aligned}$$

$$+\tilde{\alpha}\left(\frac{1}{\alpha^*(t)^2}\frac{1}{\sum_{i=1}^n(1-\Delta_i)Y_i(t)}-\frac{1}{\tilde{\alpha}^2E((1-\Delta_i)Y_i(t))}\right)\left[n^{-1/2}\sum_{i=1}^n(1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})Y_i(t)\right]$$

$$=\frac{1}{\tilde{\alpha}E((1-\Delta_i)Y_i(t))}\left[n^{-1/2}\sum_{i=1}^n(1-\Delta_i)(1-\frac{\xi_i}{\tilde{\alpha}})Y_i(t)\right]+o_P(1). \quad (6.20)$$

Similarly, $\hat{q}(t)^{-1}-\tilde{q}^{-1}=-\frac{1}{q^*(t)^2}(\hat{q}(t)-\tilde{q})=\frac{\tilde{q}}{q^*(t)^2}\frac{1}{\sum_{i=1}^n\Delta_i(1-\xi_i)Y_i(t)}\left[\sum_{i=1}^n\Delta_i(1-\xi_i)(1-\frac{\eta_i}{\tilde{q}})Y_i(t)\right]$, where $q^*(t)$ is on the line between $\hat{q}(t)$ and $\tilde{q}$. Proceeding as before, we can see that $\hat{q}(t)$, $q^*(t)$ and $\tilde{q}$ converges to the same limit. Using Slutsky's theorem, we have

$$n^{1/2}(\hat{q}(t)^{-1}-\tilde{q}^{-1})=\frac{1}{\tilde{q}E(\Delta_i(1-\xi_i)Y_i(t))}\left[n^{-1/2}\sum_{i=1}^n\Delta_i(1-\xi_i)(1-\frac{\eta_i}{\tilde{q}})Y_i(t)\right]+o_P(1). \quad (6.21)$$

B.2 Proof of Theorem 3

: Let us define $U_n(\beta)=\frac{1}{n}U^{II}(\beta)$. Based on similar arguments, as in Foutz(1977), the consistency of $\hat{\beta}^{II}$ can be shown by proving the following : (a) $\frac{\delta}{\delta\beta'}U_n(\beta)$ exists and is continuous in an open neighborhood of β_0 in $\mathcal{B}$, (b) $\frac{\delta}{\delta\beta_0'}U_n(\beta_0)$ is negative definite w.p. $\to 1$ as $n\to\infty$, (c) - $\frac{\delta}{\delta\beta'}U_n(\beta)\xrightarrow{P}A(\beta_0)$ uniformly for β in a neighborhood of β_0 and (d) $U_n(\beta)\xrightarrow{P}0$.

Taking derivative of the expression, we get

$$\frac{\delta}{\delta\beta'}U_n(\beta)=-\frac{1}{n}\sum_{i=1}^n\int_0^\tau w_i^{II}(t)\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}Y_i(t)Z_i(t)'dt \quad (6.22)$$

We need to show that this goes to A in probability. Using the above formula, one can rewrite

$$-\frac{\delta}{\delta\beta'}U_n(\beta)=\frac{1}{n}\sum_{i=1}^n\int_0^\tau w_i^{II}(t)\{Z_i(t)^{\otimes 2}-\bar{Z}^{II}(t)^{\otimes 2}\}Y_i(t)dt. \quad (6.23)$$

Hence, condition (a) of the consistency of $\hat{\beta}^{II}$ is satisfied. To prove the other conditions, we will first start with proving

$$\sup_{t\in[0,\tau]}\|\bar{Z}^{II}(t)-e(t)\|\xrightarrow{P}0 \text{ as n } \to\infty.$$

It suffices to show that

$$\sup_{t\in[0,\tau]} \|\frac{1}{n}\sum_{i=1}^{n} w_i^{II}(t)Y_i(t)Z_i(t)^{\otimes d} - \frac{1}{n}\sum_{i=1}^{n} Y_i(t)Z_i(t)^{\otimes d}\| \xrightarrow{P} 0 \text{ as n } \to \infty \text{ for } d = 0, 1.$$

Let us start with $n^{-1/2}\sum_{i=1}^{n} w_i^{II}(t)Y_i(t)Z_i(t)^{\otimes d} - n^{-1/2}\sum_{i=1}^{n} Y_i(t)Z_i(t)^{\otimes d}$. We can note that

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n} w_i^{II}(t)Y_i(t)Z_i(t)^{\otimes d} - n^{-1/2}\sum_{i=1}^{n} Y_i(t)Z_i(t)^{\otimes d}\\
&= n^{-1/2}\sum_{i=1}^{n}\{\Delta_i\xi_i + \frac{(1-\Delta_i)\xi_i}{\hat{\alpha}(t)} + \frac{\Delta_i(1-\xi_i)\eta_i}{\hat{q}(t)}\}Y_i(t)Z_i(t)^{\otimes d} - n^{-1/2}\sum_{i=1}^{n} Y_i(t)Z_i(t)^{\otimes d}\\
&= n^{-1/2}\sum_{i=1}^{n}(1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}} - 1\}Y_i(t)Z_i(t)^{\otimes d} + n^{-1/2}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}} - 1\}Y_i(t)Z_i(t)^{\otimes d}\\
&- n^{-1/2}\sum_{i=1}^{n}\frac{Y_i(t)}{E\left[(1-\Delta_1)Y_1(t)\right]}(1-\Delta_i)\left(\frac{\xi_i}{\tilde{\alpha}} - 1\right)\left\{\frac{1}{n}\sum_{j=1}^{n}(1-\Delta_j)\frac{\xi_j}{\tilde{\alpha}}Y_j(t)Z_j(t)^{\otimes d}\right\}\\
&- n^{-1/2}\sum_{i=1}^{n}\frac{Y_i(t)}{E\left[\Delta_1(1-\xi_1)Y_1(t)\right]}\Delta_i(1-\xi_i)\left(\frac{\eta_i}{\tilde{q}} - 1\right)\left\{\frac{1}{n}\sum_{j=1}^{n}\Delta_j(1-\xi_j)\frac{\eta_j}{\tilde{q}}Y_j(t)Z_j(t)^{\otimes d}\right\} + o_P(1).
\end{aligned} \tag{6.24}$$

For each j, from (iii) of the regularity conditions, one can note that the total variation of $\Delta_i(1-\xi_i)Y_i(t)Z_i(t)^{\otimes 2}$ and $(1-\Delta_i)Y_i(t)Z_i(t)^{\otimes 2}$ are finite for $t \in [0,\tau]$. From Lemma A1, $\frac{1}{n}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\frac{\eta_i}{\tilde{q}}Y_i(t)Z_i(t)^{\otimes d}$ converges to some finite quantity $E\left(\Delta_i(1-\xi_i)Y_i(t)Z_i(t)^{\otimes d}\right)$, in probability uniformly in t. Similarly, we know that $\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\frac{\xi_i}{\tilde{\alpha}}Y_i(t)Z_i(t)^{\otimes d} \xrightarrow{P} E\left((1-\Delta_i)Y_i(t)Z_i(t)^{\otimes d}\right)$. Therefore,

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n} w_i^{II}(t)Y_i(t)Z_i(t)^{\otimes d} - n^{-1/2}\sum_{i=1}^{n} Y_i(t)Z_i(t)^{\otimes d}\\
&= n^{-1/2}\sum_{i=1}^{n}(1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}} - 1\}Y_i(t)\left\{Z_i(t)^{\otimes d} - \frac{E\left((1-\Delta_i)Y_i(t)Z_i(t)^{\otimes d}\right)}{E\left[(1-\Delta_1)Y_1(t)\right]}\right\}\\
&+ n^{-1/2}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}} - 1\}Y_i(t)\left\{Z_i(t)^{\otimes d} - \frac{E\left(\Delta_i(1-\xi_i)Y_i(t)Z_i(t)^{\otimes d}\right)}{E\left[\Delta_1(1-\xi_1)Y_1(t)\right]}\right\} + o_P(1).
\end{aligned} \tag{6.25}$$

Now, note that, $E\left[(1-\Delta_i)Y_i(t)\left\{Z_i(t)^{\otimes d}-\frac{E\left((1-\Delta_i)Y_i(t)Z_i(t)^{\otimes d}\right)}{E[(1-\Delta_1)Y_1(t)]}\right\}\right]=0$ and $E\left[\Delta_i(1-\xi_i)Y_i(t)\times\left\{Z_i(t)^{\otimes d}-\frac{E\left(\Delta_i(1-\xi_i)Y_i(t)Z_i(t)^{\otimes d}\right)}{E[\Delta_1(1-\xi_1)Y_1(t)]}\right\}\right]=0$. Hence, by Lemma A1, the terms on the RHS of (6.25): $n^{-1/2}\sum_{i=1}^n(1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}}-1\}Y_i(t)\left\{Z_i(t)^{\otimes d}-\frac{E\left((1-\Delta_i)Y_i(t)Z_i(t)^{\otimes d}\right)}{E[(1-\Delta_1)Y_1(t)]}\right\}$ and $n^{-1/2}\sum_{i=1}^n\Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}}-1\}Y_i(t)\times\left\{Z_i(t)^{\otimes d}-\frac{E\left(\Delta_i(1-\xi_i)Y_i(t)Z_i(t)^{\otimes d}\right)}{E[\Delta_1(1-\xi_1)Y_1(t)]}\right\}$ converge weakly to zero-mean Gaussian processes. Hence, $\frac{1}{n}\sum_{i=1}^n(1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}}-1\}Y_i(t)\left\{Z_i(t)^{\otimes d}-\frac{E\left((1-\Delta_i)Y_i(t)Z_i(t)^{\otimes d}\right)}{E[(1-\Delta_1)Y_1(t)]}\right\}$ and $\frac{1}{n}\sum_{i=1}^n\Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}}-1\}Y_i(t)\times\left\{Z_i(t)^{\otimes d}-\frac{E\left(\Delta_i(1-\xi_i)Y_i(t)Z_i(t)^{\otimes d}\right)}{E[\Delta_1(1-\xi_1)Y_1(t)]}\right\}$ converge to zero in probability uniformly in t. Therefore, by Slutsky's theorem, we have

$$\left\|\frac{1}{n}\sum_{i=1}^n w_i^{II}(t)Y_i(t)Z_i(t)^{\otimes d}-\frac{1}{n}\sum_{i=1}^n Y_i(t)Z_i(t)^{\otimes d}\right\|\xrightarrow{P}0\text{ n }\rightarrow\infty\text{ for d }=0,1. \tag{6.26}$$

Therefore, $\frac{1}{n}\sum_{i=1}^n w_i^{II}(t)Y_i(t)Z_i(t)^{\otimes d}$ and $\frac{1}{n}\sum_{i=1}^n Y_i(t)Z_i(t)^{\otimes d}$ converges to the same limit uniformly. Also, since $E(Y_i(t))$ is bounded away from zero by regularity condition (ii), we have from the above convergence results,

$$\sup_{t\in[0,\tau]}\|\bar{Z}^{II}(t)-e(t)\|\xrightarrow{P}0\text{ as n }\rightarrow\infty$$

Now, one can rewrite $-\frac{\delta}{\delta\beta_0'}U_n(\beta)$ as the following :

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\int_0^\tau w_i^{II}(t)\{Z_i(t)^{\otimes 2}-\bar{Z}^{II}(t)^{\otimes 2}\}Y_i(t)dt\\
&=\frac{1}{n}\sum_{i=1}^n\int_0^\tau(1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}}-1\}Y_i(t)\{Z_i(t)^{\otimes 2}-\bar{Z}^{II}(t)^{\otimes 2}\}dt\\
&+\frac{1}{n}\sum_{i=1}^n\int_0^\tau(1-\Delta_i)\xi_i\{\hat{\alpha}(t)^{-1}-\tilde{\alpha}^{-1}\}Y_i(t)\{Z_i(t)^{\otimes 2}-\bar{Z}^{II}(t)^{\otimes 2}\}dt\\
&+\frac{1}{n}\sum_{i=1}^n\int_0^\tau\Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}}-1\}Y_i(t)\{Z_i(t)^{\otimes 2}-\bar{Z}^{II}(t)^{\otimes 2}\}dt\\
&+\frac{1}{n}\sum_{i=1}^n\int_0^\tau\Delta_i(1-\xi_i)\eta_i\{\hat{q}(t)^{-1}-\tilde{q}^{-1}\}Y_i(t)\{Z_i(t)^{\otimes 2}-\bar{Z}^{II}(t)^{\otimes 2}\}dt\\
&+\frac{1}{n}\sum_{i=1}^n\int_0^\tau\{Z_i(t)^{\otimes 2}-\bar{Z}^{II}(t)^{\otimes 2}\}Y_i(t)dt.
\end{aligned}$$

By the fact that $\bar{Z}^{II}(t)$ converges uniformly in t to $e(t)$, the first term of the right hand side quantity is asymptotically equivalent to $\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau (1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}}-1\}Y_i(t)\{Z_i(t)^{\otimes 2}-e(t)^{\otimes 2}\}dt$ and the third term converges to $\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}}-1\}Y_i(t)\{Z_i(t)^{\otimes 2}-e(t)^{\otimes 2}\}dt$. Further, $\hat{\alpha}(t)^{-1}-\tilde{\alpha}^{-1}\xrightarrow{P}0$ and $\hat{q}(t)^{-1}-\tilde{q}^{-1}\xrightarrow{P}0$. Both $(1-\Delta_i)\xi_i Y_i(t)Z_i(t)^{\otimes 2}$ and $\Delta_i(1-\xi_i)\eta_i Y_i(t)Z_i(t)^{\otimes 2}$ have bounded variation and are monotone functions in t. Hence, $\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)\xi_i Y_i(t)Z_i(t)^{\otimes 2}\xrightarrow{P}0$ and $\frac{1}{n}\sum_{i=1}^{n}\Delta_i(1-\xi_i)\eta_i Y_i(t)Z_i(t)^{\otimes 2}\xrightarrow{P}0\ \forall t$. Therefore, we have the second term and fourth terms converging to zero in probability. In other words,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau w_i^{II}(t)\{Z_i(t)^{\otimes 2}-\bar{Z}^{II}(t)^{\otimes 2}\}Y_i(t)dt\\
&=\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau (1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}}-1\}Y_i(t)\{Z_i(t)^{\otimes 2}-e(t)^{\otimes 2}\}dt\\
&+\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}}-1\}Y_i(t)\{Z_i(t)^{\otimes 2}-e(t)^{\otimes 2}\}dt+\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \{Z_i(t)^{\otimes 2}-e(t)^{\otimes 2}\}Y_i(t)dt\\
&+o_P(1).
\end{aligned}
$$

Since, $(1-\Delta_i)Y_i(t)\left[Z_i(t)^{\otimes 2}-e(t)^{\otimes 2}\right]$ and $\Delta_i Y_i(t)\left[Z_i(t)^{\otimes 2}-e(t)^{\otimes 2}\right]$ have bounded variation and each of these terms are independent and identically distributed. Further, note that $\frac{1}{n}\sum_{i=1}^{n}\Delta_i Y_i(t)\left[Z_i(t)^{\otimes 2}-e(t)^{\otimes 2}\right]=\frac{1}{n(1-\tilde{\alpha})}\sum_{i=1}^{n}(1-\xi_i)\Delta_i Y_i(t)\left[Z_i(t)^{\otimes 2}-e(t)^{\otimes 2}\right]$ since ξ_i's are simple random sample from the finite population. It follows from Lemma A1, $\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau (1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}}-1\}Y_i(t)\{Z_i(t)^{\otimes 2}-e(t)^{\otimes 2}\}dt\xrightarrow{P}E\left[\int_0^\tau (1-\Delta_i)\{\frac{\xi_i}{\tilde{\alpha}}-1\}Y_i(t)\{Z_i(t)^{\otimes 2}\right.$ $\left.-e(t)^{\otimes 2}\}\right]dt$ and $\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \Delta_i(1-\xi_i)\{\frac{\eta_i}{\tilde{q}}-1\}Y_i(t)\{Z_i(t)^{\otimes 2}-e(t)^{\otimes 2}\}dt\xrightarrow{P}E\left[\int_0^\tau \Delta_i(1-\xi_i)\right.$ $\left.\{\frac{\eta_i}{\tilde{q}}-1\}Y_i(t)\ \times\{Z_i(t)^{\otimes 2}-e(t)^{\otimes 2}\}\right]dt$, which are equal to zero. Hence,

$$
-\frac{\delta}{\delta\beta_0'}U_n(\beta)\xrightarrow{P}A \quad \text{as n } \to\infty. \tag{6.27}
$$

Thus, conditions (b) and (c) of the consistency of $\hat{\beta}^{II}$ are satisfied. For condition (d), we start by looking at $n^{1/2}U_n(\beta)$. Note that $n^{1/2}U_n(\beta)$ can be decomposed into

$$
n^{-1/2}\sum_{i=1}^{n}\int_0^\tau w_i^{II}(t)\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}\left(dM_i(t)+Y_i(t)d\mu_0(t)\right)
$$

$$= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}dM_i(t)+n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left(w_i^{II}(t)-1\right)\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}dM_i(t). \tag{6.28}$$

The first part of the above equation can be further decomposed into

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}dM_i(t) \\ & = n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{Z_i(t)-\bar{Z}(t)\right\}dM_i(t)+n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\bar{Z}(t)-\bar{Z}^{II}(t)\right\}dM_i(t). \end{aligned} \tag{6.29}$$

The first quantity on the right hand side is the pseudo partial likelihood score function for the entire cohort data. Note that for fixed t, $n^{-1/2}\sum_{i=1}^{n}M_i(t)$ is the sum of zero-mean independent and identically distributed random variables. From the regularity conditions (iii) and (v) and by CLT, $M_i(t)$ has bounded variation and hence, from Lemma A3, we have that $n^{-1/2}\sum_{i=1}^{n}M_i(t)$ converges weakly to a zero-mean Gaussian process, say $\mathcal{W}_M(t)$ (Schaubel et al. 2006). From the functional Central Limit Theorem (Pollard, 1990 page 53) we can conclude that $n^{-1/2}\sum_{i=1}^{n}M_i(t)$ is tight and converges weakly to $\mathcal{W}_M(t)$. It can be shown that $E\left\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\right\}^4 = 3\left(E\left\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\right\}^2\right)^2$ by properties of a Gaussian distribution. Further, $E\left\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\right\}^2 = E\left\{\mathcal{W}_M(t)\right\}^2+E\left\{\mathcal{W}_M(s)\right\}^2-2E\left\{\mathcal{W}_M(t)\mathcal{W}_M(s)\right\}^2 = E\left\{\mathcal{W}_M(t)\right\}^2-E\left\{\mathcal{W}_M(s)\right\}^2$ for $s\le t$, since $E\left\{\mathcal{W}_M(t)\right\}^2 = E\left\{M_i(t)\right\}^2 = E\left(\int_0^t Y_i(u)\left\{d\mu_0(u) + \beta_0' Z_i(u)du\right\}\right)$ and $E\left\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\right\}^2 = E\left(\int_s^t Y_i(u)\left\{d\mu_0(u)+\beta_0' Z_i(u)du\right\}\right)$. Note that the conditions (iii) and (v) ensure boundedness of $\mu_0(t)$ and $\beta_0' Z_i(t)$ in $[0,\tau]$. Thus, by the Mean Value theorem, there is a constant such that, $E\left\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\right\}^2 = E\left(\int_s^t Y_i(u)\left\{d\mu_0(u)+\beta_0' Z_i(u)du\right\}\right) \le K(t-s)$ for $s\le t$ and $E\left\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\right\}^4 \le 3\left(E\left\{\mathcal{W}_M(t)-\mathcal{W}_M(s)\right\}^2\right)^2 \le K^*(t-s)^2$. Then, by the Kolmogorov-Centsov Theorem (Karatzas and Shereve 1988), $\mathcal{W}_M(t)$ has continuous sample paths. We can also note that, $\frac{1}{n}\sum_{i=1}^{n}Y_i(t)Z_i(t)$ and $\frac{1}{n}\sum_{i=1}^{n}Y_i(t)$ have bounded variations and $\frac{1}{n}\sum_{i=1}^{n}Y_i(t)$ is bounded away from zero from the regularity conditions. By Lin et al. (2000), $\bar{Z}(t)$ is of bounded variation and can be written as the difference of two monotone functions in t. Hence, one can write $\bar{Z}(t) = Z_1^*(t) - Z_2^*(t)$ where each of the functions are non-negative, monotone in t and bounded.

Since, $\bar{Z}^{II}(t)$ is also of bounded variation by the same arguments and equation (6.26), one can express that as the difference of two monotone bounded functions in t. We can rewrite $n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{\bar{Z}(t)-\bar{Z}^{II}(t)\right\}dM_i(t) = n^{1/2}\int_0^{\tau}\left\{\bar{Z}(t)-e(t)\right\}d\bar{M}(t) - n^{1/2}\int_0^{\tau}\left\{\bar{Z}^{II}(t)-e(t)\right\}d\bar{M}(t)$, where $\bar{M}(t)=\frac{1}{n}\sum_{i=1}^{n}M_i(t)$. Noting that $n^{1/2}\bar{M}(t)$ is bounded and the uniform convergence of $\bar{Z}(t)$ and $\bar{Z}^{II}(t)$ to $e(t)$, both the terms on the RHS of (6.29) converges to zero in probability by Lemma A2 as n $\rightarrow\infty$. Therefore, $n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}\times dM_i(t)$ is asymptotically equivalent to $n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\{Z_i(t)-e(t)\}dM_i(t)$. The second term of (6.28) can be rewritten as

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left(w_i^{II}(t)-1\right)\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}dM_i(t) \\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left(\frac{\xi_i}{\tilde{\alpha}}-1\right)(1-\Delta_i)\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}dM_i(t) \\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left(\frac{\eta_i}{\tilde{q}}-1\right)\Delta_i(1-\xi_i)\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}dM_i(t) \\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left(\hat{\alpha}(t)^{-1}-\tilde{\alpha}^{-1}\right)\xi_i(1-\Delta_i)\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}dM_i(t) \\
&+ n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left(\hat{q}(t)^{-1}-\tilde{q}^{-1}\right)\eta_i\Delta_i(1-\xi_i)\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}dM_i(t). \qquad (6.30)
\end{aligned}
$$

From the uniform convergence of $\bar{Z}^{II}(t)$ to $e(t)$, the first term on the right hand side of 6.30 can be written as

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left(\frac{\xi_i}{\tilde{\alpha}}-1\right)(1-\Delta_i)\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}dM_i(t) \\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left(\frac{\xi_i}{\tilde{\alpha}}-1\right)(1-\Delta_i)\{Z_i(t)-e(t)\}dM_i(t)+o_P(1) \\
&= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left(1-\frac{\xi_i}{\tilde{\alpha}}\right)(1-\Delta_i)\{Z_i(t)-e(t)\}Y_i(t)\left[d\mu_0(t)+\beta_0' Z_i(t)dt\right]+o_P(1).
\end{aligned}
$$

The last equality holds because only those individuals with $\Delta_i=0$ contribute to the summation, i.e., only individuals who did not experience a single event. By (6.20), (6.21), and the fact that $\bar{Z}^{II}(t)$ converges uniformly to $e(t)$ in probability, we have the third term of (6.30):

$$
\begin{aligned}
& n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \left(\hat{\alpha}(t)^{-1} - \tilde{\alpha}^{-1}\right) \xi_i (1-\Delta_i) \left\{Z_i(t) - \bar{Z}^{II}(t)\right\} dM_i(t) \\
& = n^{-1/2} \sum_{j=1}^{n} \int_0^{\tau} (1-\Delta_j)(\frac{\xi_j}{\tilde{\alpha}} - 1) \left(\frac{Y_j(t)}{E((1-\Delta_1)Y_1(t))}\right) \\
& \times \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\xi_i}{\tilde{\alpha}} (1-\Delta_i) \left\{Z_i(t) - e(t)\right\} Y_i(t) \left[d\mu_0(t) + \beta_0' Z_i(t) dt\right] \right\} + o_P(1).
\end{aligned}
$$

Similarly, we can show that

$$
\begin{aligned}
& n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \left(\hat{q}(t)^{-1} - \tilde{q}^{-1}\right) \eta_i \Delta_i (1-\xi_i) \left\{Z_i(t) - \bar{Z}^{II}(t)\right\} dM_i(t) \\
& = n^{-1/2} \sum_{j=1}^{n} \int_0^{\tau} \Delta_j (1-\xi_j)(\frac{\eta_j}{\tilde{q}} - 1) \left(\frac{Y_j(t)}{(1-\tilde{\alpha})E(\Delta_1 Y_1(t))}\right) \\
& \times \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\eta_i}{\tilde{q}} \Delta_i (1-\xi_i) \left\{Z_i(t) - e(t)\right\} dM_i(t) \right\} + o_P(1).
\end{aligned}
$$

From Lemma A1, we have $\frac{1}{n}\sum_{i=1}^{n} \frac{\eta_i}{\tilde{q}} \Delta_i (1-\xi_i) \left\{Z_i(t) - e(t)\right\} dM_i(t)$ converges to $E\left[\left\{Z_i(t) - e(t)\right\} dM_i(t) \mid \Delta_i = 1, \xi_i = 0\right] P(\Delta_i = 1, \xi_i = 0)$ in probability uniformly in t. Thus,

$$
\begin{aligned}
& n^{-1/2} \sum_{j=1}^{n} \int_0^{\tau} \Delta_j (1-\xi_j)(\frac{\eta_j}{\tilde{q}} - 1) \left(\frac{Y_j(t)}{(1-\tilde{\alpha})E(\Delta_1 Y_1(t))}\right) \times \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\eta_i}{\tilde{q}} \Delta_i (1-\xi_i) \left\{Z_i(t) - e(t)\right\} \right. \\
& \left. \times dM_i(t) \right\} \\
& = n^{-1/2} \sum_{j=1}^{n} \int_0^{\tau} \Delta_j (1-\xi_j)(\frac{\eta_j}{\tilde{q}} - 1) \left(\frac{Y_j(t)}{E(\Delta_1 Y_1(t))}\right) E\left[\left\{Z_i(t) - e(t)\right\} dM_i(t) \mid \Delta_i = 1, \xi_i = 0\right] \\
& \times P(\Delta_i = 1) + o_P(1) \\
& = n^{-1/2} \sum_{j=1}^{n} \int_0^{\tau} \Delta_j (1-\xi_j)(\frac{\eta_j}{\tilde{q}} - 1) \left(\frac{Y_j(t)}{E(Y_i(t) \mid \Delta_i = 1)}\right) E\left[\left\{Z_i(t) - e(t)\right\} dM_i(t) \mid \Delta_i = 1, \right. \\
& \left. \xi_i = 0\right] + o_P(1).
\end{aligned}
$$

Rewrite $dM_{\tilde{Z},i}(t) = (Z_i(t) - e(t))dM_i(t)$, the left hand side of (6.30) is asymptotically equivalent to

$$n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}(w_i^{II}(t)-1)\left\{Z_i(t)-\bar{Z}^{II}(t)\right\}dM_i(t)$$
$$= n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left(1-\frac{\xi_i}{\tilde{\alpha}}\right)(1-\Delta_i)\left\{Z_i(t)-e(t)\right\}Y_i(t)\left[d\mu_0(t)+\beta_0' Z_i(t)dt\right]$$
$$+ n^{-1/2}\sum_{j=1}^{n}\int_0^{\tau}(1-\Delta_j)(\frac{\xi_j}{\tilde{\alpha}}-1)\left(\frac{Y_j(t)}{E((1-\Delta_1)Y_1(t))}\right)\left\{E\left[(1-\Delta_i)\left\{Z_i(t)-e(t)\right\}Y_i(t)\right]d\mu_0(t)\right.$$
$$\left.+E\left[1-\Delta_i)\left\{Z_i(t)-e(t)\right\}Y_i(t)\beta_0' Z_i(t)dt\right]\right\} + n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\left(1-\frac{\eta_i}{\tilde{q}}\right)\Delta_i(1-\xi_i)\left\{Z_i(t)-e(t)\right\}$$
$$\times dM_i(t) + n^{-1/2}\sum_{j=1}^{n}\int_0^{\tau}\Delta_j(1-\xi_j)(\frac{\eta_j}{\tilde{q}}-1)\left(\frac{Y_j(t)}{E(Y_i(t)\mid\Delta_i=1)}\right)$$
$$\times E\left[\left\{Z_i(t)-e(t)\right\}dM_i(t)\mid\Delta_i=1,\xi_i=0\right]+o_P(1)$$
$$= n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\tilde{\alpha}}\right)(1-\Delta_i)\int_0^{\tau}\left\{\left\{Z_i(t)-e(t)\right\}\left[Y_i(t)d\mu_0(t)+Y_i(t)\beta_0' Z_i(t)dt\right]\right.$$
$$\left.-\frac{Y_i(t)E\left[(1-\Delta_i)\left\{Z_i(t)-e(t)\right\}(Y_i(t)d\mu_0(t)+Y_i(t)\beta_0' Z_i(t)dt)\right]}{E((1-\Delta_i)Y_i(t))}\right\}$$
$$+ n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\eta_i}{\tilde{q}}\right)\Delta_i(1-\xi_i)\left\{M_{\tilde{Z},i}(\tau)-\int_0^{\tau}Y_i(t)\frac{E\left[dM_{\tilde{Z},i}(t)\mid\Delta_i=1,\xi_i=0\right]}{E(Y_i(t)\mid\Delta_i=1)}\right\}+o_P(1).$$

Now, defining $dR_i(\beta,t) = Y_i(t)\left\{Z_i(t)-e(t)\right\}\left\{d\mu_0(t)+\beta' Z_i(t)dt\right\}$, we have the above expression as

$$n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\tilde{\alpha}}\right)(1-\Delta_i)\int_0^{\tau}\left\{dR_i(\beta_0,t)-\frac{Y_i(t)E\left((1-\Delta_i)dR_i(\beta,t)\right)}{E\left((1-\Delta_i)Y_i(t)\right)}\right\}$$

$$+n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\eta_i}{\tilde{q}}\right)\Delta_i(1-\xi_i)\left\{M_{\tilde{Z},i}(\tau)-\int_0^{\tau}Y_i(t)\frac{E\left[dM_{\tilde{Z},i}(t)\mid\Delta_i=1,\xi_i=0\right]}{E(Y_i(t)\mid\Delta_i=1)}\right\}.$$

Further, denote $dR_i(\beta_0,t)-\frac{Y_i(t)E((1-\Delta_i)dR_i(\beta_0,t))}{E((1-\Delta_i)Y_i(t))}$ by $dR_i^*(\beta,t)$ and $dM_{\tilde{Z},i}(t)-Y_i(t)\frac{E\left[dM_{\tilde{Z},i}(t)\mid\Delta_i=1,\xi_i=0\right]}{E(Y_i(t)\mid\Delta_i=1)}$

as $dR_i^{**}(t)$, $R_i^*(\beta) = \int_0^\tau dR_i^*(\beta, t)$ and $R_i^{**}(\tau) = \int_0^\tau dR_i^{**}(t)$. Hence, $n^{-1/2}U^{II}(\beta_0)$ is asymptotically equivalent to

$$n^{-1/2}\sum_{i=1}^{n} M_{\tilde{Z},i}(\beta_0)+n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\xi_i}{\tilde{\alpha}}\right)(1-\Delta_i)R_i^*(\beta_0)++n^{-1/2}\sum_{i=1}^{n}\left(1-\frac{\eta_i}{\tilde{q}}\right)\Delta_i(1-\xi_i)R_i^{**}(\tau). \tag{6.31}$$

The first term of the (6.31) is asymptotically normally distributed with mean zero and variance, $Q(\beta) = E\left(M_{\tilde{Z},1}(\beta)\right)^{\otimes 2}$ by Schaubel et al. (2006). To study the variance terms of the second and third terms of the (6.31), we use Hájek's (1960) central limit theorem for finite population sampling. Using (6.11) and (6.12), we can show that the following conditions for the Hájek's Theorem are satisfied.

(a) $\tilde{\alpha}$ converges to $\alpha \in (0, 1)$.

(b) $\frac{1}{n}\left|a'\left(\int_0^\tau dR_i^*(\beta,t) - \frac{1}{n}\sum_{i=1}^n \int_0^\tau dR_i^*(\beta,t)\right)\right|$ and $\frac{1}{n}\left|a'\left(\int_0^\tau dR_i^{**}(t) - \frac{1}{n}\sum_{i=1}^n \int_0^\tau dR_i^{**}(t)\right)\right|$ converges to 0 in probability.

(c) $\frac{1}{n-1}\sum_{i=1}^n \left(a'\int_0^\tau dR_i^*(\beta,t) - \frac{1}{n}\sum_{i=1}^n a'\int_0^\tau dR_i^*(\beta,t)\right)^2$ and $\frac{1}{n-1}\sum_{i=1}^n \left(a'\int_0^\tau dR_i^{**}(t)d\mu_0(t)\right.$ $\left.-\frac{1}{n}\sum_{i=1}^n a'\int_0^\tau dR_i^{**}(t)\right)^2$ converges to some quantity, $\sigma^{*2} > 0$ and $\sigma^{**2} > 0$ respectively.

Hence, based on Hájek's Theorem for finite population sampling, the second term of (6.31) can be shown to be asymptotically normally distributed with mean zero and covariance $\frac{1-\alpha}{\alpha}V^I(\beta)$ where

$$V^I(\beta) = var\left((1-\Delta_i)\int_0^\tau \left\{dR_i(\beta,t) - \frac{Y_i(t)E\left((1-\Delta_i)dR_i(\beta,t)\right)}{E\left((1-\Delta_i)Y_i(t)\right)}\right\}\right).$$

Further, based on the regularity condition and Cramer- Wold device, we have $n^{-1/2}\sum_{i=1}^n$ $\left(\{1-\frac{\eta_i}{\tilde{q}}\}\Delta_i(1-\xi_i)R_i^{**}(\tau)\right)$ converges to a zero-mean Normal random variable with variance $\frac{1-q}{q}(1-\alpha)P(\Delta_1 = 1)V^{II}(\beta)$, where

$$V^{II}(\beta) = var\left[M_{\tilde{Z},i}(\tau) - \int_0^\tau Y_i(t)\frac{E\left[dM_{\tilde{Z},i}(t) \mid \Delta_i = 1, \xi_i = 0\right]}{E(Y_i(t) \mid \Delta_i = 1)}\right].$$

Next let us consider the covariance between $n^{-1/2}\sum_{i=1}^n M_{\tilde{Z},i}(\beta)$, $n^{-1/2}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)R_i^*(\beta)$ and $n^{-1/2}\sum_{i=1}^n(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)R_i^{**}(\tau)$. We have already proved that $cov(n^{-1/2}\sum_{i=1}^n$

$M_{\tilde{Z},i}(\beta), n^{-1/2}\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)R_i^*(\tau)) = 0$ in Chapter 3. Defining the marginal filtration as $\mathcal{F}_i(t) = \sigma\{N_i(t), Y_i(t), Z_i(t)\}$ for the i-th individual at time t, we have

$$
\begin{aligned}
& cov(n^{-1/2}\sum_{i=1}^{n} M_{\tilde{Z},i}(\beta), n^{-1/2}\sum_{i=1}^{n}(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)R_i^{**}(\tau)) \\
&= E\left(n^{-1/2}\sum_{i=1}^{n} M_{\tilde{Z},i}(\beta) n^{-1/2}\sum_{i=1}^{n}(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)R_i^{**}(\tau)\right) \\
&= E\left[E\left(n^{-1/2}\sum_{i=1}^{n} M_{\tilde{Z},i}(\beta) n^{-1/2}\sum_{i=1}^{n}(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)R_i^{**}(\tau) \,\Big|\, \mathcal{F}(\tau)\right)\right] \\
&= E\left[n^{-1/2}\sum_{i=1}^{n} M_{\tilde{Z},i}(\beta) n^{-1/2}\sum_{i=1}^{n} E\left((1-\frac{\eta_i}{\tilde{q}})(1-\xi_i)\Delta_i \,\Big|\, \mathcal{F}(\tau)\right) R_i^{**}(\tau)\right] \\
&= E\left[n^{-1/2}\sum_{i=1}^{n} M_{\tilde{Z},i}(\beta) n^{-1/2}\sum_{i=1}^{n} \Big(P(\eta_i = 0, \xi_i = 0, \Delta_i = 1 \,\Big|\, \mathcal{F}(\tau)) + \right. \\
&\left. \frac{\tilde{q}-1}{\tilde{q}} P(\eta_i = 1, \xi_i = 0, \Delta_i = 1 \,\Big|\, \mathcal{F}(\tau))\Big) R_i^{**}(\tau)\right] \\
&= 0.
\end{aligned}
$$

The last step is a direct consequence of the fact that $(P(\eta_i = 0, \xi_i = 0, \Delta_i = 1 \,\Big|\, \mathcal{F}(\tau)) = P(\eta_i = 0 \,\Big|\, \xi_i = 0, \Delta_i = 1, \mathcal{F}(\tau)) P(\xi_i = 0, \Delta_i = 1 \,\Big|\, \mathcal{F}(\tau)) = 1 - \tilde{q} = 1 - P(\eta_i = 1 \,\Big|\, \xi_i = 0, \Delta_i = 1, \mathcal{F}(\tau)) P(\xi_i = 0, \Delta_i = 1 \,\Big|\, \mathcal{F}(\tau))$. Looking at the covariance between $n^{-1/2}\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)R_i^*(\beta)$ and $n^{-1/2}\sum_{i=1}^{n}(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)R_i^{**}(\tau)$, we have

$$
\begin{aligned}
& cov(n^{-1/2}\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)R_i^*(\beta), n^{-1/2}\sum_{i=1}^{n}(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)R_i^{**}(\tau)) \\
&= E\left(n^{-1/2}\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)R_i^*(\beta) n^{-1/2}\sum_{i=1}^{n}(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)R_i^{**}(\tau)\right) \\
&= E\left[E\left(n^{-1/2}\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)R_i^*(\beta) n^{-1/2}\sum_{i=1}^{n}(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)R_i^{**}(\tau) \,\Big|\, \mathcal{F}(\tau)\right)\right] \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i)R_i^*(\beta) E\left((1-\frac{\eta_i}{\tilde{q}})(1-\frac{\xi_i}{\tilde{\alpha}})(1-\xi_i) \,\Big|\, \mathcal{F}(\tau)\right) \Delta_i R_i^{**}(\tau)\right] \\
&+ \sum_{i \neq j} E\left[\frac{1}{n}(1-\Delta_i)R_i^*(\beta) E\left((1-\frac{\eta_j}{\tilde{q}})(1-\frac{\xi_i}{\tilde{\alpha}})(1-\xi_j) \,\Big|\, \mathcal{F}(\tau)\right) \Delta_j R_j^{**}(\tau)\right]
\end{aligned}
$$

Now, the first term is zero as $\Delta_i(1-\Delta_i) = 0 \forall i$ and the second term $E\Big((1-\frac{\eta_j}{\tilde{q}})(1-\frac{\xi_i}{\tilde{\alpha}})(1-\xi_j) \,\Big|$

$\mathcal{F}(\tau)) = E\left((1-\frac{\eta_j}{\tilde{q}})(1-\xi_j) \,\middle|\, \mathcal{F}_j(\tau)\right) E\left((1-\frac{\xi_i}{\tilde{\alpha}}) \,\middle|\, \mathcal{F}_i(\tau)\right)$ as the i-th and j-th subjects are independent and we can easily show that this term equals 0. Therefore, the variance of $n^{1/2}U_n(\beta)$ is given by $E(M_{\tilde{Z},i}(\beta))^{\otimes 2} + \frac{1-\alpha}{\alpha}V^I(\beta) + (1-\alpha)\frac{1-q}{q}P(\Delta_1 = 1)V^{II}(\beta)$. This implies that $U_n(\beta) \xrightarrow{P} 0$ as $n \to \infty$. Further, we can use Taylor Series expansion of $n^{1/2}U_n(\hat{\beta}^{II})$ around β_0 to get

$$n^{1/2}U_n(\hat{\beta}^{II}) \approx n^{1/2}U_n(\beta) + n^{1/2}(\hat{\beta}^{II} - \beta_0)' \frac{\delta}{\delta\beta}U_n(\beta^*),$$

where β^* belongs to the line segment joining β_0 and $\hat{\beta}^{II}$. Hence, we can say that $n^{1/2}(\hat{\beta}^{II} - \beta_0)$ converges to a normal distribution with mean 0 and variance

$$A^{-1}\left(E(M_{\tilde{Z},i}(\beta))^{\otimes 2} + \frac{1-\alpha}{\alpha}V^I(\beta) + (1-\alpha)\frac{1-q}{q}P(\Delta_1 = 1)V^{II}(\beta)\right)A^{-1}.$$

B.3 Proof of Theorem 4

: We can decompose $\hat{\mu}_0^{II}(\beta, t)$ in the following way:

$$
\begin{aligned}
& n^{1/2}\{\hat{\mu}_0^{II}(\hat{\beta}^{II}, t) - \mu_0(t)\} \\
& = n^{1/2}\{\hat{\mu}_0^{II}(\hat{\beta}^{II}, t) - \hat{\mu}_0^{II}(\beta_0, t) + \hat{\mu}_0^{II}(\beta_0, t) - \mu_0(t)\} \\
& = n^{1/2}\left\{\int_0^t \frac{\sum_{i=1}^n w_i^{II}(u)\left[dN_i(u) - Y_i(u)Z_i(u)'\hat{\beta}^{II}du\right]}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)}\right. \\
& - \int_0^t \frac{\sum_{i=1}^n w_i^{II}(u)\left[dN_i(u) - Y_i(u)\beta_0' Z_i(u)du\right]}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)} \\
& \left. + \int_0^t \frac{\sum_{i=1}^n w_i^{II}(u)\left[dN_i(u) - Y_i(u)\beta_0' Z_i(u)du\right]}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)} - \int_0^t \frac{\sum_{i=1}^n w_i^{II}(u)Y_i(u)d\mu_0(u)}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)}\right\} \\
& = n^{1/2}\left\{\int_0^t \frac{\sum_{i=1}^n w_i^{II}(u)\left[\beta_0 - \hat{\beta}^{II}\right]' Y_i(u)Z_i(u)du}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)}\right\} + n^{1/2}\left\{\int_0^t \frac{\sum_{i=1}^n w_i^{II}(u)dM_i(u)}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)}\right\} \\
& = n^{1/2}\int_0^t \frac{\sum_{i=1}^n w_i^{II}(u)\left[\beta_0 - \hat{\beta}^{II}\right]' Y_i(u)Z_i(u)du}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)} + n^{1/2}\int_0^t \frac{\sum_{i=1}^n dM_i(u)}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)} \\
& + n^{1/2}\int_0^t \frac{\sum_{i=1}^n \{w_i^{II}(u) - 1\}dM_i(u)}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)}. \qquad (6.32)
\end{aligned}
$$

By the uniform convergence of $\bar{Z}^{II}$ to $e(t)$, the first term on the right hand side of (6.32) is asymptotically equivalent to $n^{1/2}r(t)'\{\hat{\beta}^{II} - \beta_0\}$ where $r(t) = -\int_0^t e(u)du$. $(\frac{1}{n}\sum_{i=1}^n w_i^{II}(t)Y_i(t))^{-1}$ can be written as the sum of two monotone functions in t and converges uniformly to $E(Y_i(t))^{-1}$ where $E(Y_i(t))$ is bounded away from zero (Lin et al. 2000). $n^{-1/2}\sum_{i=1}^n M_i(t)$ converges to a zero-mean Gaussian process with continuous sample paths. Hence, by Lemma A2, $n^{1/2}\int_0^t \frac{\sum_{i=1}^n dM_i(u)}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)} = \int_0^t \frac{n^{-1/2}\sum_{i=1}^n dM_i(u)}{E(Y_i(u))} + o_P(1)$. The last term of (6.32) is

$$
\begin{aligned}
& n^{1/2}\int_0^t \frac{\sum_{i=1}^n \{w_i^{II}(u) - 1\}dM_i(u)}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)} \\
& = n^{1/2}\int_0^t \frac{\sum_{i=1}^n \{\Delta_i\xi_i + \frac{\xi_i(1-\Delta_i)}{\hat{\alpha}(t)} + \frac{\Delta_i(1-\xi_i)\eta_i}{\hat{q}(t)} - 1\}dM_i(u)}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)} \\
& = n^{1/2}\int_0^t \frac{\sum_{i=1}^n \left(\frac{\xi_i}{\tilde{\alpha}} - 1\right)(1-\Delta_i)dM_i(u)}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)} + n^{1/2}\int_0^t \frac{\sum_{i=1}^n \left(\frac{\eta_i}{\tilde{q}} - 1\right)\Delta_i(1-\xi_i)dM_i(u)}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)}
\end{aligned}
$$

$$
\begin{aligned}
&+ n^{1/2} \int_0^t \frac{\sum_{i=1}^n \left(\hat{\alpha}(u)^{-1} - \tilde{\alpha}^{-1}\right) \xi_i (1-\Delta_i) dM_i(u)}{\sum_{i=1}^n w_i^{II}(u) Y_i(u)} \\
&+ n^{1/2} \int_0^t \frac{\sum_{i=1}^n \left(\hat{q}(u)^{-1} - \tilde{q}^{-1}\right) \eta_i \Delta_i (1-\xi_i) dM_i(u)}{\sum_{i=1}^n w_i^{II}(u) Y_i(u)} \\
&= n^{-1/2} \int_0^t \frac{\sum_{i=1}^n \left(1 - \frac{\xi_i}{\tilde{\alpha}}\right) (1-\Delta_i) Y_i(u) \left[d\mu_0(u) + \beta_0' Z_i(u) du\right]}{E(Y_1(u))} \\
&+ n^{-1/2} \int_0^t \frac{\sum_{i=1}^n \left(\frac{\eta_i}{\tilde{q}} - 1\right) \Delta_i (1-\xi_i) dM_i(u)}{E(Y_1(u))} \\
&- n^{-1/2} \int_0^t \frac{\sum_{i=1}^n \left(\hat{\alpha}(u)^{-1} - \tilde{\alpha}^{-1}\right) \xi_i (1-\Delta_i) Y_i(u) \left[d\mu_0(u) + \beta_0' Z_i(u) du\right]}{E(Y_1(u))} \\
&+ n^{-1/2} \int_0^t \frac{\sum_{i=1}^n \left(\hat{q}(u)^{-1} - \tilde{q}^{-1}\right) \eta_i \Delta_i (1-\xi_i) dM_i(u)}{E(Y_1(u))} + o_P(1) \qquad (6.33)
\end{aligned}
$$

Using (6.20) and Lemma A1, we have the third term of (6.33) as:

$$
\begin{aligned}
&\int_0^t \frac{1}{E(Y_1(u))} n^{1/2} \left(\frac{1}{\hat{\alpha}(u)} - \frac{1}{\tilde{\alpha}}\right) \times n^{-1} \sum_{i=1}^n (1-\Delta_i) \xi_i Y_i(u) \left[d\mu_0 + \beta_0' Z_i(u) du\right] \\
&= \int_0^t \frac{1}{E(Y_1(u))} \frac{1}{E(1-\Delta_1) Y_1(u))} \left[n^{-1/2} \sum_{i=1}^n (1-\Delta_i)(1 - \frac{\xi_i}{\tilde{\alpha}}) Y_i(u)\right] \\
&\times n^{-1} \sum_{i=1}^n (1-\Delta_i) \xi_i Y_i(u) \left[d\mu_0(u) + \beta_0' Z_i(u) du\right] \\
&= n^{-1/2} \sum_{i=1}^n \int_0^t \frac{1}{E(Y_1(u))} (1-\Delta_i)(1 - \frac{\xi_i}{\tilde{\alpha}}) Y_i(u) \frac{E\left[Y_1(u)(1-\Delta_1)(d\mu_0(u) + \beta_0' Z_1(u) du)\right]}{E((1-\Delta_1) Y_1(u))} + o_P(1).
\end{aligned}
$$

(6.34)

Using (6.21) and Lemma A1, we have the fourth term of (6.33) as:

$$
\begin{aligned}
&\int_0^t \frac{1}{E(Y_1(u))} n^{1/2} \left(\frac{1}{\hat{q}(u)} - \frac{1}{\tilde{q}}\right) \times n^{-1} \sum_{i=1}^n \Delta_i \{1-\xi_i\} \eta_i dM_i(u) \\
&= \int_0^t \frac{1}{E(Y_1(u))} \frac{1}{E(\Delta_1(1-\xi_1)) Y_1(u))} \left[n^{-1/2} \sum_{i=1}^n \Delta_i (1-\xi_i)(1 - \frac{\eta_i}{\tilde{q}}) Y_i(u)\right] n^{-1} \sum_{i=1}^n \Delta_i \{1-\xi_i\} \\
&\times \eta_i dM_i(u) \\
&= n^{-1/2} \sum_{i=1}^n \int_0^t \frac{1}{E(Y_1(u))} \Delta_i (1-\xi_i)(1 - \frac{\eta_i}{\tilde{q}}) Y_i(u) \frac{E\left[dM_1(u) \mid \Delta_1 = 1, \xi_1 = 0\right]}{E(Y_1(u) \mid \Delta_1 = 1, \xi_1 = 0)} + o_P(1).
\end{aligned}
$$

(6.35)

Therefore,

$$
\begin{aligned}
& n^{1/2}\int_0^t \frac{\sum_{i=1}^n \{w_i^{II}(u)-1\}dM_i(u)}{\sum_{i=1}^n w_i^{II}(u)Y_i(u)} \\
& = n^{-1/2}\sum_{i=1}^n\left(1-\frac{\xi_i}{\tilde{\alpha}}\right)(1-\Delta_i)\int_0^t \frac{Y_i(u)}{E(Y_1(u))}\times\left\{\beta_0' Z_i(u)du-\frac{E\left[(1-\Delta_1)Y_1(u)\beta_0' Z_1(u)du\right]}{E((1-\Delta_1)Y_1(u))}\right\} \\
& + n^{-1/2}\sum_{i=1}^n\left(1-\frac{\eta_i}{\tilde{q}}\right)\Delta_i(1-\xi_i)\int_0^\tau \frac{1}{E(Y_1(u))}\times\left\{dM_i(u)-Y_i(u)\frac{E\left[dM_1(u)\mid\Delta_1=1,\xi_1=0\right]}{E(Y_1(u)\mid\Delta_1=1)}\right\} \\
& + o_P(1). \qquad (6.36)
\end{aligned}
$$

Hence, using (6.36), (6.32) can be written as

$$
\begin{aligned}
& n^{1/2}\{\hat{\mu}_0^{II}(\hat{\beta}^{II},t)-\mu_0(t)\} \\
& = n^{1/2}r(t)'\left\{\hat{\beta}^{II}-\beta_0\right\}+n^{-1/2}\left\{\int_0^t \frac{\sum_{i=1}^n dM_i(u)}{E(Y_1(u))}\right\} \\
& + n^{-1/2}\sum_{i=1}^n\left(1-\frac{\xi_i}{\tilde{\alpha}}\right)(1-\Delta_i)\int_0^t \frac{Y_i(u)}{E(Y_1(u))}\times\left\{\beta_0' Z_i(u)du-\frac{E\left[(1-\Delta_1)Y_1(u)\beta_0' Z_1(u)du\right]}{E((1-\Delta_1)Y_1(u))}\right\} \\
& + n^{-1/2}\sum_{i=1}^n\left(1-\frac{\eta_i}{\tilde{q}}\right)\Delta_i(1-\xi_i)\int_0^\tau \frac{1}{E(Y_1(u))}\times\left\{dM_i(u)-Y_i(u)\frac{E\left[dM_1(u)\mid\Delta_1=1,\xi_1=0\right]}{E(Y_1(u)\mid\Delta_1=1)}\right\} \\
& + o_P(1). \qquad (6.37)
\end{aligned}
$$

By Taylor Series expansion of $U_n(\hat{\beta}^{II})$ around β_0, we have

$$
\begin{aligned}
n^{1/2}\{\hat{\beta}^{II}-\beta_0\} &= A^{-1}\times n^{-1/2}\{U_n(\beta_0)\}+o_P(1) \\
&= A^{-1}\left[n^{-1/2}\sum_{i=1}^n M_{\tilde{Z},i}(\beta_0)+n^{-1/2}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)R_i^*(\beta_0)\right. \\
&\left. + n^{-1/2}\sum_{i=1}^n(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)R_i^{**}(\tau)\right]+o_p(1).
\end{aligned}
$$

Plugging the above in (6.37), we have

$$
\begin{aligned}
& n^{1/2}\{\hat{\mu}_0^{II}(\hat{\beta}^{II},t)-\mu_0(t)\} \\
& = n^{-1/2}\sum_{i=1}^n\left[r(t)'A^{-1}M_{\tilde{Z},i}(\beta_0)+\int_0^t \frac{dM_i(u)}{E(Y_1(u))}\right]+n^{-1/2}\sum_{i=1}^n(1-\frac{\xi_i}{\tilde{\alpha}})(1-\Delta_i)\left(r(t)'A^{-1}R_i^*(\beta_0)\right.
\end{aligned}
$$

$$+\int_0^t \frac{1}{E\left(Y_1(u)\right)} Y_i(u) \times \left[\beta_0' Z_i(u) - \frac{E\left(\{1-\Delta_1\}Y_1(u)\beta_0' Z_1(u)\right)}{E((1-\Delta_1)Y_1(u))}\right] du\right) + n^{-1/2}\sum_{i=1}^{n}(1-\frac{\eta_i}{\tilde{q}})\Delta_i$$

$$(1-\xi_i)\left(r(t)'A^{-1}R_i^{**}(\tau) + \int_0^t \frac{1}{E\left(Y_1(u)\right)}\left\{dM_i(u) - Y_i(u)\frac{E\left[dM_1(u) \mid \Delta_1 = 1, \xi_1 = 0\right]}{E(Y_1(u) \mid \Delta_1 = 1)}\right\}\right) + o_P(1)$$

$$= n^{-1/2}\sum_{i=1}^{n}\nu_i(\beta_0,t) + n^{-1/2}\sum_{i=1}^{n}(1-\frac{\xi_i}{\tilde{\alpha}})\psi^I(\beta_0,t) + n^{-1/2}\sum_{i=1}^{n}(1-\frac{\eta_i}{\tilde{q}})\psi^{II}(\beta_0,t) + o_P(1). \tag{6.38}$$

$$\begin{aligned}\nu_i(\beta_0,t) &= r(t)'A^{-1}M_{\tilde{Z},i}(\beta_0) + \int_0^t \frac{dM_i(u)}{E(Y_1(u))} \\ &= r(t)'A^{-1}\int_0^\tau \left(Z_i(t) - e(t)\right) dM_i(t) + \int_0^t \frac{dM_i(u)}{E(Y_1(u))}.\end{aligned}$$

$$\begin{aligned}&\psi_i^I(\beta_0,t) \\ &= (1-\Delta_i)\left[r(t)'A^{-1}R_i^*(\beta_0) + \int_0^t \frac{1}{E\left(Y_1(u)\right)}Y_i(u)\left[\beta_0' Z_i(u) - \frac{E\left(\{1-\Delta_1\}Y_1(u)\beta_0' Z_1(u)\right)}{E((1-\Delta_1)Y_1(u))}\right] du\right] \\ &= (1-\Delta_i)\left[r(t)'A^{-1}\left(\int_0^\tau dR_i(\beta_0,t) - \int_0^\tau \frac{Y_i(t)E((1-\Delta_1)dR_i(\beta_0,t))}{E((1-\Delta_1)Y_i(t))}\right)\right. \\ &+ \left.\int_0^t \frac{1}{E\left(Y_1(u)\right)}Y_i(u)\left[\beta_0' Z_i(u) - \frac{E\left(\{1-\Delta_1\}Y_1(u)\beta_0' Z_1(u)\right)}{E((1-\Delta_1)Y_1(u))}\right] du\right]\end{aligned}$$

and

$$\begin{aligned}\psi_i^{II}(\beta_0,t) = \Delta_i(1-\xi_i)r(t)'A(\beta_0)^{-1}\int_0^\tau &\left\{dM_{\tilde{Z},i}(u) - Y_i(u)\frac{E\left[dM_{\tilde{Z},1}(u) \mid \Delta_1 = 1, \xi_1 = 0\right]}{E(Y_1(u) \mid \Delta_1 = 1)}\right\} \\ + \Delta_i(1-\xi_i)\int_0^t \frac{1}{E\left(Y_1(u)\right)}&\left\{dM_i(u) - Y_i(u)\frac{E\left[dM_1(u) \mid \Delta_1 = 1, \xi_1 = 0\right]}{E(Y_1(u) \mid \Delta_1 = 1)}\right\}\end{aligned}$$

Let $W_1(t) = n^{-1/2}\sum_{i=1}^n \nu_i(\beta_0,t)$, $W_2(t) = n^{-1/2}\sum_{i=1}^n (1-\frac{\xi_i}{\tilde{\alpha}})\psi^I(\beta_0,t)$ and $W_3(t) = n^{-1/2}\sum_{i=1}^n$ $(1-\frac{\eta_i}{\tilde{q}})\psi^{II}(\beta_0,t)$. Since, $r(t)$, $E\left(\{1-\Delta_1\}Y_1(u)\beta_0' Z_1(u)\right)$, $E\left[dM_{\tilde{Z},1}(u) \mid \Delta_1 = 1, \xi_1 = 0\right]$ and $E\left[dM_1(u) \mid \Delta_1 = 1, \xi_1 = 0\right]$ are of bounded variations, $E(Y_1(t))$ is bounded away from zero and A is positive definite, along with the asymptotic normality of the rest of the terms, we can say that $\nu_i(\beta_0,t)$, $\psi_i^I(\beta_0,t)$ and $\psi_i^{II}(\beta_0,t)$ are bounded processes. Hence, by multivariate CLT,

$W_1(t)$, $W_2(t)$ and $W_3(t)$ would converge to a zero-mean Guassian processes, which has continuous sample paths under the Euclidean distance(Lin et al. 2000). The covariance function for the three Gaussian processes are given by $E\left(\nu_i^{II}(\beta_0,t)\nu_i^{II}(\beta_0,s)\right)+\frac{1-\alpha}{\alpha}E\left(\psi_i^{I}(\beta_0,t)\psi_i^{I}(\beta_0,s)\right)$ $\frac{1-q}{q}E\left(\Delta_i(1-\xi_i)\psi_i^{II}(\beta_0,t)\psi_i^{II}(\beta_0,s)\right)$. One can note that, $W_n(t) = W_1(t)+W_2(t)+W_3(t)$ converges to a zero-mean Gaussian process, $\mathcal{W}(t)$ with continuous sample paths, as it is the sum of independent and identically distributed zero-mean terms for fixed t(Lin et al. 2000). The covariance between $W_1(t)$ and $W_3(t) = 0$ since

$$
\begin{aligned}
& cov(n^{-1/2}\sum_{i=1}^{n}\nu_i(\beta_0,t), n^{-1/2}\sum_{i=1}^{n}(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)\psi_i^{II}(\beta_0,t)) \\
&= E\left(n^{-1/2}\sum_{i=1}^{n}\nu_i(\beta_0,t)n^{-1/2}\sum_{i=1}^{n}(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)\psi_i^{II}(\beta_0,t)\right) \\
&= E\left[E\left(n^{-1/2}\sum_{i=1}^{n}\nu_i(\beta_0,t)n^{-1/2}\sum_{i=1}^{n}(1-\frac{\eta_i}{\tilde{q}})\Delta_i(1-\xi_i)\psi_i^{II}(\beta_0,t)\;\middle|\;\mathcal{F}(\tau)\right)\right] \\
&= E\left[n^{-1/2}\sum_{i=1}^{n}\nu_i(\beta_0,t)n^{-1/2}\sum_{i=1}^{n}E\left(1-\frac{\eta_i}{\tilde{q}}(1-\xi_i)\;\middle|\;\mathcal{F}(\tau)\right)\Delta_i\psi_i^{II}(\beta_0,t)\right] = 0.
\end{aligned}
$$

Similarly, the covariance between $W_1(t)$ and $W_2(t)$ and $W_2(t)$ and $W_3(t)$ are equal to zero $\forall t$. Hence, $W_n(t)$ converges weakly to a zero-mean Gaussian process with the covariance function between $\mathcal{W}(t)$ and $\mathcal{W}(s)$ is given by $E\left(\nu_i(\beta_0,t)\nu_i(\beta_0,s)\right)+\frac{1-\alpha}{\alpha}E\left((1-\Delta_i)\psi_i^{I}(\beta_0,t))\psi_i^{I}(\beta_0,s)\right)$ $+\ \frac{1-q}{q}E\left(\Delta_i(1-\xi_i)\psi_i^{II}(\beta_0,t))\psi_i^{II}(\beta_0,s)\right)$.

APPENDIX C: TECHNICAL DETAILS FOR CHAPTER 5

Here is the Appendix containing the simplification of the power formula in Chapter 5.

C.1 Asymptotic Distribution of Test Statistic

Self and Prentice (1988) proposed a score function whose asymptotic distribution is given by

$$n_j^{-1/2}\tilde{U}_j(\beta_j^0) \underset{D}{\longrightarrow} N(0, \sigma_j^2(\beta_j^0) + \delta_j(\beta_j^0)) \quad \forall j = 0, 1$$

By Taylor series expansion, $n^{1/2}\left(\hat{\beta}_j - \beta_j\right) \underset{D}{\longrightarrow} N(0, r_j^{-1}\left(\sigma_j^{-2}(\beta_j) + \sigma_j^{-4}(\beta_j)\delta_j(\beta_j)\right)), \quad j = 0, 1$ and $\hat{\beta}_j$ are independent. Hence, we have under $H_0 : \beta_0 = \beta_1$, the asymptotic distribution of $n^{1/2}\left(\hat{\beta}_1 - \hat{\beta}_0\right)$ follows:

$$n^{1/2}\left(\hat{\beta}_1 - \hat{\beta}_0\right) \underset{D}{\longrightarrow} N\left(0, \sum_{j=0}^{1} r_j^{-1}\left(\sigma_j^{-2}(\beta_j) + \sigma_j^{-4}(\beta_j)\delta_j(\beta_j)\right)\right)$$

C.2 Consistent Estimator Of The Variance Components

To obtain consistent estimators of different variance components, note that

$$\frac{\sum_{i=1}^{n_j} Y_{ijk}(x)}{\tilde{n}_j} \underset{P}{\longrightarrow} (1-p_j)^{1-k} p_j^k \pi_{jk}(x) S_j(x)^{e^{k\beta_j}} \quad \forall j, k = 0, 1$$

$$\tilde{Q^{(l)}}_j(\beta_j, x, w) = \frac{1}{\tilde{n}_j} \sum_{i \in \tilde{C}_j} Y_i(x \vee w) Z_i^l e^{2Z_i\beta_j}, \quad \forall l = 0, 1, j = 0, 1$$

$$S^{(l)}(\beta_j, x) = \frac{1}{\tilde{n}_j} \sum_{i \in \tilde{C}_j} Y_i(x) Z_i^l e^{Z_i\beta_j}, \quad \forall l = 0, 1, j = 0, 1$$

$$\frac{\tilde{n}_j}{n_j} \to \psi_j, \tilde{Q^{(l)}}_j(\beta_j, x, w) \xrightarrow{P} q^{(l)}(\beta_j, x, w), S^{(l)}(\beta_j, x) \xrightarrow{P} s^{(l)}(\beta_j, x) \quad \forall l = 0, 1, j = 0, 1$$

Note that the sub-cohort averages converges to the full cohort averages as $n \to \infty$. Hence, the variance terms can be written as

$$\sigma_j^2 = \int_0^1 \left(\frac{s^{(2)}(\beta_j, t)}{s^{(0)}(\beta_j, t)} - (\frac{s^{(1)}(\beta_j, t)}{s^{(0)}(\beta_j, t)})^{\otimes 2} \right) s^{(0)}(\beta_j, t)\lambda_j(t)dt$$
$$= \int_0^1 \left(\frac{s^{(2)}(\beta_j, t)s^{(0)}(\beta_j, t) - s^{(1)}(\beta_j, t)^{\otimes 2}}{[s^{(0)}(\beta_j, t)]} \right) \lambda_j(t)dt$$

$$\text{and} \quad \delta_j = \int_0^1 \int_0^1 G(\beta_j, x, w)s^{(0)}(\beta_j, x)s^{(0)}(\beta_j, w)\lambda_j(x)\lambda_j(w)dxdw,$$

Since,

$$h^{(k+l)}(\beta_j, x, w) = q^{(k+l)}(\beta_j, x, w) - s^{(k)}(\beta_j, w)s^{(l)}(\beta_j, x)' \quad \forall l \leq k = 0, 1$$

we can rewrite $G(\beta_j, x, w)$ as

$$\begin{aligned} G(\beta_j, x, w) = & \frac{1-\psi_j}{\psi_j}\{s^{(0)}(\beta_j, x)s^{(0)}(\beta_j, w)\}^{-2} \Big[\Big(q^{(2)}(\beta_j, x, w) - s^{(1)}(\beta_j, x)s^{(1)}(\beta_j, w) \Big) \\ & \times s^{(0)}(\beta_j, x)s^{(0)}(\beta_j, w) + \Big(q^{(0)}(\beta_j, x, w) - s^{(0)}(\beta_j, x)s^{(0)}(\beta_j, w) \Big) s^{(1)}(\beta_j, x)s^{(1)}(\beta_j, w) \\ & - s^{(0)}(\beta_j, x)s^{(1)}(\beta_j, w) \Big(q^{(1)}(\beta_j, x, w) - s^{(0)}(\beta_j, w)s^{(1)}(\beta_j, x) \Big) - s^{(0)}(\beta_j, w)s^{(1)}(\beta_j, x) \\ & \times \Big(q^{(1)}(\beta_j, x, w) - s^{(0)}(\beta_j, x)s^{(1)}(\beta_j, w) \Big) \Big] \\ = & \frac{1-\psi_j}{\psi_j}\{s^{(0)}(\beta_j, x)s^{(0)}(\beta_j, w)\}^{-2} \Big[q^{(2)}(\beta_j, x, w)s^{(0)}(\beta_j, x)s^{(0)}(\beta_j, w) + q^{(0)}(\beta_j, x, w) \\ & s^{(1)}(\beta_j, x)s^{(1)}(\beta_j, w) - \Big(s^{(0)}(\beta_j, x)s^{(1)}(\beta_j, w) + s^{(0)}(\beta_j, w)s^{(1)}(\beta_j, x) \Big) q^{(1)}(\beta_j, x, w) \Big] \end{aligned}$$

Hence, $G(\beta_j, x, w)$ can be further simplified as

$$\begin{aligned} & G(\beta_j, x, w) \\ & = \frac{1-\psi_j}{\psi_j}\{s^{(0)}(\beta_j, x)s^{(0)}(\beta_j, w)\}^{-2} \Big[q^{(1)}(\beta_j, x, w) \Big(s^{(0)}(\beta_j, x)s^{(0)}(\beta_j, w) - s^{(0)}(\beta_j, x)s^{(1)}(\beta_j, w) \\ & - s^{(0)}(\beta_j, w)s^{(1)}(\beta_j, x) \Big) + q^{(0)}(\beta_j, x, w)s^{(1)}(\beta_j, x)s^{(1)}(\beta_j, w) \Big] \end{aligned}$$

Now, for the variance term of the statistic, we examined each of the variances of the score functions separately. $G(\beta_j, x, w)$ can be simplified as

$$
\begin{aligned}
&\frac{\psi_j}{1-\psi_j} G(\beta_j, x, w) \\
&= \{s^{(0)}(\beta_j, x) s^{(0)}(\beta_j, w)\}^{-2} \Big[q^{(1)}(\beta_j, x, w)(1-p_j)^2 \pi_{j0}(x)\pi_{j0}(w) S_j(x) S_j(w) \\
&+ (1-p_j)\pi_{j0}(x \vee w) S_j(x \vee w) s^{(1)}(\beta_j, x) s^{(1)}(\beta_j, w) \Big] \\
&= \{s^{(0)}(\beta_j, x) s^{(0)}(\beta_j, w)\}^{-2} \Big[e^{2\beta_j} p_j \pi_{j1}(x \vee w) S_j(x \vee w)^{e^{\beta_j}} \times (1-p_j)^2 \pi_{j0}(x)\pi_{j0}(w) S_j(x) S_j(w) \\
&+ (1-p_j)\pi_{j0}(x \vee w) S_j(x \vee w) \times e^{2\beta_j} p_j^2 \pi_{j1}(x)\pi_{j1}(w) \left(S_j(x) S_j(w)\right)^{e^{\beta_j}} \Big] \\
&= \frac{e^{2\beta_j}(1-p_j)p_j \sum_{k=0}^{1} (1-p_j)^{1-k} p_j^k \pi_{j,(1-k)}(x \vee w) \pi_{j,k}(x) \pi_{j,k}(w) S_j(x \vee w)^{e^{(1-k)\beta_j}}}{\left(\sum_{k=0}^{1} e^{k\beta_j}(1-p_j)^{1-k} p_j^k \pi_{jk}(x) S_j(x)^{e^{k\beta_j}}\right)^2 \left(\sum_{k=0}^{1} e^{k\beta_j}(1-p_j)^{1-k} p_j^k \pi_{jk}(w) S_j(w)^{e^{k\beta_j}}\right)^2} \\
&\times S_j(x)^{e^{k\beta_j}} S_j(w)^{e^{k\beta_j}}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\delta_j = \frac{1-\psi_j}{\psi_j} \int_0^1 \int_0^1 & \frac{\sum_{k=0}^{1} (1-p_j)^{1-k} p_j^k \pi_{j,(1-k)}(x \vee w) \pi_{j,k}(x) \pi_{j,k}(w) S_j(x \vee w)^{e^{(1-k)\beta_j}}}{\left(\sum_{k=0}^{1} e^{k\beta_j}(1-p_j)^{1-k} p_j^k \pi_{jk}(x) S_j(x)^{e^{k\beta_j}}\right)} \\
\times S_j(x)^{e^{k\beta_j}} S_j(w)^{e^{k\beta_j}} & \frac{e^{2\beta_j}(1-p_j)p_j d\Lambda_j(x) d\Lambda_j(w)}{\left(\sum_{k=0}^{1} e^{k\beta_j}(1-p_j)^{1-k} p_j^k \pi_{jk}(w) S_j(w)^{e^{k\beta_j}}\right)}
\end{aligned}
$$

and

$$
\begin{aligned}
\sigma_j^2 &= \int_0^1 \left(\frac{s^{(1)}(\beta_j, t)\left(s^{(0)}(\beta_j, t) - s^{(1)}(\beta_j, t)\right)}{[s^{(0)}(\beta_j, t)]} \right) \lambda_j(t) dt \\
&= \int_0^1 \left(\frac{s^{(1)}(\beta_j, t)\left((1-p_j)\pi_{j0}(t) S_j(t)\right)}{[s^{(0)}(\beta_j, t)]} \right) \lambda_j(t) dt \\
&= \int_0^1 \frac{e^{\beta_j}(1-p_j)p_j \pi_{j0}(t)\pi_{j1}(t) S_j(t)^{e^{\beta_j}}}{\sum_{k=0}^{1} e^{k\beta_j}(1-p_j)^{1-k} p_j^k \pi_{jk}(t) S_j(t)^{e^{k\beta_j}}} S_j(t) d\Lambda_j(t)
\end{aligned}
$$

Under $H_0 : \beta_0 = \beta_1$, and $\tilde{\Lambda}_j(t) = \int_0^t \frac{d\bar{N}_{j0}(u)+\bar{N}_{j1}(u)}{\tilde{Y}_{j0}(u)+e^{\beta_j}\tilde{Y}_{j1}(u)} \approx \Lambda_j(t)/\psi_j$, the consistent estimates of the variance components, δ_j and σ_j^2 for X = j are :

$$\hat{\delta}_j = \frac{\hat{\psi}_j(1-\hat{\psi}_j)}{\tilde{n}_j} \int_0^1 \int_0^1 \frac{e^{2\hat{\beta}_j}\Big(\sum_{k=0}^1 \tilde{Y}_{j,1-k}(x \vee w)\tilde{Y}_{jk}(x)\tilde{Y}_{jk}(w)\Big) d\tilde{\Lambda}_j(x) d\tilde{\Lambda}_j(w)}{(\sum_{k=0}^1 e^{k\hat{\beta}_j}\tilde{Y}_{j,k}(x))(\sum_k e^{k\hat{\beta}_j}\tilde{Y}_{jk}(w))}.$$

and $d\tilde{\Lambda}_j(x) = \frac{\sum_k d\bar{N}_{jk}(x)}{\sum_k e^{k\hat{\beta}_j}\tilde{Y}_{jk}(x)}$ imply

$$\hat{\delta}_j = \frac{\hat{\psi}_j(1-\hat{\psi}_j)}{\tilde{n}_j} \int_0^1 \int_0^1 \frac{e^{2\hat{\beta}_j}\Big(\sum_{k=0}^1 \tilde{Y}_{j,1-k}(x \vee w)\tilde{Y}_{jk}(x)\tilde{Y}_{jk}(w)\Big)\Big(\sum_{k=0}^1 d\bar{N}_{jk}(x)\Big)}{(\sum_{k=0}^1 e^{k\hat{\beta}_j}\tilde{Y}_{jk}(x))^2(\sum_{k=0}^1 e^{k\hat{\beta}_j}\tilde{Y}_{jk}(w))^2} \times \left(\sum_{k=0}^1 d\bar{N}_{jk}(w)\right)$$

Also, we have the estimate of σ_j^2 as

$$\begin{aligned} \hat{\sigma_j^2} &= \frac{\psi_j}{\tilde{n}_j} \int_0^1 \frac{e^{\hat{\beta}_j}\tilde{Y}_{j0}(x)\tilde{Y}_{j1}(x)\left(\sum_k d\bar{N}_{jk}(x)\right)}{\Big(\sum_{k=0}^1 e^{k\hat{\beta}_j}\tilde{Y}_{j,k}(x)\Big)^2} \\ &\approx \frac{1}{n_j} \int_0^1 \frac{\tilde{Y}_{j0}(x)^2 d\bar{N}_{j1}(x) + e^{2\hat{\beta}_j}\tilde{Y}_{j1}(x)^2 d\bar{N}_{j0}(x)}{\Big(\sum_{k=0}^1 e^{k\hat{\beta}_j}\tilde{Y}_{jk}(x)\Big)^2} \end{aligned} \tag{6.39}$$

C.3 Power/Sample Size for Rare Event

Assume that the distribution of the censoring distribution C_i does not depend on Z_i and X_i, we have

$$\pi_{jk}(x) = \pi_j(x) \quad \forall j, k = 0, 1$$

and all the results hold as earlier.

We can rewrite $\hat{\delta}_j$ as

$$\hat{\delta}_j = \frac{\hat{\psi}_j(1-\hat{\psi}_j)}{\tilde{n}_j} \sum_{k,k'=0}^{1} \sum_{i=1}^{n_k} \sum_{i'=1}^{n_{k'}} \Big[\frac{e^{2\hat{\beta}_j}\Delta_{ijk}\Delta_{i'jk'}}{\Big(\tilde{Y}_{j0}(T_{ijk}) + e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk})\Big)^2\Big(\tilde{Y}_{j0}(T_{ijk'}) + e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk'})\Big)^2} \times \Big(\tilde{Y}_{j0}(T_{ijk})\tilde{Y}_{j0}(T_{ijk'})\tilde{Y}_{j1}(T_{ijk} \vee T_{ijk'}) + \tilde{Y}_{j1}(T_{ijk})\tilde{Y}_{j1}(T_{ijk'})\tilde{Y}_{j0}(T_{ijk} \vee T_{ijk'})\Big)\Big]$$

$$
\begin{aligned}
&= 2\frac{e^{2\hat{\beta}_j}\hat{\psi}_j(1-\hat{\psi}_j)}{\tilde{n}_j}\sum_{k,k'=0}^{1}\sum_{i=1}^{n_k}\sum_{i'=1}^{n_{k'}}\left[\frac{\Delta_{ijk}\Delta_{i'jk'}\mathbf{1}(T_{ijk'}\le T_{ijk})\tilde{Y}_{j1}(T_{ijk})\tilde{Y}_{j0}(T_{ijk})}{\left(\tilde{Y}_{j0}(T_{ijk})+e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk})\right)^2\left(\tilde{Y}_{j0}(T_{ijk'})+e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk'})\right)^2}\right.\\
&\left.\times\left(\tilde{Y}_{j1}(T_{ijk'})+\tilde{Y}_{j0}(T_{ijk'})\right)\right]-\frac{e^{2\hat{\beta}_j}\hat{\psi}_j(1-\hat{\psi}_j)}{\tilde{n}_j}\sum_{k=0}^{1}\sum_{i=1}^{n_k}\frac{\Delta_{ijk}\tilde{Y}_{j1}(T_{ijk})\tilde{Y}_{j0}(T_{ijk})}{\left(\tilde{Y}_{j0}(T_{ijk})+e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk})\right)^4}\\
&\times\left(\tilde{Y}_{j1}(T_{ijk})+\tilde{Y}_{j0}(T_{ijk})\right)\\
&= 2\frac{e^{2\hat{\beta}_j}(1-\hat{\psi}_j)}{n_j}\sum_{k,k'=0}^{1}\sum_{i,\Delta_{ijk}=1}\sum_{i,\Delta_{i'jk'}=1}\left[\frac{\mathbf{1}(T_{ijk'}\le T_{ijk})\tilde{Y}_{j1}(T_{ijk})\tilde{Y}_{j0}(T_{ijk})}{\left(\tilde{Y}_{j0}(T_{ijk})+e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk})\right)^2}\right.\\
&\left.\times\frac{\left(\tilde{Y}_{j1}(T_{ijk'})+\tilde{Y}_{j0}(T_{ijk'})\right)}{\left(\tilde{Y}_{j0}(T_{ijk'})+e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk'})\right)^2}\right]-\frac{e^{2\hat{\beta}_j}(1-\hat{\psi}_j)}{n_j}\sum_{k=0}^{1}\sum_{i,\Delta_{ijk}=1}\frac{\tilde{Y}_{j1}(T_{ijk})\tilde{Y}_{j0}(T_{ijk})}{\left(\tilde{Y}_{j0}(T_{ijk})+e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk})\right)^4}\\
&\times\left(\tilde{Y}_{j1}(T_{ijk})+\tilde{Y}_{j0}(T_{ijk})\right) \qquad (6.40)
\end{aligned}
$$

In order to derive the formula which can be used in the design stage of the study, we need to simplify (6.40) so that it only involves quantities that can be assumed in the design. We order the failure times from smallest to largest for each of the two treatment groups for biomarker j in the full cohort and the corresponding risk sets are $m_{1jk}, m_{2jk}, \ldots, j,k = 1,2$. The size of the risk set in the sub-cohort is about q times that of the full cohort. In the case of rare events, $Y_{ijk}(T_{ijk}) = Y_{ijk}(\tau) \quad \forall i,j,k, \quad \tau =$ last observed time. In other words, $m_{ljk} = m_{*jk} \quad \forall l$, m_{*jk} is the number of individuals at risk at the last observed failure time and $\frac{m_{lj0}}{m_{lj1}} \approx \frac{(1-p_j)(1-p_D^{j0})}{p_j(1-p_D^{j1})} \quad \forall j = 0,1$, d_{lj} is the number of failure times $\ge$ l-th failure time; D_k^j is the number of failure for the k-th treatment in the j-th biomarker group. Hence, $\hat{\delta}_j$ can be rewritten as

$$
\begin{aligned}
\hat{\delta}_j &\approx 2\frac{e^{2\hat{\beta}_j}(1-\psi_j)}{n_j}\sum_l\frac{\psi_j^2 m_{lj0}m_{lj1}}{\psi_j^2\left(m_{lj0}+e^{\hat{\beta}_j}m_{lj1}\right)^2}\sum_{l'\le l}\frac{\psi_j\left(m_{l',j,0}+m_{l',j,1}\right)}{\psi_j^2\left(m_{l',j,0}+e^{\hat{\beta}_j}m_{l',j,1}\right)^2}\\
&-\frac{e^{2\hat{\beta}_j}(1-\psi_j)}{n_j}\sum_l\frac{\psi_j^2 m_{lj0}m_{lj1}\left(m_{lj0}+m_{lj1}\right)\psi_j}{\psi_j^4\left(m_{lj0}+e^{\hat{\beta}_j}m_{lj1}\right)^4}\\
&= 2\frac{e^{2\hat{\beta}_j}(1-\psi_j)}{n_j\psi_j}\sum_l\frac{1}{\left(1+e^{\hat{\beta}_j}\frac{m_{lj1}}{m_{lj0}}\right)\left(e^{\hat{\beta}_j}+\frac{m_{lj0}}{m_{lj1}}\right)}\sum_{l'\le l}\frac{\left(1+\frac{m_{l',j,1}}{m_{l',j,0}}\right)}{\left(1+e^{\hat{\beta}_j}\frac{m_{l',j,1}}{m_{l',j,0}}\right)\left(m_{l',j,0}+e^{\hat{\beta}_j}m_{l',j,1}\right)}
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{e^{2\hat{\beta}_j}(1-\psi_j)}{\psi_j n_j} \sum_l \frac{\left(1+\frac{m_{lj1}}{m_{lj0}}\right)}{\left(1+e^{\hat{\beta}_j}\frac{m_{lj1}}{m_{lj0}}\right)^2 \left(e^{\hat{\beta}_j}+\frac{m_{lj0}}{m_{lj1}}\right)\left(m_{lj0}+e^{\hat{\beta}_j}m_{lj1}\right)} \\
&\approx 2\frac{e^{2\hat{\beta}_j}(1-\psi_j)}{n_j\psi_j} \sum_l \frac{1}{\left(1+e^{\hat{\beta}_j}\frac{p_j(1-p_D^{j1})}{(1-p_j)(1-p_D^{j0})}\right)\left(e^{\hat{\beta}_j}+\frac{(1-p_j)(1-p_D^{j0})}{p_j(1-p_D^{j1})}\right)} \\
&\times \sum_{l'\le l} \frac{\left(1+\frac{p_j(1-p_D^{j1})}{(1-p_j)(1-p_D^{j0})}\right)}{\left(1+e^{\hat{\beta}_j}\frac{p_j(1-p_D^{j1})}{(1-p_j)(1-p_D^{j0})}\right)} \frac{1}{\left(m_{l',j,0}+e^{\hat{\beta}_j}m_{l',j,1}\right)} - \frac{e^{2\hat{\beta}_j}(1-\psi_j)}{\psi_j n_j} \sum_l \frac{\left(1+\frac{p_j(1-p_D^{j1})}{(1-p_j)(1-p_D^{j0})}\right)}{\left(1+e^{\hat{\beta}_j}\frac{p_j(1-p_D^{j1})}{(1-p_j)(1-p_D^{j0})}\right)^2} \\
&\times \frac{1}{\left(e^{\hat{\beta}_j}+\frac{(1-p_j)(1-p_D^{j0})}{p_j(1-p_D^{j1})}\right)\left(m_{lj0}+e^{\hat{\beta}_j}m_{lj1}\right)} \\
&= 2\frac{e^{2\hat{\beta}_j}(1-\psi_j)\left((1-p_j)(1-p_D^{j0})+p_j(1-p_D^{j1})\right)}{n_j\psi_j(1-p_j)(1-p_D^{j0})\left(1+e^{\hat{\beta}_j}\frac{p_j(1-p_D^{j1})}{(1-p_j)(1-p_D^{j0})}\right)^2\left(e^{\hat{\beta}_j}+\frac{(1-p_j)(1-p_D^{j0})}{p_j(1-p_D^{j1})}\right)} \sum_l \frac{d_l^j-1/2}{\left(m_{lj0}+e^{\hat{\beta}_j}m_{lj1}\right)} \\
&= 2\frac{e^{2\hat{\beta}_j}(1-\psi_j)p_j(1-p_j)(1-p_D^{j1})(1-p_D^{j0})\left((1-p_j)(1-p_D^{j0})+p_j(1-p_D^{j1})\right)}{n_j\psi_j\left((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\right)^3} \\
&\times \sum_l \frac{d_{lj}-1/2}{\left(m_{lj0}+e^{\hat{\beta}_j}m_{lj1}\right)} \\
&\approx \frac{e^{2\hat{\beta}_j}(1-\psi_j)p_j(1-p_j)(1-p_D^{j1})(1-p_D^{j0})\left((1-p_j)(1-p_D^{j0})+p_j(1-p_D^{j1})\right)}{n_j\psi_j\left((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\right)^3} \frac{\left(D_0^j+D_1^j\right)^2}{\left(m_{*j0}+e^{\hat{\beta}_j}m_{*j1}\right)}
\end{aligned}
\tag{6.41}
$$

Further, $\frac{m_{*j1}}{n_j} \approx p_j(1-p_D^{j1})$ and $\frac{m_{*j0}}{n_j} \approx (1-p_j)(1-p_D^{j0})$ and therefore, $\hat{\delta}_j$ is given by

$$
\begin{aligned}
\hat{\delta}_j &\approx \frac{e^{2\hat{\beta}_j}(1-\psi_j)p_j(1-p_j)(1-p_D^{j1})(1-p_D^{j0})\left((1-p_j)(1-p_D^{j0})+p_j(1-p_D^{j1})\right)}{n_j\psi_j\left((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\right)^3} \frac{\left(D_0^j+D_1^j\right)^2}{\left(m_{*j0}+e^{\hat{\beta}_j}m_{*j1}\right)} \\
&\approx \frac{e^{2\hat{\beta}_j}(1-\psi_j)p_j(1-p_j)(1-p_D^{j1})(1-p_D^{j0})\left((1-p_j)(1-p_D^{j0})+p_j(1-p_D^{j1})\right)}{\psi_j\left((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\right)^3} \\
&\times \frac{\left((1-p_j)p_D^{j0}+p_jp_D^{j1}\right)^2}{\left(\frac{m_{*j0}}{n_j}+e^{\hat{\beta}_j}\frac{m_{*j1}}{n_j}\right)}
\end{aligned}
$$

$$\approx \frac{e^{2\hat{\beta}_j}(1-\psi_j)p_j(1-p_j)(1-p_D^{j1})(1-p_D^{j0})\left((1-p_j)(1-p_D^{j0})+p_j(1-p_D^{j1})\right)}{\psi_j\left((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\right)^4}$$
$$\times\left((1-p_j)p_D^{j0}+p_jp_D^{j1}\right)^2. \tag{6.42}$$

Similarly, $\hat{\sigma_j^2}$ can be written as

$$\begin{aligned}
\hat{\sigma_j^2} &= \frac{1}{n_j}\sum_k\sum_{i=1}^{n_k}\frac{\tilde{Y}_{j0}(T_{ijk})^2\Delta_{ij1}+e^{2\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk})^2\Delta_{ij0}}{\left(\tilde{Y}_{j0}(T_{ijk})+e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ijk})\right)^2}\\
&= \frac{1}{n_j}\sum_{i=1}^{n_0}\frac{e^{2\hat{\beta}_j}\tilde{Y}_{j1}(T_{ij0})^2\Delta_{ij0}}{\left(\tilde{Y}_{j0}(T_{ij0})+e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ij0})\right)^2}+\frac{1}{n_j}\sum_{i=1}^{n_1}\frac{\tilde{Y}_{j0}(T_{ij1})^2\Delta_{ij1}}{\left(\tilde{Y}_{j0}(T_{ij1})+e^{\hat{\beta}_j}\tilde{Y}_{j1}(T_{ij1})\right)^2}\\
&\approx \frac{1}{n_j}\sum_l\frac{e^{2\hat{\beta}_j}\psi_j^2m_{lj1}^2\mathbf{I}(l\in L_j^0)+\psi_j^2m_{lj0}^2\mathbf{I}(l\in L_j^1)}{\psi_j^2\left(m_{lj0}+e^{\hat{\beta}_j}m_{lj1}\right)^2}\\
&= \frac{1}{n_j}\sum_l\frac{e^{2\hat{\beta}_j}(\frac{m_{lj1}}{m_{lj0}})^2\mathbf{I}(l\in L_j^0)+\mathbf{I}(l\in L_j^1)}{\left(1+e^{\hat{\beta}_j}\frac{m_{lj1}}{m_{lj0}}\right)^2}\\
&\approx \frac{1}{n_j}\frac{e^{2\hat{\beta}_j}(\frac{p_j(1-p_D^{j1})}{(1-p_j)(1-p_D^{j0})})^2D_0^j+D_1^j}{\left(1+e^{\hat{\beta}_j}\frac{p_j(1-p_D^{j1})}{(1-p_j)(1-p_D^{j0})}\right)^2}=\frac{1}{n_j}\frac{e^{2\hat{\beta}_j}(p_j(1-p_D^{j1}))^2D_0^j+((1-p_j)(1-p_D^{j0}))^2D_1^j}{\left((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\right)^2}
\end{aligned}$$

where L_j^k, k = 0,1 is the index set of individuals experiencing an event who are in treatment group k and biomarker group j. Therefore, approximating D_k^j by $n_jp_j^k(1-p_j)^{1-k}(p_D^{j1})^k(p_D^{j0})^{1-k}$, $k=0,1$, we have

$$\hat{\sigma_j^2}\approx p_j(1-p_j)\frac{e^{2\hat{\beta}_j}p_j(1-p_D^{j1})^2p_D^{j0}+(1-p_j)(1-p_D^{j0})^2p_D^{j1}}{\left((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\right)^2}. \tag{6.43}$$

Based on the approximations, the asymptotic variance of $n_j^{1/2}(\hat{\beta}_j-\beta_j)$ is given by

$$\mathrm{var}_j\approx r_j^{-1}\left(\left[p_j(1-p_j)\frac{e^{2\hat{\beta}_j}p_j(1-p_D^{j1})^2p_D^{j0}+(1-p_j)(1-p_D^{j0})^2p_D^{j1}}{\left((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\right)^2}\right]^{-1}\right.$$

$$+\left[p_j(1-p_j)\frac{e^{2\hat{\beta}_j}p_j(1-p_D^{j1})^2p_D^{j0}+(1-p_j)(1-p_D^{j0})^2p_D^{j1}}{\Big((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\Big)^2}\right]^{-2}$$
$$\times\frac{e^{2\hat{\beta}_j}(1-\psi_j)p_j(1-p_j)(1-p_D^{j1})(1-p_D^{j0})\Big((1-p_j)(1-p_D^{j0})+p_j(1-p_D^{j1})\Big)}{\psi_j\Big((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\Big)^4}$$
$$\times\Big((1-p_j)p_D^{j0}+p_jp_D^{j1}\Big)^2\Bigg)$$
$$=r_j^{-1}\frac{1}{p_j(1-p_j)\left[e^{2\hat{\beta}_j}p_j(1-p_D^{j1})^2p_D^{j0}+(1-p_j)(1-p_D^{j0})^2p_D^{j1}\right]}$$
$$\times\Bigg[\Big((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\Big)^2+\frac{\left[e^{2\hat{\beta}_j}(1-\psi_j)(1-p_D^{j1})(1-p_D^{j0})\right]}{\psi_j\Big(e^{2\hat{\beta}_j}p_j(1-p_D^{j1})^2p_D^{j0}+(1-p_j)(1-p_D^{j0})^2p_D^{j1}\Big)}$$
$$\times\Big\{(1-p_j)(1-p_D^{j0})+p_j(1-p_D^{j1})\Big\}\Big((1-p_j)p_D^{j0}+p_jp_D^{j1}\Big)^2\Bigg] \tag{6.44}$$

Approximating $\frac{1-p_D^{j1}}{1-p_D^{j0}}\approx 1$ and $\frac{m_{*j0}}{n_j}\approx(1-p_j)\left[p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right]$ and $\frac{m_{*j1}}{n_j}\approx p_j\left[p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right]$, (6.40) can be written as

$$\hat{\delta}_j\approx\frac{e^{2\hat{\beta}_j}(1-\psi_j)p_j(1-p_j)\Big((1-p_j)p_D^{j0}+p_jp_D^{j1}\Big)^2}{\psi_j\Big((1-p_j)+e^{\hat{\beta}_j}p_j\Big)^4\Big(p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\Big)} \tag{6.45}$$

and (6.39) is

$$\hat{\sigma_j^2}\approx p_j(1-p_j)\frac{e^{2\hat{\beta}_j}p_jp_D^{j0}+(1-p_j)p_D^{j1}}{\Big((1-p_j)+e^{\hat{\beta}_j}p_j\Big)^2}. \tag{6.46}$$

Noting that the censoring distribution is a mixture distribution with probability $1-p_C$ being degenerate at τ and the event proportion being low, we can approximate $\frac{m_{*j0}}{n_j}$ by $(1-p_C)(1-p_j)\left[p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right]$ and $\frac{m_{*j1}}{n_j}$ by $(1-p_C)p_j\left[p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right]$, which relates to the proportion of people remaining at the risk set at the end of the study. Therefore,

$$\hat{\delta}_j\approx\frac{e^{2\hat{\beta}_j}(1-\psi_j)p_j(1-p_j)\Big((1-p_j)p_D^{j0}+p_jp_D^{j1}\Big)^2}{\psi_j\Big((1-p_j)+e^{\hat{\beta}_j}p_j\Big)^4(1-p_C)\Big(p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\Big)} \tag{6.47}$$

$$\text{and} \quad \hat{\sigma_j^2} \approx p_j(1-p_j)\frac{e^{2\hat{\beta}_j}p_jp_D^{j0} + (1-p_j)p_D^{j1}}{\Big((1-p_j) + e^{\hat{\beta}_j}p_j\Big)^2}. \tag{6.48}$$

C.4 Bounds of Power/Sample Size under Non-Rare Event Assumption

We propose the following lower bound and upper bound for the power calculation. Let us denote the risk set for the treatment group biomarker group j, at the l_{j1}-th time-point of the index set, by m_{lj1} and that for the control group in biomarker group j, at the l_{j0}-th time-point of the index set, by m_{lj0}. Based on Cai and Zeng (2007), we have the following bound for treatment group :

$$n_jp_j - n_{j1C} - D_1^j \le m_{lj1} \le n_jp_j - l_{j1} + 1, \quad j = 0, 1,$$

and for control group, it is

$$n_j(1-p_j) - n_{j0C} - D_0^j \le m_{lj0} \le n_j(1-p_j) - l_{j0} + 1, \quad j = 0, 1.$$

where n_{j1C} and n_{j0C} are the total number of dropouts in the treated group and non-treated group in biomarker group j respectively, in $[0, \tau)$. If all individuals experience the event at the beginning of the study, then the risk set is the lower bound. Similarly, if at all the event times there is only one event, we have the upper bound of the risk sets. Using the above conditions, we define $\delta_{j,ub}$ and $\delta_{j,lb}$ as the upper and lower bounds of $\hat{\delta}_j$ respectively.

$$\begin{aligned}
\hat{\delta}_{j,ub} \approx\ & 2\frac{e^{2\hat{\beta}_j}(1-\psi_j)p_j(1-p_j)(1-p_D^{j1})(1-p_D^{j0})\Big(p_j(1-p_D^{j1}) + (1-p_j)(1-p_D^{j0})\Big)}{n_j\psi_j\Big((1-p_j)(1-p_D^{j0}) + e^{\hat{\beta}_j}p_j(1-p_D^{j1})\Big)^3} \\
& \times \sum_{l=1}^{m}\sum_{k=0}^{1}\frac{d_{lj} - 1/2}{\Big(\{n_j(1-p_j) - l_{j0} + 1\}(1 + e^{\hat{\beta}_j}\frac{p_j(1-p_D^{j1})}{(1-p_j)(1-p_D^{j0})})\Big)^{1-k}} \\
& \times \frac{1}{\Big(\{n_jp_j - l_{j1} + 1\}(e^{\hat{\beta}_j} + \frac{(1-p_j)(1-p_D^{j0})}{p_j(1-p_D^{j1})})\Big)^k}
\end{aligned}$$

$$\text{and } \hat{\delta}_{j,lb} \approx 2\frac{e^{2\hat{\beta}_j}(1-\psi_j)p_j(1-p_j)(1-p_D^{j1})(1-p_D^{j0})\left(p_j(1-p_D^{j1})+(1-p_j)(1-p_D^{j0})\right)}{n_j\psi_j\left((1-p_j)(1-p_D^{j0})+e^{\hat{\beta}_j}p_j(1-p_D^{j1})\right)^3}$$

$$\times\sum_{l=1}^{m}\sum_{k=0}^{1}\frac{d_{lj}-1/2}{\left(\{n_j(1-p_j)-n_{j0C}-D_0^j\}(1+e^{\hat{\beta}_j}\frac{p_j(1-p_D^{j1})}{(1-p_j)(1-p_D^{j0})})\right)^{1-k}}$$

$$\times\frac{1}{\left(\{n_jp_j-n_{j1C}-D_1^j\}(e^{\hat{\beta}_j}+\frac{(1-p_j)(1-p_D^{j0})}{p_j(1-p_D^{j1})})\right)^{k}},$$

Using this, one can get the bounds for power as

$$\Phi\left[\frac{\sqrt{n}\,(\beta_1-\beta_0)}{\sqrt{\sum_{j=0}^{1}r_j^{-1}\left(\sigma_j^{-2}+\sigma_j^{-4}\delta_{j,ub}\right)}}-Z_{1-\alpha}\right]\leq \text{Power}\leq\Phi\left[\frac{\sqrt{n}\,(\beta_1-\beta_0)}{\sqrt{\sum_{j=0}^{1}r_j^{-1}\left(\sigma_j^{-2}+\sigma_j^{-4}\delta_{j,lb}\right)}}-Z_{1-\alpha}\right],$$

where $\sigma_j^2 \approx (1-p_j)p_j\frac{e^{2\beta_j}p_j(1-p_D^{j1})^2p_D^{j0}+(1-p_j)(1-p_D^{j0})^2p_D^{j1}}{\left((1-p_j)(1-p_D^{j0})+e^{\beta_j}p_j(1-p_D^{j1})\right)^2}$.

REFERENCES

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. *Lecture notes in statistics* **2,** 25.

Albert, P. S. (1991). A two-state markov mixture model for a time series of epileptic seizure counts. *Biometrics* **47,** 1371–1381.

Amorim, L. D. and Cai, J. (2015). Modelling recurrent events : a tutorial for analysis in epidemiology. *International Journal of Epidemiology* pages 324–333.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* **10,** 1110–1120.

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

Barlow, W. E., Ichikawa, L., Rosner, D., and Izumi, S. (1999). Analysis of case-cohort designs. *Journal of Clinical Epidemiology* **52,** 1165–1172.

Barreto, M. L., Farenzena, G. G., Fiaccone, R. L., Santos, L. M. P., Assis, A. M. O., Araujo, M. P. N., and Santos, P. A. B. (1994). Effect of vitamin a supplementation on diarrhoea and acute lower-respiratory-tract infections in young children in brazil. *Lancet* **344,** 228–231.

Bilias, Y., Gu, M., and Ying, Z. (1997). Towards a general asymptotic theory for cox model with staggered entry. *The Annals of Statistics* **25,** 662–682.

Borgan, O., Goldstein, L., and Langholz, B. (1995). Methods for the analysis of sampled cohort data in the cox proportional hazards model. *The Annals of Statistics* **23,** 1749–1778.

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

Breslow, N. E. and Wellner, J. A. (2006). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to cox regression. *Scandinavian Journal of Statistics* **34,** 86–102.

Cai, J. and Schaubel, D. E. (2004). Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Analysis* **10,** 121–138.

Cai, J. and Zeng, D. (2004). Sample size/power calculation for case-cohort studies. *Biometrics* **60,** 1015–1024.

Cai, J. and Zeng, D. (2007). Power calculation for case-cohort studies with non-rare events. *Biometrics* **63,** 1288–1295.

Cao, Y. and Yu, J. (2017). Optimal generalized case-cohort sampling design under the additive hazard model. *Communications in Statistics - Theory and Methods* **46,** 4484–4493.

Chang, S.-H. and Wang, M.-C. (1999). Conditional regression analysis for recurrence time data. *Journal of the American Statistical Association* **94,** 1221–1230.

Chen, F. and Chen, K. (2014). Case-cohort analysis of clusters of recurrent events. *Lifetime Data Analysis* **20,** 1 – 15.

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

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

Cole, E. H., Cattran, D. C., Farewell, V. T., Aprile, M., Bear, R. A., Pei, Y. P., Fenton, S. S., Tober, J. A. L., and Cardella, C. J. (1994). A comparison of rabbit antithymocyte serum and okt3 as prophylaxis against renal allograft rejection. *Transplantation* **57,** 60–67.

Cook, R. J., Lawless, J. F., Lakhal-Chaieb, L., and Lee, K.-A. (2009). Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination: Application to skeletal complications in cancer metastatic to bone. *Journal of the American Statistical Association* **104,** 60–75.

Cook, R. J., Lawless, J. F., and Nadeau, C. (1996). Robust tests for treatment comparisons based on recurrent event responses. *Biometrics* **52,** 557–571.

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

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

Cui, J., Forbesa, A., Kirby, A., Marschner, I., Simes, J., Weste, M., and Tonkina, A. (2008). Parametric conditional frailty models for recurrent cardiovascular events in the lipid study. *Clinical Trials* **5,** 565–574.

Dong, C.-L., Zhou, J., and Sun, L.-Q. (2014). A class of weighted estimators for additive hazards model in case-cohort studies. *Acta Mathematicae Applicatae Sinica, English Series* **30,** 1153–1168.

Folsom, A., Aleksic, N., Park, E., Salomaa, V., Juneja, H., and Wu, K. K. (2001). Prospective study of fibrinolytic factors and incident coronary heart disease : The atherosclerosis risk in communities (aric) study. *Arteriosclerosis, Thrombosis, and Vascular Biology* **21,** 611–617.

Gail, M. H., Santner, T. J., and Brown, C. C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* **36,** 255–266.

He, S., Wang, F., and Sun, L. (2013). A semiparametric additive rates model for clustered recurrent event data. *Acta Mathematicae Applicatae Sinica, English Series* **29,** 53–62.

Hu, W., Cai, J., and Zeng, D. (2014). Sample size/power calculation for stratified case-cohort design. *Statisics in Medicine* **33,** 3973–3985.

Huang, Y. and Chen, Y. Q. (2003). Marginal regression of gaps between recurrent events. *Lifetime Data Analysis* **9,** 293–303.

Huffer, F. W. and McKeague, I. W. (1991). Weighted least squares estimation for aalen's additive risk model. *Journal of the American Statistical Association* **86,** 114–129.

Jahn-Eimermacher, A., Ingel, K., Ozga, A.-K., Preussler, S., and Binder, H. (2015). Simulating recurrent event data with hazard functions defined on a total time scale. *BMC Medical Research Methodology* **15,** 16–24.

Jiang, J. and Zhou, H. (2007). Additive hazard regression with auxiliary covariates. *Biometrika* **94,** 359–369.

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

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data, 2nd Ed.* John Wiley, New York.

Kang, S. and Cai, J. (2009a). Marginal hazards model for case-cohort studies with multiple disease outcomes. *Biometrika* **96,** 887–901.

Kang, S. and Cai, J. (2009b). Marginal hazards regression for retrospective studies within cohort with possibly correlated failure time data. *Biometrics* **65,** 405–414.

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.

Karatzas, I. and Shereve, S. E. (1988). *Brownian Motion and Stochastic Calculus*. Springer New York.

Kim, S., Cai, J., and Couper, D. (2016). Improving the efficiency of estimation in the additive hazards model for stratified casecohort design with multiple diseases. *Statistics in Medicine* **35,** 282–293.

Klien, J. (1992). Semiparametric estimation of the random effects using the cox model based on the em algorithnm. *Biometrics* **48,** 795–806.

Kulich, M. and Lin, D. (2000). Additive hazards regression 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,** 832–844.

Langholz, B. and Thomas, D. C. (1990). Nested case-control and case-cohort methods of sampling from a cohort : A critical comparison. *American Journal of Epidemiology* **131,** 169–176.

Lawless, J. F. (1995). The analysis of recurrent events for multiple subjects. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **44,** 487–498.

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

Li, Q. H. and Lagakos, S. W. (1997). Use of the wei-lin-weissfeld method for the analysis of a recurring and a terminating event. *Statisics in Medicine* **16,** 925–940.

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

Liao, D., Cai, J., Rosamond, W. D., Barnes, R. W., Hutchinson, R. G., Whrtsel, E. A., Rautaharju, P., and Heiss, G. (1997). Cardiac autonomic function and incident coronary heart disease: A population-based case-cohort study. *American Journal of Epidemiology* **145,** 696–706.

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

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

Lin, D. Y., Sun, W., and Ying, Z. (1999). Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika* **86,** 59–70.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of the recurrent events. *Journal of the Royal Statistical Society (Series B)* **62,** 711–730.

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

Liu, Y., Sun, L., and Zhou, Y. (2013). Additive transformation models for recurrent events. *Communications in Statistics - Theory and Methods* **42,** 4043–4055.

Liu, Y., Wu, Y., Cai, J., and Zhou, H. (2010). Additiv-multiplicative rates model for recurrent events. *Lifetime Data Analysis* **16,** 353–373.

Lu, S.-E. and Shih, J. H. (2006). Case-cohort designs and analysis for clustered failure time data. *Biometrics* **62,** 1138–1148.

Lu, W. and Tsiatis, A. (2006). Semiparametric transformation models for the case-cohort study. *Biometrika* **93,** 207–214.

Ma, S. (2007). Additive risk model with case-cohort sampled current status data. *Statistical Papers* **48,** 595–608.

Martinussen, T. and Scheike, T. H. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika* **89,** 649–658.

Nan, B. (2004). Efficient estimation of case-cohort studies. *The Canadian Journal of Statistics* **32,** 403–419.

Nelson, W. (1988). Graphical analysis of system repair data. *Journal of Quality Technology* **20,** 24–35.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association* **88,** 811–820.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics.

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

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68,** 373–379.

Savitz, D. A., Cai, J., Wijngaarden, E. V., Loomis, D., Mihlan, G., Dufort, V., Kleckner, R. C., Nylander-French, L., Kromhout, H., and Zhou, H. (2000). Case-cohort analysis of brain cancer and leukemia in electric utility workers using a rened magnetic field job-exposure matrix. *American Journal of Industrial Medicine* **38,** 417–425.

Schaubel, D. E. and Cai, J. (2004). Regression methods for gap time hazard functions of sequentially ordered multivariate failure time data. *Biometrika* **91,** 291–303.

Schaubel, D. E. and Cai, J. (2005). Semiparametric methods for clustered recurrent event data. *Lifetime Data Analysis* **11,** 405–425.

Schaubel, D. E., Zeng, D., and Cai, J. (2006). A semiparametric additive rates model for recurrent event data. *Lifetime Data Analysis* **12,** 389–406.

Schouten, E. G., Dekker, J. M., Kok, F. J., Cessie, S. L., Van Houwelingen, H. C., Pool, J., and Vanderbroucke, J. P. (1993). Risk ratio and rate ratio estimation in case-cohort designs : Hypertension and cardiovascular mortality. *Statistics in Medicine* **12,** 1733–1745.

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

Shepherd, F. A., Domerg, C., Hainaut, P., Jönne, P. A., Pignon, J.-P., Graziano, S., Douillard, J.-Y., Brambilla, E., Chevalier, T. L., Seymour, L., Bourredjem, A., Teuff, G. L., Pirker, R., Filipits, M., Rosell, r., Kratzke, R., Bandarchi, B., Ma, X., Capelletti, M., Soria, J.-C., and Tsao, M.-S. (2013). Pooled analysis of the prognostic and predictive effects of kras mutation status and kras mutation subtype in early-stage resected nonsmall-cell lung cancer in four trials of adjuvant chemotherapy. *Journal of Clinical Oncology* **31,** 2173–2181.

Spiekerman, C. F. and Lin, D. Y. (1999). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* **93,** 1164–1175.

Sun, J., Sun, L., and Flournoy, N. (2004). Additive hazards model for competing risks analysis of the case-cohort design. *Communications in Statistics - Theory and Methods* **33,** 351–366.

Sun, L., Park, D.-H., and Sun, J. (2006). The additive hazards model for recurrent gap times. *Statistica Sinica* **16,** 919–932.

Therneau, T. M. and Grambsch, P. M. (2001). *Modeling Survival Data*. Springer New York.

Therneau, T. M. and Hamilton, S. A. (1997). rhdnase as an example of recurrent event analysis. *Statistics in Medicine* **16,** 2029–2047.

Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer New York.

Wacholder, S. (1991). Practical considerations in choosing between the case-cohort and nested case-control designs. *Epidemiology* **2,** 155–158.

Wacholder, S., Gail, M. H., Pee, D., and Brookmeyer, R. (1989). Alternative variance and efficiency calculations for the case-cohort design. *Biometrika* **76,** 117–123.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84,** 1065–1073.

Yin, G. and Cai, J. (2004). Additive hazards model with multivariate failure time data. *Biometrika* **91,** 801–818.

Yu, J. C., Shi, Y. Y., Yang, Q. L., and Liu, Y. Y. (2014). Additive hazards regression under generalized case-cohort sampling. *Acta Mathematica Sinica, English Series* **30,** 251–260.

Zeng, D. and Cai, J. (2010). A semiparametric additive rate model for recurrent events with an informative terminal event. *Biometrika* **97,** 699–712.

Zeng, D., Lin, D. Y., Avery, C. L., North, K. E., and Bray, M. S. (2006). Efficient semiparametric estimation of haplotype-disease associations in case-cohort and nested case-control studies. *Biostatistics* **7,** 486–502.

Zhang, H., Schaubel, D., and Kalbeisch, J. D. (2011). Proportional hazards regression for the analysis of clustered survival data from case-cohort studies. *Biometrics* **67,** 18–28.