國 立 交 通 大 學

統 計 學 研 究 所

博 士 論 文

在競爭風險下，

目標事件的累積發生機率的回歸分析

Regression Analysis for Cumulative Incidence Probability under Competing Risks

研 究 生 ：張瑋華

指導教授 ：王維菁博士

中華民國九十六年七月

摘 要

許多疾病的病程包含數種以上的風險，例如乳癌患者可能會經歷癌細胞的復發，轉移甚至病情嚴重導致死亡。當研究者感興趣的主題事件為非終端事件時(例如癌細胞的復發)，若終端事件先發生，就無法觀察到主題事件。在存活分析的架構下，死亡被視為復發事件的競爭風險。此時如何推估主題事件的發生機率是個熱門的研究主題。本篇論文在迴歸的架構下探討自變數對主題事件發生機率的影響。

分析實證資料中常會因設限(censoring)而只能記錄到不完整的資訊。針對此現象，本論文提出兩種偏誤修正的方法以估計迴歸模式的參數。第一個方法利用設限機率的倒數做為權數以改正因設限造成的偏誤，稱之為 IPCW；第二個方法則以缺失值的條件期望值做為填補不完整的資訊，稱之為 Imputation。論文中我們推導迴歸參數估計量的大樣本性質，並藉由模擬以驗證所提出的估計方法在有限樣本時的表現。本論文亦將所提的迴歸模型和估計方法應用在史丹佛心臟移植資料和非典型肺炎(SARS)資料做為實例的佐證。

關鍵字：競爭風險，累積發生率函數，加權法，補插法，估計方程式

Abstract

In the dissertation, we consider regression analysis for the cumulative incidence probability under the framework of competing risks. Instead of modeling the whole function which usually involves making stronger assumptions, we investigate the effect of covariates on the cumulative incidence rate at a pre-specified time point.

The information of incidence may be missing due to censoring. We apply two approaches to handle incomplete data. The first method utilizes the technique of the inverse probability of censoring weighting (IPCW) to correct the sampling bias. The other approach is to impute missing variables by an estimate of its conditional mean. Both methods are popular and useful tools in handling missing data. Large-sample properties of the proposed methods are also derived. Simulations are performed to examine finite-sample performances of the proposed methods. The Stanford Heart Transplant data and the severe acute respiratory syndrome (SARS) data are analyzed to illustrate the applicability of the proposed model and inference methods.

Key words : Cause-specific hazard; Mixture model; Transformation model; Imputation; Cumulative incidence function; Inverse probability of censoring; Imputation; Logistic regression; Missing Data

致 謝

這條研究的路走得很漫長，回首又彷彿一切只在昨日。從大學進入統計系到現在即將取得統計博士學位，在這過程中，我曾自認為學會了統計，哪裡都混得下去；也曾沮喪的懷疑除了課本上的數學式子，我其實什麼也不懂。但我相信在那些複雜的統計模型和高深的數學方程式的背後，應該有個道理涵蓋了一切也解釋了一切。就像老子的“道可道，非常道”，只是我還未能體會而已。

衷心感謝我的指導教授王維菁老師對我的教誨和悉心指導，並在論文研究過程中不斷為我指點迷津，心中謝意實難以隻字片語表達。感謝陳志榮老師和洪慧念老師在論文口試階段所提供的寶貴意見，兩位老師平日在課業上的指導和啟發也讓我獲益良多。感謝陳珍信老師和程毅豪老師撥空前來參加論文口試，並提出許多寶貴的建議以資本篇論文的修正。

謝謝所上老師們在我求學期間的指導與教誨，另外特別感謝陳珍信老師和程毅豪老師在我擔任中研院統計所臨時助理的期間所給予的指導和關心，僅向諸位老師深深的一鞠躬已表示我誠摯的敬意與謝忱。

謝謝班上的同學與已畢業或未畢業的學長姐和學弟妹所給予的鼓勵；也謝謝在中研院的同事們對我的鼓勵和幫助。

最後感謝我的家人，有你們的支持，我才能走到今天。

Contents

Chapter 1

Introduction

In biomedical studies, researchers may encounter analysis of multiple events data which are often formulated under the framework of competing risks. Here a competing risk is defined as an event whose occurrence prevents the occurrence of other events. For example, when the relapse of leukemia is the outcome of interest, death without relapse is a competing risk event. Other examples can be found in clinical trials in which subjects are usually subject to multiple risks. In breast cancer studies, a patient may experience multiple events, such as local recurrence, distant metastasis, a second primary cancer other than the original one and death. Researchers are sometimes more concerned about an event of particular type.

Under the framework of competing risks, let T be the failure time and $\tilde{B}$ be the corresponding cause of failure taking values in the set $\{1, \ldots, J\}$. Competing risks data can be summarized naturally by the following two quantities. One is the cause-specific hazard function defined as

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{\Pr(T \in [t, t + \Delta t), \tilde{B} = j | T \geq t)}{\Delta t}$$

which is the rate of occurrence for type-j failure in the presence of all causes of failure. The other is the cumulative incidence function or the crude failure probability defined as

$$F_j(t) = \Pr(T \leq t, \tilde{B} = j)$$

which describes the cumulative probability of developing type-j failure by time t (Pepe and Mori, 1993). Note that both quantities make no assumption about the relationship between the competing risks events and hence can be estimated nonparametrically. The cause-specific hazard function is useful in quantifying the instantaneous risk of a particular cause for alive individuals. However when the interest is on cumulative or overall risks of a particular cause, the cumulative incidence function is more intuitively appealing and easily explained to the clinicians.

We aim to study the effect of covariates on the cumulative incidence function. There exist several regression models which are constructed based on different decompositions of $F_j(t)$. Some authors including Cheng, Fine and Wei (1998) suggested making inference for the cumulative incidence function by modeling the cause-specific hazards of all causes. Specifically, they considered the following decomposition

$$F_j(t) = \int_0^t \lambda_j(u) \exp\left[-\int_0^u \sum_{j=1}^{J} \lambda_j(v) dv \right] du.$$

However, since the effect of a covariate on $\lambda_j(t)$ can be very different from its effect on $F_j(t)$, such an indirect approach can be misleading. Also the parameters in the models for the cause-specific hazards may lack a simple interpretation in terms of the crude failure probabilities.

In this dissertation we review existing papers which are more closely related to our framework based on a mixture formulation. This approach has also received growing attention due to its relation with cure models. Section 2 contains a review of related models. In Section 3, we review inference methods which have been developed for handling incomplete data. Our proposal is presented in Section 4, and simulations, the analysis of Stanford Heart Transplant data and the analysis of severe acute respiratory syndrome (SARS) data are presented in Section 5. The SARS example highlights the applicability of the proposed regression model. The proposed inference methods are useful when complete data are not available. For example, when the epidemic disease is still ongoing, interim analysis based on incomplete data is still useful for making timely decision.

Chapter 2

Models for Cumulative Incidence Function – A Review

2.1 Transformation Models

In recent years, statistical analysis based on transformation models has received substantial attention due to its wide applicability. Under the classical formulation without competing risk (i.e., $J = 1$), the model can be expressed as

$$m(T) = -\mathbf{Z}^T\boldsymbol{\gamma} + \epsilon,$$

where $m(t)$ is an unspecified strictly increasing function mapping from $(0, \infty)$ to $(-\infty, \infty)$, $\mathbf{Z}$ is a $p \times 1$ vector of covariates, $\boldsymbol{\gamma}$ is the corresponding vector of parameters and ϵ is a random error with a completely known distribution F_ϵ. Alternatively, the model can also be expressed as

$$H(\Pr(T \leq t|\mathbf{Z})) = m(t) + \mathbf{Z}^T\boldsymbol{\gamma}, \tag{2.1}$$

where $H = F_\epsilon^{-1}$. This class of models contains useful members. For example if F_ϵ is the extreme value distribution with $F_\epsilon(s) = 1 - \exp\{-\exp(s)\}$, $H(\cdot)$ becomes a complementary log-log link and model (2.1) is the Cox proportional hazard model. If F_ϵ is the standard logistic distribution, $H(\cdot)$ becomes the logit link and the model is the

proportional odds model.

The class of transformation models mentioned above has been modified to describe the covariate effect on the cumulative incidence function. When $J > 1$, the model assumes that

$$H(F_1(t|\mathbf{Z})) = m(t) + \mathbf{Z}^T\boldsymbol{\theta}, \tag{2.2}$$

where H is an increasing link function mapping from $(0,1)$ to $(-\infty,\infty)$, $\boldsymbol{\theta}$ is the $p \times 1$ vector of parameters and $m(t)$ is the baseline function. It is easy to see that $m(t) = H\{F_0(t)\}$ in which $F_0(t)$ represents the baseline cumulative incidence function when $\mathbf{Z} = 0_{p\times 1}$ or $\boldsymbol{\theta}$ vanishes. Model (2.2) describes the situation that under a suitable transformation, the cumulative incidence functions for subjects with different covariate values are "parallel" over the entire time span with distance being measured by a linear combination of $\boldsymbol{\theta}$.

Depending on whether F_0 is specified in an explicit form or not, model (2.2) can be classified as a parametric or semi-parametric transformation model. Fine and Gray (1999) first proposed the semi-parametric transformation model under the complementary log-log link function with $H(u) = \log[-\log(1-u)]$; Fine (2001) and Klein and Andersen (2005) extended the class of H to include any well-defined monotone function, say the logistic function, $H(u) = \log[u/(1-u)]$. Jeong and Fine (2006) proposed a parametric transformation model by specifying the baseline $F_0(t)$ with the improper Gompertz distribution, $B(t;\rho,\psi) = 1 - \exp\{\psi[1-\exp(\rho t)]/\rho\}$ where $\rho < 0$ and $0 < \psi < \infty$, and adopted the odds rate transformation as the link function,

$$H(v;\phi) = \begin{cases} \log\left[\frac{(1-v)^{-\phi}-1}{\phi}\right] & \text{if} \quad \phi \neq 0, \\ \log[-\log(1-v)] & \text{if} \quad \phi = 0, \end{cases}$$

which includes the complementary log-log link and the logistic link when $\phi = 0$ and $\phi = 1$, respectively.

Model (2.2) can be viewed under the context of a cure model (Fine and Gray, 1999; Fine, 2001) in which $F_1(t|\mathbf{Z})$ is the distribution function of the improper failure time

$$T_1 = T \cdot I(\tilde{B} = 1) + \infty \cdot I(\tilde{B} \neq 1). \tag{2.3}$$

Note that $\Pr(T_1 \leq t|\mathbf{Z}) = \Pr(T \leq t, \tilde{B} = 1|\mathbf{Z}) = F_1(t|\mathbf{Z})$. The difference between T_1 and the usual failure time is that $\Pr(T_1 = \infty) > 0$ which measures the proportion of individuals failing from causes other than the first type. The way of treating other competing risks as "cure" or "immune" sometimes lacks interpretability. Specifically, consider the "hazard" of the improper variable T_1 defined by

$$\tilde{\lambda}_1(t|\mathbf{Z}) = -d\log[1 - F_1(t|\mathbf{Z})]/\,dt = \frac{dF_1(t|\mathbf{Z})/dt}{1 - F_1(t|\mathbf{Z})}$$

which is also called the subdistribution hazard. We find that the denominator in the last identity which indicates the at-risk probability for failure of the first type at time t, always includes the quantity $\Pr(\tilde{B} \neq 1)$. It seems not very sensible to view those who have failed from other causes as always being "at risk" later on for a failure type that will never occur.

2.2 Mixture Models

The cumulative incidence function can also be written as the mixture form,

$$F_j(t) = \pi_j[1 - Q_j(t)], \quad j = 1, \ldots, J, \tag{2.4}$$

where $\pi_j = \lim_{t\to\infty} F_j(t) = \Pr(\tilde{B} = j)$ measures the marginal probability of type-j failure, and $1-Q_j(t) = \Pr(T \leq t|\tilde{B} = j)$ describes the corresponding latency distribution for the sub-population with $\tilde{B} = j$. Such a mixture formula was originally developed for the improper distribution function encountered in analysis of failure time with long-term survivors (Maller and Zhou, 1996). Nonparametric analysis of model (2.4) based on the competing risks data has been studied by Betensky and Schoenfeld (2001) in which the acute respiratory distress syndrome (ARDS) data with two competing events (cure or death) was used as an example, and by Wang (2003) under a two-path framework. In presence of right censoring, it has been mentioned that nonparametric maximum likelihood estimators of π_j and $Q_j(t)$ are consistent only if the support of failure time is shorter than the support of censoring time. That is, the follow-up must be long enough so that failure times of all individuals have positive probabilities to be observed.

Maller and Zhou (2002) called such a condition "sufficient follow-up" which assesses how "heavy" the tail of the censoring distribution is, relative to the survival distribution, while still permitting consistent estimation. Specifically, define the endpoint of the support of failure time T as τ_T=sup$\{t$: $\Pr(T > t) > 0\}$ and that of censoring time C as τ_C =sup$\{t$: $\Pr(C > t) > 0\}$. Under model (2.4), the "sufficient follow-up" for nonparametric analysis is $\tau_T \leq \tau_C$. When $\tau_T > \tau_C$ which happens in most longitudinal studies, additional assumptions are needed to avoid underestimation of π_j and $Q_j(t)$. Wang (2003) assumed that $\Pr(\tilde{B} = j|T > \tau_C) = \Pr(\tilde{B} = j)$ and then derived the estimators of π_j and $Q_j(t)$. However, such an assumption is quite subjective and hard to check empirically. The "sufficient follow-up" assumption can be much relaxed if one imposes a parametric or semiparametric model on the latency distribution. Maller and Zhou (2002) showed that under the parameterization that $Q_j(t) = Q(t; \boldsymbol{\varphi}_j)$, consistent estimators of π_j and $\boldsymbol{\varphi}_j$ can be obtained through the maximum likelihood approach if $1 - Q(\tau_C; \boldsymbol{\varphi}_j) > 0$ for $j = 1, \ldots, J$, which only requires that each cause of failure has positive probability to be observed. The parameterization of $Q_j(t)$ can also be extended to a regression setting. See Ghitany, Maller and Zhou (1994) and Vu, Maller and Zhou (1998) for further references.

Model (2.4) can be expressed under the following regression framework,

$$F_j(t|\mathbf{Z}) = \pi_j(\mathbf{Z})[1 - Q_j(t|\mathbf{Z})]. \tag{2.5}$$

Larson and Dinse (1985) assumed a multinomial logit model for $\pi_j(\mathbf{Z})$ and a parametric proportional hazards model for $Q_j(t|\mathbf{Z})$ with

$$Q_j(t|\mathbf{Z}) = \exp\left[-\int_0^t h_j(u)\exp(\mathbf{Z}^T\boldsymbol{\varphi})du\right],$$

where the baseline hazard $h_j(t)$ is specified as a piecewise exponential function. Kuk (1992) and Ng and McLachlan (2003) generalized the mixture model of Larson and Dinse (1985) by assuming $h_j(t)$ is unknown. To remedy the support problem, Fine (1999) considered the representation

$$\begin{aligned} F_j(t \wedge \tau) &= \Pr(T \leq \tau, \tilde{B} = j)\Pr(T \leq t|T \leq \tau, \tilde{B} = j) \\ &= F_j(\tau)[1 - Q_j(t|\tau)], \end{aligned}$$

where $U \wedge V = \min(U, V)$ and τ is a pre-determined time point located inside the support of the observed time variable. Statistical inference of $F_j(\tau)$ and $Q_j(t|\tau)$ are no longer subject to the potential problem of non-identifiability if τ is chosen properly such that $\Pr(T \wedge C > \tau) > 0$. Accordingly one can consider the regression model

$$\begin{aligned} F_j(t \wedge \tau|\mathbf{Z}) &= \Pr(T \leq \tau, \tilde{B} = j|\mathbf{Z})\Pr(T \leq t|T \leq \tau, \tilde{B} = j, \mathbf{Z}) \\ &= F_j(\tau|\mathbf{Z})[1 - Q_{j,\mathbf{Z}}(t|\tau)]. \end{aligned} \tag{2.6}$$

Fine (1999) assumed a binary regression model, namely the logistic model, for $F_j(\tau|\mathbf{Z})$ and the transformation model stated in (2.1) for the latency distribution $1 - Q_{j,\mathbf{Z}}(t|\tau)$.

Chapter 3

Review of Inference Methods in the Presence of Right Censoring

In this section we review inference methods that have been applied to solving incomplete data due to right censoring. In Section 3.1, we review how the maximum likelihood approach is applied to the analysis of mixture models. In Sections 3.2 and 3.3, we review two moment-based approaches which use different techniques to handle missing data. From this section on, the notations are unified as follows. Let $\mathbf{Z}$ be the $p \times 1$ vector of covariates, T be the failure time with the survival function $S(t) = \Pr(T > t)$ and C be the censoring time with $G(t) = \Pr(C \geq t)$. In the presence of right censoring, one observes $(X, \delta, \mathbf{Z})$, where $X = T \wedge C$ and $\delta = I(T \leq C)$.

3.1 Likelihood Estimation

In presence of right censoring, the competing risks data can be denoted as

$$(X, B, \mathbf{Z}) = \{(X_i, B_i, \mathbf{Z}_i) : i = 1, \ldots, n\}, \tag{3.1}$$

where $B = \tilde{B} \cdot \delta$. Note that $B = 0$ refers to the censored case in which the value of $\tilde{B}$ is unknown. Assuming that T and C are independent given $\mathbf{Z}$, the likelihood function

based on the data in (3.1) can be written as

$$L_F = \prod_{i=1}^{n} \left\{ \left[\prod_{j=1}^{J} f_j(X_i; \mathbf{Z}_i)^{I(B_i=j)} \right] S(X_i; \mathbf{Z}_i)^{I(B_i=0)} \right\}, \tag{3.2}$$

where $f_j(t; \mathbf{Z}) = dF_j(t; \mathbf{Z})/\, dt$ and $S(t; \mathbf{Z}) = \Pr(T > t|\mathbf{Z}) = 1 - \sum_{j=1}^{J} F_j(t; \mathbf{Z})$. Due to censoring, the likelihood function has to combine the information of failure of all types and can not be factored into separate pieces for each type as in the case of complete data. When the mixture model (2.5) is assumed, the likelihood function can be expressed as

$$L_M = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{J} \left[\pi_j(\mathbf{Z}_i) q_j(X_i; \mathbf{Z}_i) \right]^{I(B_i=j)} \left[\sum_{j=1}^{J} \pi_j(\mathbf{Z}_i) Q_j(X_i; \mathbf{Z}_i) \right]^{I(B_i=0)} \right\},$$

where $q_j(t; \mathbf{Z}) = -dQ_j(t; \mathbf{Z})/\, dt$ and $\sum_{j=1}^{J} \pi_j(\mathbf{Z}) = 1$. Because the summation term within the products on the right-hand side makes the maximization of L_M very difficult, Larson and Dinse (1985) suggested to apply the EM algorithm to facilitating maximum likelihood estimation. This algorithm has also been utilized by Kuk (1992) and Ng and McLachlan (2003) in estimation of the semi-parametric proportional hazard/logistic mixture model.

To illustrate the use of the EM algorithm, it is assumed temporarily that indicators of the failure types (i.e., $\tilde{B}_i$'s) are available and the likelihood based on such pseudo-observations is given by

$$L_s \propto \prod_{i=1}^{n} \prod_{j=1}^{J} \left\{ \pi_j(\mathbf{Z}_i) \left[q_j(X_i; \mathbf{Z}_i) \right]^{\delta_i} \left[Q_j(X_i; \mathbf{Z}_i) \right]^{1-\delta_i} \right\}^{I(\tilde{B}_i=j)}.$$

The algorithm consists of two steps. First, the E-step is to compute

$$\begin{aligned} l_s &= E\left(\log(L_s) \,\middle|\, X, B, \mathbf{Z}, (\pi_j^{(m)}, Q_j^{(m)}) \text{ for } j = 1, \ldots, J \right), \\ &= \sum_{i=1}^{n} \sum_{j=1}^{J} \left\{ I(B_i = j) \cdot \log[h_j(X_i; \mathbf{Z}_i)] + w_{i,j}^{(m)} \cdot \log[\pi_j(\mathbf{Z}_i) Q_j(X_i; \mathbf{Z}_i)] \right\}, \end{aligned}$$

where $h_j(t; \mathbf{Z}) = q_j(t; \mathbf{Z})/\, Q_j(t; \mathbf{Z})$, $(\pi_j^{(m)}, Q_j^{(m)})$ denotes the model expressions for $\pi_j(\mathbf{Z})$ and $Q_j(t; \mathbf{Z})$ with the parameters being replaced by the corresponding estimated values

obtained at the mth iteration, and

$$\begin{aligned} w_{i,j}^{(m)} &= E\left(I(\tilde{B}_i = j) \,\middle|\, X, B, \mathbf{Z}, (\pi_j^{(m)}, Q_j^{(m)}) \text{ for } j = 1, \ldots, J\right), \\ &= I(B_i = j) + I(B_i = 0) \cdot \frac{\pi_j^{(m)}(\mathbf{Z}_i) Q_j^{(m)}(X_i; \mathbf{Z}_i)}{\sum_{k=1}^{J} \pi_k^{(m)}(\mathbf{Z}_i) Q_k^{(m)}(X_i; \mathbf{Z}_i)}. \end{aligned} \tag{3.3}$$

Note that the last term of $w_{i,j}^{(m)}$ is the conditional probability that the ith patient will experience the event of type j given that the failure of all types not occurred by time X_i. Next, the M-step of the algorithm involves maximizing, regarding $w_{i,j}^{(m)}$ as fixed, the log-likelihood l_s which can be expressed as

$$l_s = l_\pi + l_{Q_1} + \cdots + l_{Q_J},$$

where

$$l_\pi = \sum_{i=1}^{n} \sum_{j=1}^{J} w_{i,j}^{(m)} \log[\pi_j(\mathbf{Z}_i)]$$

and

$$l_{Q_j} = \sum_{i=1}^{n} \left\{ I(B_i = j) \log[h_j(X_i; \mathbf{Z}_i)] + w_{i,j}^{(m)} \log[Q_j(X_i; \mathbf{Z}_i)] \right\}.$$

The EM procedure is iterative in a way that the estimates obtained previously are used to update the value of $w_{i,j}^{(m)}$ in the current maximization step. One of its attractive features is that $J + 1$ components of l_s can be maximized separately. The convergence properties of the estimators obtained from the EM procedure have been discussed in Dempster *et al.* (1977), Wu (1983) and Louis (1982).

3.2 Inverse Probability of Censoring Weighting (IPCW)

Under right censoring, individuals with larger failure times have more chance to be censored than those with smaller ones. However, uncensored observations are still useful proxies if their bias can be corrected. The inverse probability of censoring weighting (IPCW) has been used to correct such bias. Let $\tilde{V}$ be a function of failure time T, say $\tilde{V} = I(T \leq t, \tilde{B} = j)$ or $I(T \leq t)$. Denote $V = \tilde{V}\delta$ which can be viewed as an observed

proxy of $\tilde{V}$. Assume that T and C are independent and the support of T is shorter than the support of C, it can be shown that

$$\begin{aligned} E\left(\frac{V}{G(X)}\right) &= E\left(\tilde{V}E\left(\frac{I(T \leq C)}{G(T)}\middle| T\right)\right) \\ &= E\left(\tilde{V}\right), \end{aligned} \tag{3.4}$$

which implies that by taking an inverse-probability-weighting adjustment on V, we can obtain an unbiased proxy of $\tilde{V}$. This idea has been applied to sample surveys. Specifically, if we know the sampling scheme well, we can correct the sampling bias and make valid inference about the true population.

For the following discussions, we refer to $\hat{G}(t)$ as the Kaplan-Meier estimator of $G(t)$. Many well-known nonparametric estimators in survival analysis can be re-expressed in terms of weighted averages. For example, the Kaplan-Meier estimator of the survival function of T, $\hat{S}^{KM}(t)$, can be expressed as

$$\hat{S}^{KM}(t) = 1 - \frac{1}{n}\sum_{i=1}^{n}\frac{I(X_i \leq t)\delta_i}{\hat{G}(X_i)}. \tag{3.5}$$

For the competing-risks analysis, nonparametric maximum likelihood estimators of cumulative incidence functions can be written as the following explicit expression:

$$\hat{F}_j^{NPMLE}(t) = \frac{1}{n}\sum_{i=1}^{n}\frac{I(X_i \leq t, B_i = j)}{\hat{G}(X_i)}. \tag{3.6}$$

The technique of IPCW also plays a useful role in the regression analysis for survival data. Consider the following linear transformation model

$$m(T) = \mathbf{Z}^T\gamma + \epsilon,$$

where $m(\cdot)$ is an unknown strictly increasing function and ϵ has a completely known distribution function F_ϵ. Under independent censoring and the assumption that the censoring distribution is independent of $\mathbf{Z}$, it can be shown that

$$\begin{aligned} E\left(\frac{\delta_j I(X_i \geq X_j)}{G^2(X_j)}\middle|\mathbf{Z}\right) &= E\left(I(T_i \geq T_j)\ E\left(\frac{\delta_j I(C_i \geq T_j)}{G^2(T_j)}\middle| T_i, T_j\right)\middle|\mathbf{Z}\right) \\ &= \Pr(T_i \geq T_j|\mathbf{Z}) \quad = \quad \Pr(\epsilon_i - \epsilon_j \geq \mathbf{Z}_{ij}^T\gamma) \end{aligned}$$

if $G(t) > 0$ for any t located within the support of T, where $\mathbf{Z}_{ij} = \mathbf{Z}_j - \mathbf{Z}_i$. Based on the equation mentioned above, Cheng, Wei and Ying (1995) proposed the following estimating function of γ

$$U(\gamma) = \sum_{i=1}^{n} \sum_{j \neq i} \left[\frac{\delta_j I(X_i \geq X_j)}{\hat{G}^2(X_j)} - \eta(\mathbf{Z}_{ij}^T \gamma) \right] w(\mathbf{Z}_{ij}^T \gamma) \mathbf{Z}_{ij}, \tag{3.7}$$

where $w(\cdot)$ is a positive weight function and $\eta(\mathbf{Z}_{ij}^T \gamma) = \Pr(\epsilon_i - \epsilon_j \geq \mathbf{Z}_{ij}^T \gamma)$. Jung (1996) considered the regression analysis for the long-term survival probability in which the word "long-term" refers to a patient having successfully survived over the specified interval. He constructed an estimating function of the regression parameters by utilizing the technique of IPCW. Chen *et al.* (2005) analyzed the mean residual life model in a similar way.

Despite that the technique of IPCW is a convenient tool and easy to be understood, it has some drawbacks. The first is that this method highly depends on a consistent estimator of G. When the censoring time depends on continuous covariates, a consistent estimator of G is not easy to be obtained and usually requires using smoothing techniques or making additional model assumptions. Another crucial point is about the support condition. Let τ_C denote the endpoint of the support of C. Then we have

$$E\left(\frac{\delta}{G(X)} \middle| T \right) = \begin{cases} E\left(I(T \leq C)/G(T) \middle| T \right) = 1 & \text{if } T < \tau_C, \\ 0 & \text{otherwise,} \end{cases}$$

and accordingly

$$E\left(\frac{V}{G(X)} \right) = E\left(I(T < \tau_C) \tilde{V} \right) \leq E(\tilde{V}).$$

That is, when $\Pr(T \geq \tau_C) > 0$ which results in a heavier tail of $\hat{S}^{KM}$, the quantity $V/G(X)$ is no longer an unbiased proxy for $\tilde{V}$ and consequently any estimating function constructed based on equation (3.4), such as equation (3.7), may not lead to a consistent estimator of γ.

To overcome this problem, one may consider the following related equation

$$E\left(\frac{I(X \leq \tau) V}{G(X)} \right) = E(I(T \leq \tau) \tilde{V}),$$

where τ is pre-specified such that $\Pr(X > \tau) > 0$. Fine, Ying and Wei (1998) suggested to modify the estimating function (3.7) as

$$U(\gamma) = \sum_{i=1}^{n} \sum_{j \neq i} \left[\frac{\delta_j I(X_i \wedge \tau \geq X_j)}{\hat{G}^2(X_j)} - \eta_\alpha(\mathbf{Z}_{ij}^T \gamma) \right] w(\mathbf{Z}_{ij}^T \gamma) \mathbf{Z}_{ij},$$

where $\alpha = m(\tau)$ and $\eta_\alpha(\mathbf{Z}_{ij}^T \gamma) = \int_{-\infty}^{\alpha} [1 - F_\epsilon(t - \mathbf{Z}_i^T \gamma)] dF_\epsilon(t - \mathbf{Z}_j^T \gamma)$. Such a modification is no longer subject to the support problem.

3.3 Imputation

Another popular method for handling missing data is by imputation. Specifically, censored (or missing) variables can be imputed by their conditional means given the observed data and the statistical analysis can be proceeded based on the imputed values. This approach has been widely used in the analysis of missing or censored data. For example, Buckley and James (1979) considered the following linear regression model,

$$T = \mathbf{Z}^T \gamma + \epsilon,$$

where T may be the transformed failure time. When the value of T is not observed, it can be imputed by an estimator of

$$\begin{aligned} E(T|X, \delta, \mathbf{Z}) &= \delta X + (1 - \delta) \left[\mathbf{Z}^T \gamma + E(\epsilon | \epsilon > X - \mathbf{Z}^T \gamma, X, \mathbf{Z}) \right] \\ &= \delta X + (1 - \delta) \left[\mathbf{Z}^T \gamma - \int_{X - \mathbf{Z}^T \gamma}^{\infty} u \, dS_\epsilon(u) \Big/ S_\epsilon(X - \mathbf{Z}^T \gamma) \right], \end{aligned}$$

where S_ϵ is the survival function of the error term ϵ. To estimate γ, Buckley and James (1979) proposed a self-consistency approach in which the estimators $\hat{\gamma}$ and $\hat{S}_\epsilon$ are updated iteratively until the convergence criterion is obtained. Li, Wang and Chen (1999) applied the sliced inverse regression (SIR) to analyze right censored data in which the sliced mean of a specified interval $[t_j, t_{j+1})$, denoted as $E[\mathbf{Z} I(t_j \leq T < t_{j+1})]$, was estimated by utilizing the following relationship

$$E[\mathbf{Z} I(T \geq t) | X, \delta, \mathbf{Z}] = \mathbf{Z} I(X \geq t) + (1 - \delta) \mathbf{Z} I(X < t) \frac{\Pr(T \geq t | \mathbf{Z})}{\Pr(T > X | \mathbf{Z})}.$$

In the regression analysis for the long-term survival probability at a given time point, Jung (1996) suggested to use the IPCW approach while Subramanian (2001) proposed the imputation approach. It implies that these two methods may be applied to solving the same inference problem.

Wang (2003) considered a nonparametric setting under the framework of a two-path model. To illustrate, consider a study of bone marrow transplants for leukaemia patients in which some patients will experience recurrence of the malignancy before death but others may die without relapse. Let T_1 and T_2 be the times to recurrence, an intermediate state, and to death, a terminal endpoint, respectively. Then $I(T_1 \leq T_2)$ denotes the path indicator in which 1 refers to the path of recurrence and 0 refers to the path of death without relapse. In presence of right censoring, observed data can be expressed as $(X_1, X_2, \delta_1, \delta_2)$, where $X_1 = T_1 \wedge T_2 \wedge C$, $X_2 = T_2 \wedge C$, $\delta_1 = I(T_1 \leq T_2 \wedge C)$ and $\delta_2 = I(T_2 \leq C)$. Under the assumption that (T_1, T_2) and C are independent, Wang (2003) proposed a nonparametric procedure for estimating the path probability $\Pr(T_1 \leq T_2)$ which imputes $I(T_1 \leq T_2)$ by an estimator of

$$E[I(T_1 \leq T_2)|X_1, X_2, \delta_1, \delta_2] = I(\delta_1 = 1) + I(\delta_1 = \delta_2 = 0)p(C),$$

where $p(C)$ is the conditional path probability which can be further expressed as

$$\begin{aligned} p(x) &= \Pr(T_1 \leq T_2 | T_1 \wedge T_2 > x), \\ &= \frac{1}{D(x)} \left[\Pr(x < T_1 \leq \tau_C, T_1 \leq T_2) + D(\tau_C) \Pr(T_1 \leq T_2 | T_1 \wedge T_2 > \tau_C)\right] \\ &= \frac{1}{D(x)} \left[\int_x^\infty \frac{\Pr(X_1 \in [v, v+dv), \delta_1 = 1)}{G(v)} + D(\tau_C) p(\tau_C)\right] \end{aligned} \tag{3.8}$$

with $D(x) = \Pr(T_1 \wedge T_2 > x)$ and τ_C being the endpoint of the support of censoring time C. All the components in equation (3.8) are estimable nonparametrically except for the last term $p(\tau_C)$ which is not identifiable if $\tau_C < \tau_{T_1}$. To overcome this problem, Wang (2003) made an additional assumption that $p(\tau_C) = \Pr(T_1 \leq T_2)$ and then derived an explicit estimator of the path probability.

Chapter 4

The Proposed Approach

4.1 Model Assumption

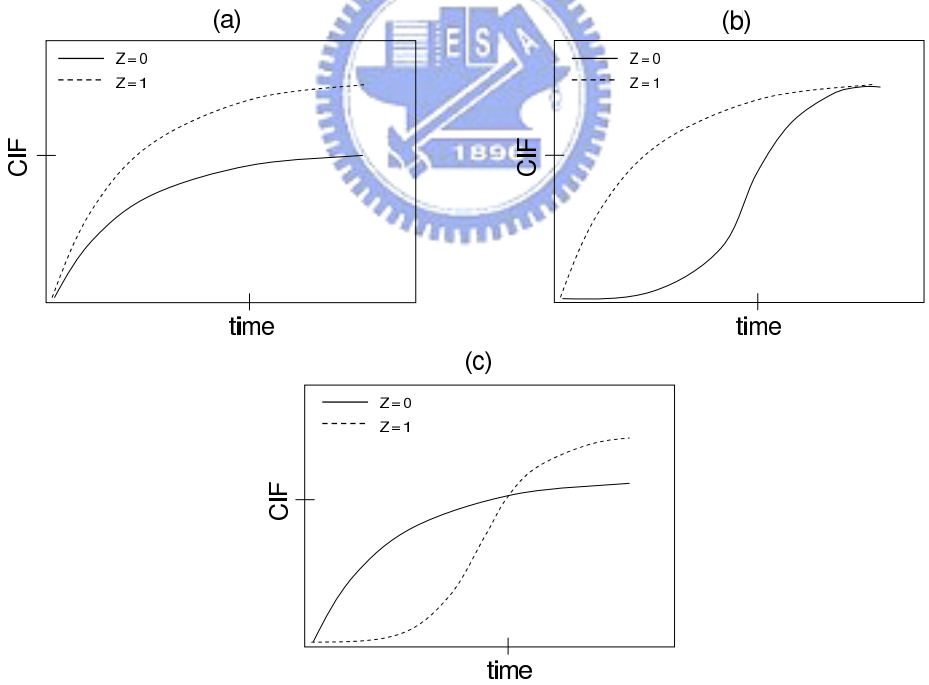
In this proposal, we consider model (2.6) in which the marginal failure probability is the focus and the latency distribution is of less interest and hence not specified. Without loss of generality, we consider only two causes of failures, namely $\tilde{B} = 1$ or 2. Suppose that failure of the first type is of main interest. Specifically, we consider the following regression formulation:

$$\begin{aligned} F_1(t \wedge \tau|\mathbf{Z}) &= \Pr(T \leq \tau, \tilde{B} = 1|\mathbf{Z})\Pr(T \leq t|T \leq \tau, \tilde{B} = 1, \mathbf{Z}) \\ &= \pi(\tilde{\mathbf{Z}}^T\boldsymbol{\beta}(\tau))[1 - Q_{1,\mathbf{Z}}(t|\tau)], \end{aligned} \tag{4.1}$$

where $\tilde{\mathbf{Z}} = (1, \mathbf{Z}^T)^T$ is the $(p+1)\times 1$ vector of covariates, $\pi(\cdot)$ is a known function mapping from $(-\infty, \infty)$ to $(0, 1)$, $\boldsymbol{\beta}(\tau)$ is a $(p + 1) \times 1$ vector of parameters and τ lies within the data support such that $\Pr(T \wedge C > \tau|\mathbf{Z}) > 0$. The main objective is to estimate $\boldsymbol{\beta}(\tau)$ which measures the covariate effect on the cumulative probability of incidence by time τ. The severe acute respiratory syndrome (SARS) provides an example for the motivation. SARS is an epidemic and life-threatening acute disease that resulted in a global outbreak in 2003. Clinicians and the pubic were most concerned with finding out which characteristics of a patient would affect his/her probability of being discharged from the hospital and alive by a target time point.

Comparing the regression formulation in (2.2) with that in (4.1), we find that setting $t = \tau$ and $H(u) = \pi^{-1}(u)$, model (2.2) coincides with model (4.1) and $\boldsymbol{\beta}(\tau) = [m(\tau),\ \boldsymbol{\theta}^T]^T$. In other words, model (4.1) fits the data at a single time point τ while model (2.2) considers modeling the entire time span. If model (2.2) is appropriate, then the last p components of $\boldsymbol{\beta}(\tau)$ derived from model (4.1) will be similar for different choices of τ. Therefore results obtained from model (4.1) can be used to verify the assumption of model (2.2) or help choosing time-dependent covariates in that model. Figure 4.1 provides a graphical illustration to highlight the difference of the two models with a binary covariate. In Figures 4.1(b) and 4.1(c), $F_1(t|0)$ and $F_1(t|1)$ have a crossing point which obviously violates model (2.2). Model (4.1) can include all the three situations. Therefore the dependency of $\boldsymbol{\beta}(\tau)$ on τ is not a subjective restriction but provides the flexibility to detect possible change of covariate effect on the cumulative incidence probability at different time points.

Figure 4.1: Illustration of the cumulative incidence function, $F_1(t|Z)$, for a binary Z.



4.2 Proposed Inference Methods

4.2.1 Preliminary

Without loss of generality, we consider only two causes of failures, namely $\tilde{B} = 1$ or 2. Suppose that failure of the first type is of main interest. Denote $\{(T_i, \tilde{B}_i, \tilde{\mathbf{Z}}_i) : i = 1, \ldots, n\}$ as a random sample of $(T, \tilde{B}, \tilde{\mathbf{Z}})$. Let $\Delta_{ji} = I(T_i \leq \tau, \tilde{B}_i = j)$ for $i = 1, \ldots, n$ and $j = 1, 2$. Under model (4.1), we have $E(\Delta_{1i}|\tilde{\mathbf{Z}}_i) = \pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})$, where $\boldsymbol{\beta} = \boldsymbol{\beta}(\tau)$ is the parameter of interest. With the complete data, the likelihood function of $\boldsymbol{\beta}$ is given by

$$\tilde{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\right]^{\Delta_{1i}} \left[\bar{\pi}(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\right]^{1-\Delta_{1i}}, \tag{4.2}$$

where $\bar{\pi}(t) = 1 - \pi(t)$, and the resulting score function becomes

$$\tilde{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[\Delta_{1i} - \pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\right] \frac{\pi_\phi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})}{\pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\bar{\pi}(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})}\tilde{\mathbf{Z}}_i, \tag{4.3}$$

where $\pi_\phi(t) = \partial\pi(t)/\partial t$.

Under right censoring, observed variables can be written as $\{(X_i, B_i, \tilde{\mathbf{Z}}_i) : i = 1, \ldots, n\}$, which are *i.i.d.* replications of $(X, B, \tilde{\mathbf{Z}})$, where $B = \tilde{B} \cdot I(T \leq C)$ and $X = T \wedge C$. Note that the value of Δ_{1i} may be unknown due to censoring. It turns out that the likelihood function of $\boldsymbol{\beta}$ becomes very complicated and involves specification of several nuisance functions such as $\Pr(T > t|\tilde{B} = j, T \leq \tau, \mathbf{Z})$ and $\Pr(T > t|\tilde{B} = j, T > \tau, \mathbf{Z})$ for $j = 1, 2$. We propose to directly modify the score function $\tilde{U}(\boldsymbol{\beta})$ by applying two useful principles to handling missing data.

The first approach utilizes observable proxies of Δ_{1i} by applying the technique of the IPCW to adjusting for their biases. The second proposal adopts the imputation approach which imputes Δ_{1i} by an estimator of $E(\Delta_{1i}|X_i, B_i, \mathbf{Z}_i)$. Both methods are popular and useful tools for handling missing data in statistical literature. We assume that, given $\mathbf{Z}$, C is independent of $(T, \tilde{B})$. To simplify the analysis, T and C are both continuous variables.

4.2.2 Inverse Probability Weighting

Assume temporarily that the distribution of C does not depend on $\mathbf{Z}$. That is, $\Pr(C \geq t|\mathbf{Z}) = \Pr(C \geq t) = G(t)$. We will discuss possible modifications when this assumption does not hold. In the presence of right censoring, we can find observable proxies of Δ_{1i} and then apply the technique of IPCW to correcting their biases. Specifically, it follows that

$$\begin{aligned} E\left(\frac{I(X \leq \tau, B=1)}{G(X)}\middle|\mathbf{Z}\right) &= E\left(I(T \leq \tau, \tilde{B}=1)E\left(\frac{I(T \leq C)}{G(T)}\middle|T, \tilde{B}, \mathbf{Z}\right)\middle|\mathbf{Z}\right) \\ &= E\left(I(T \leq \tau, \tilde{B}=1)\middle|\mathbf{Z}\right) = \pi(\tilde{\mathbf{Z}}^T\boldsymbol{\beta}) \end{aligned} \tag{4.4}$$

and

$$E\left(\frac{I(X > \tau)}{G(\tau+)} + \frac{I(X \leq \tau, B=2)}{G(X)}\middle|\mathbf{Z}\right) = 1 - \pi(\tilde{\mathbf{Z}}^T\boldsymbol{\beta}) = \bar{\pi}(\tilde{\mathbf{Z}}^T\boldsymbol{\beta}), \tag{4.5}$$

where $G(\tau+) = \Pr(C > \tau)$. These two moment conditions can be utilized to construct estimating functions of $\boldsymbol{\beta}$. Set

$$H_{1i} = \frac{I(X_i \leq \tau, B_i = 1)}{G(X_i)} - \pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta}) \tag{4.6}$$

and

$$H_{2i} = \frac{I(X_i > \tau)}{G(\tau+)} + \frac{I(X_i \leq \tau, B_i = 2)}{G(X_i)} - \bar{\pi}(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta}) \tag{4.7}$$

for $i = 1, \ldots, n$. Replacing $G(t)$ with the Kaplan-Meier estimator

$$\hat{G}(t) = \prod_{u<t}\left[1 - \frac{\sum_{k=1}^n I(X_k = u, B_k = 0)}{\sum_{k=1}^n I(X_k \geq u)}\right], \tag{4.8}$$

the resulting estimating functions become

$$U_{\mathrm{w1}}(\boldsymbol{\beta}) = \sum_{i=1}^n \hat{H}_{1i}\frac{\pi_\phi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})}{V_{1i}}\tilde{\mathbf{Z}}_i \tag{4.9}$$

and

$$U_{\mathrm{w2}}(\boldsymbol{\beta}) = \sum_{i=1}^n \hat{H}_{2i}\frac{\pi_\phi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})}{V_{2i}}\tilde{\mathbf{Z}}_i, \tag{4.10}$$

where $\hat{H}_{ji}$ are H_{ji} $(j = 1, 2)$ with G being replaced by $\hat{G}$ and V_{ji} is a weight function, which measures the variation of $\hat{H}_{ji}$. A natural choice for V_{ji} is $Var(H_{ji}|\mathbf{Z}_i)$. Specifically,

$$Var(H_{1i}|\mathbf{Z}_i) = E\left(\frac{I(X_i \le \tau, B_i = 1)}{G^2(X_i)}\middle|\mathbf{Z}_i\right) - \pi^2(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta}), \tag{4.11}$$

which however involves some unknown quantity which does not have an analytic expression. Based on the first-order Taylor expansion, the first term in (4.11) can be approximated by

$$E\left(\frac{1}{G(X_i)}\middle|\mathbf{Z}_i\right) E\left(\frac{I(X_i \le \tau, B_i = 1)}{G(X_i)}\middle|\mathbf{Z}_i\right) \approx E\left(\frac{1}{G(X_i)}\right)\pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta}).$$

Although $E\,(1/G(X))$ can be estimated by its moment estimator, this quantity is too sensitive to the tail behavior of $\hat{G}$ which may be unstable. Hence we suggest using a related but more robust quantity instead such as the sample median of $\{1/\hat{G}(X_i) : i = 1, \ldots, n\}$, denoted as M_G. Accordingly, we suggest to set $V_{1i} = \pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})(M_G - \pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta}))$, and by the same argument, we choose $V_{2i} = \bar{\pi}(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})(M_G - \bar{\pi}(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta}))$.

The two estimating functions mentioned above may be combined by utilizing the method of constructing the optimal estimating function discussed in Heyde (1997, Chapter 2). Let $\mathbf{H}_i = (H_{1i}, H_{2i})^T$ for $i = 1, \ldots, n$, the optimal estimating function of $\boldsymbol{\beta}$ based on $\mathbf{H} = (\mathbf{H}_1^T, \ldots, \mathbf{H}_n^T)^T$ is given by

$$E\left(-\frac{\partial \mathbf{H}^T}{\partial \boldsymbol{\beta}}\middle|\mathbf{Z}\right)\Sigma_{\mathbf{H}}^{-1}\,\mathbf{H} = \sum_{i=1}^{n} E\left(-\frac{\partial \mathbf{H}_i^T}{\partial \boldsymbol{\beta}}\middle|\mathbf{Z}_i\right)\Sigma_{\mathbf{H}_i}^{-1}\,\mathbf{H}_i, \tag{4.12}$$

where $\Sigma_{\mathbf{H}} = E(\mathbf{H}\mathbf{H}^T|\mathbf{Z})$ and

$$\Sigma_{\mathbf{H}_i} = E(\mathbf{H}_i\mathbf{H}_i^T|\mathbf{Z}_i) = \begin{bmatrix} Var(H_{1i}|\mathbf{Z}_i) & -\pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\bar{\pi}(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta}) \\ -\pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\bar{\pi}(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta}) & Var(H_{2i}|\mathbf{Z}_i) \end{bmatrix}.$$

Replacing $Var(H_{ji}|\mathbf{Z}_i)$ by V_{ji} whose forms have been suggested earlier, we obtain the following estimating function

$$U_{\mathrm{w}^*}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left[(V_{2i} - V_{3i})\hat{H}_{1i} - (V_{1i} - V_{3i})\hat{H}_{2i}\right]\frac{\pi_\phi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})}{V_{1i}V_{2i} - V_{3i}^2}\tilde{\mathbf{Z}}_i, \tag{4.13}$$

where $V_{3i} = \pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\bar{\pi}(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})$.

The three estimating functions in equations (4.9), (4.10) and (4.13) all reduce to $\tilde{U}(\boldsymbol{\beta})$ in absence of right censoring. With censored data, it is reasonable to suspect that $U_{\mathrm{w}^*}(\boldsymbol{\beta})$ is the most efficient one since it utilizes more information. It is interesting to note that the estimating function proposed by Fine (1999) actually has the form of $U_{\mathrm{w}1}(\beta)$ with a different weight $V_{1i} = \pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\bar{\pi}(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})$ which does not account for the effect of censoring. Via the simulations, we will see how these different weight assignments affect the resulting estimators of $\boldsymbol{\beta}$.

Denote the solution to $U_{\mathrm{w}^*}(\boldsymbol{\beta}) = 0_{(p+1)\times 1}$ as $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{w}j}$ as the solution to $U_{\mathrm{w}j}(\boldsymbol{\beta}) = 0_{(p+1)\times 1}$ for $j = 1, 2$. In Appendices A and B, we prove the asymptotic normality of $U_{\mathrm{w}^*}(\boldsymbol{\beta}_0)$ and $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$, where $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$. Note that an asymptotic expression of $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{w}^*} - \boldsymbol{\beta}_0)$ is obtained as

$$n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{w}^*} - \boldsymbol{\beta}_0) = [A_{\mathrm{w}^*}(\boldsymbol{\beta}_0)]^{-1} n^{-1/2} U_{\mathrm{w}^*}(\boldsymbol{\beta}_0) + o_p(1),$$

where

$$A_{\mathrm{w}^*}(\boldsymbol{\beta}_0) = -\lim_{n\to\infty} \frac{1}{n} \frac{\partial U_{\mathrm{w}^*}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}.$$

Hence $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{w}^*} - \boldsymbol{\beta}_0)$ has an asymptotically normal distribution with mean $0_{(p+1)\times 1}$ and covariance matrix

$$V_{\mathrm{w}^*} = [A_{\mathrm{w}^*}(\boldsymbol{\beta}_0)]^{-1} \Gamma_{\mathrm{w}^*} [A_{\mathrm{w}^*}(\boldsymbol{\beta}_0)]^{-1}, \tag{4.14}$$

where Γ_{w^*} is the asymptotic covariance matrix of $n^{-1/2}U_{\mathrm{w}^*}(\boldsymbol{\beta}_0)$.

If the censoring variable C depends on discrete covariates, the Kaplan-Meier estimator $\hat{G}(t)$ can be evaluated for each covariate group. If the related covariate is continuous, we suggest two different ways of modification. In Section 4.2.3, we illustrate using a nonparametric smoothing technique, namely the kernel method, to estimate $\Pr(C \geq t|\tilde{\mathbf{Z}} = z)$. The other approach, which can avoid the curse of dimensionality, is to impose some parametric or semi-parametric model which describes the covariate effect on C.

4.2.3 Imputation by Conditional Mean

Alternatively, to handle possible incompleteness of $\Delta_1 = I(T \leq \tau, \tilde{B} = 1)$ due to censoring, one may impute its value by an estimate of the conditional mean given the data. Specifically, $E(\Delta_1|X, B, \mathbf{Z})$ equals

$$I(X \leq \tau, B = 1) + I(X \leq \tau, B = 0)p_{\mathbf{Z}}(X),$$

where $p_z(x) = \Pr(T \leq \tau, \tilde{B} = 1|T > x, \mathbf{Z} = z)$. Two estimators of $p_z(x)$, denoted as $\hat{p}_z^{(j)}(x)$ for $j = 1, 2$, will be proposed and their formulae will be derived later. Replacing Δ_{1i} by

$$\hat{\Delta}_{1i}^{(j)} = I(X_i \leq \tau, B_i = 1) + I(X_i \leq \tau, B_i = 0)\hat{p}_{\mathbf{Z}_i}^{(j)}(X_i),$$

in the score function (4.3), we can obtain the following estimating functions of $\boldsymbol{\beta}$

$$U_{Ij}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[\hat{\Delta}_{1i}^{(j)} - \pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\right] \frac{\pi_\phi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})}{\pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\bar{\pi}(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})}\tilde{\mathbf{Z}}_i \tag{4.15}$$

for $j = 1, 2$ depending on which $\hat{p}_z^{(j)}(x)$ will be used in.

The first proposed estimator $\hat{p}_z^{(1)}(x)$, is derived under a purely nonparametric setting which generalizes the nonparametric results in Wang (2003) and Satten and Datta (2001). Their ideas are roughly summarized in Appendix C. With covariates, it follows that

$$p_z(x) = \Pr(T \leq \tau, \tilde{B} = 1|T > x, \mathbf{Z} = z) = \frac{\Pr(x < T \leq \tau, \tilde{B} = 1|\mathbf{Z} = z)}{S_z(x)}, \tag{4.16}$$

where $S_z(t) = \Pr(T > t|\mathbf{Z} = z)$. When $\mathbf{Z}$ takes only discrete values, a model-free estimator of $p_z(x)$ is given by

$$\hat{p}_z^{(1)}(x) = \frac{1}{\hat{S}_z(x)} \frac{\sum_{i=1}^n I(x < X_i \leq \tau, B_i = 1, \mathbf{Z}_i = z)/\hat{G}(X_i)}{\sum_{i=1}^n I(\mathbf{Z}_i = z)}, \tag{4.17}$$

where $\hat{G}$ is obtained in equation (4.8) and

$$\hat{S}_z(t) = \prod_{u \leq t} \left[1 - \frac{\sum_{i=1}^n I(X_i = u, B_i \neq 0, \mathbf{Z}_i = z)}{\sum_{i=1}^n I(X_i \geq u, \mathbf{Z}_i = z)}\right]. \tag{4.18}$$

A nonparametric way of handling continuous $\mathbf{Z}$ is to utilize some smoothing technique. Using the idea of Dabrowska (1987) in the nonparametric regression, we obtain

$$\hat{p}_z^{(1)}(x) = \frac{1}{\sum_{i=1}^n \left[\frac{I(X_i>x,B_i=1)}{\hat{G}(X_i)} + \frac{I(X_i>X_{(m)})}{\hat{G}(X_{(m)}+)}\right] B_{n,i}(z)} \sum_{i=1}^n \frac{I(x < X_i \le \tau, B_i = 1)}{\hat{G}(X_i)} B_{n,i}(z), \tag{4.19}$$

where $X_{(m)}$ is the largest observed failure time and $B_{n,i}(z)$ is a random set of non-negative weights. Candidates of $B_{n,i}(z)$ include kernel-type weights, nearest neighbors or local linear weights. For example, one can use the kernel-type weight

$$B_{n,i}(z) = \frac{K(a_n^{-1}(z - \mathbf{Z}_i))}{\sum_{\ell=1}^n K(a_n^{-1}(z - \mathbf{Z}_\ell))}, \tag{4.20}$$

where $K(\cdot)$ is an appropriate kernel function and a_n is the bandwidth.

The second proposed estimator, $\hat{p}_z^{(2)}(x)$, utilizes the model assumption in equation (4.1). Specifically, equation (4.16) can be expressed as $p_z(x; \boldsymbol{\beta}, Q_{1,z}(\cdot|\tau), Q_{2,z}(\cdot|\tau), S_z(\tau))$ which equals

$$\frac{Q_{1,z}(x|\tau)\pi(\tilde{z}^T\boldsymbol{\beta})}{Q_{1,z}(x|\tau)\pi(\tilde{z}^T\boldsymbol{\beta}) + Q_{2,z}(x|\tau)\{1 - S_z(\tau) - \pi(\tilde{z}^T\boldsymbol{\beta})\} + S_z(\tau)}, \tag{4.21}$$

where $\tilde{z} = (1, z^T)^T$ and $Q_{j,z}(t|\tau) = \Pr(T > t | T \le \tau, \tilde{B} = j, \mathbf{Z} = z)$ for $j = 1, 2$.

The previous equation still involves nuisance functions, namely $Q_{j,z}(t|\tau)$ for $j = 1, 2$ and $S_z(\tau)$. Here we suggest to estimate these quantities in a nonparametric way. To simplify the presentation, we give the formula which includes both types of covariates by setting $B_{n,i}(z) = I(\mathbf{Z}_i = z)$ for discrete covariates and the formula in equation (4.20) for continuous covariates. The proposed estimator $\hat{Q}_{1,z}(t|\tau)$ can be written as

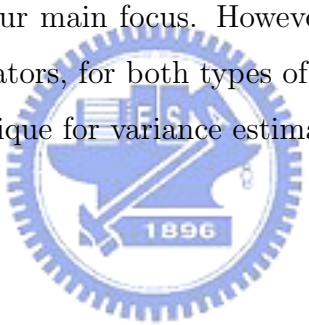
$$\prod_{u \le t} \left\{ 1 - \frac{\sum_{i=1}^n I(u = X_i \le \tau, B_i = 1) B_{n,i}(z)}{\sum_{i=1}^n [I(u \le X_i \le \tau, B_i = 1) + I(u \le X_i \le \tau, B_i = 0)\hat{p}_z^{(1)}(X_i)] B_{n,i}(z)} \right\}, \tag{4.22}$$

where the formula of $\hat{p}_z^{(1)}(x)$ is given in (4.17) or (4.19) for discrete and continuous $\mathbf{Z}$, respectively. The estimator of $Q_{2,z}(t|\tau)$, denoted as $\hat{Q}_{2,z}(t|\tau)$, has a similar expression as $\hat{Q}_{1,z}(t|\tau)$ with $B_i = 1$ being replaced by $B_i = 2$ in the corresponding formula. The proposed estimator of $S_z(\tau)$ is given by

$$\prod_{u \le \tau} \left\{ 1 - \frac{\sum_{i=1}^n I(X_i = u, B_i \ne 0) B_{n,i}(z)}{\sum_{i=1}^n I(X_i \ge u) B_{n,i}(z)} \right\}. \tag{4.23}$$

The solution to $U_{Ij}(\boldsymbol{\beta}) = 0_{(p+1)\times 1}$ is denoted as $\hat{\boldsymbol{\beta}}_{Ij}$ for $j = 1, 2$. These two estimating functions differ in the way of estimating $p_z(x)$. Via simulations, we will examine whether the second proposal which utilizes the model information has better performance. Since $U_{I2}(\boldsymbol{\beta})$ is a more complicated function of $\boldsymbol{\beta}$, to simplify the root-finding procedure, one may treat $\hat{\Delta}_{1i}^{(2)}$ as a fixed number in the mth iteration by using $p_z(x; \hat{\boldsymbol{\beta}}^{(m-1)}, \hat{Q}_{1,z}(\cdot|\tau), \hat{Q}_{2,z}(\cdot|\tau), \hat{S}_z(\tau))$ instead, where $\hat{\boldsymbol{\beta}}^{(m-1)}$ is the solution in the previous step. The final solution is obtained via an iterative procedure with $m = 1, 2, \ldots$, etc. The modified equation is a simpler function of $\boldsymbol{\beta}$ and thus convergence can be achieved by only few steps of iterations.

In Appendix D, we prove the asymptotic normality of $n^{-1/2}U_{I1}(\boldsymbol{\beta}_0)$ and that of $n^{1/2}(\hat{\boldsymbol{\beta}}_{I1} - \boldsymbol{\beta}_0)$ when $\mathbf{Z}$ is discrete. Similar arguments can be applied to establishing asymptotic properties of $n^{-1/2}U_{I2}(\boldsymbol{\beta}_0)$ and $n^{1/2}(\hat{\boldsymbol{\beta}}_{I2} - \boldsymbol{\beta}_0)$. For continuous covariates, asymptotic analysis is not provided since the method involves kernel smoothing which is a technical issue and not of our main focus. However, due to the complexity of the plugged-in nonparametric estimators, for both types of covariates, we suggest to utilize the bootstrap re-sampling technique for variance estimation.



Chapter 5

Numerical Studies

5.1 Simulation Analysis

Finite-sample performances of the proposed estimators were evaluated via simulations. The covariate Z was generated from three distributions. For the discrete case, we set $Z \sim$ Bernoulli(0.5). For the continuous case, we set $Z \sim$ Normal$(0, 1)$ or $Z \sim$ Unif$(-3, 3)$. Let $\Delta_j = I(T \leq \tau, \tilde{B} = j)$ for $j = 1, 2$. Given Z, we set $\Delta_1 \sim$Bernoulli$(\pi(\beta_0 + \beta_1 Z))$ with

$$\pi(\beta_0 + \beta_1 Z) = \frac{\exp(\beta_0 + \beta_1 Z)}{1 + \exp(\beta_0 + \beta_1 Z)}.$$

If $\Delta_1 = 1$, then $\Delta_2 = 0$; and if $\Delta_1 = 0$, Δ_2 is generated from a Bernoulli(p_2), where p_2 may depend on Z but its form is not of interest. Given (Δ_1, Δ_2), the failure time T is generated from a distribution with density function f_T which can be expressed as

$$f_T(t) = \begin{cases} f_1(t|\tau, Z) & \text{if} \quad (\Delta_1, \Delta_2) = (1, 0) \\ f_2(t|\tau, Z) & \text{if} \quad (\Delta_1, \Delta_2) = (0, 1) \\ f_3(t|\tau, Z) & \text{if} \quad (\Delta_1, \Delta_2) = (0, 0), \end{cases}$$

where $f_j(t|\tau, Z)$, $j = 1, 2$, are density functions with supports no greater than τ and $f_3(t|\tau, Z)$ is a density function whose value exceeds τ. In the simulations, we set

$$f_j(t|\tau, Z) = \frac{f_{Y_j}(t|Z)}{1 - S_{Y_j}(\tau|Z)} I(t \leq \tau) \quad \text{for } j = 1, 2 \quad \text{and} \quad f_3(t|\tau, Z) = \frac{f_{Y_3}(t|Z)}{S_{Y_3}(\tau|Z)} I(t > \tau),$$

where $f_{Y_j}(t|Z)$ and $S_{Y_j}(t|Z)$ are the density and survival functions of Y_j which follows the accelerated failure-time model. Specifically, we set

$$\log Y_j = \gamma_{0,j} + \gamma_{1,j} Z + \sigma_j \cdot W_j, \tag{5.1}$$

where $\gamma_{0,j}$, $\gamma_{1,j}$ and σ_j are (nuisance) parameters and W_j has the pre-specified error distribution.

The censoring variable is generated from Unif$(c_0, c_0 + c_1)$, where (c_0, c_1) are pre-specified constants making the censoring proportion to achieve the target value (i.e., 30% or 40%). Denoted $\{(\Delta_{1i}, \Delta_{2i}, T_i, Z_i, C_i) : i = 1, \ldots, n\}$ as a random sample of $(\Delta_1, \Delta_2, T, Z, C)$. Note that

$$I(X_i \le \tau, B_i = j) = \Delta_{ji} \cdot I(T_i \le C_i)$$

for $j = 1, 2$, where $X_i = T_i \wedge C_i$. The proposed methods can be implemented based on

$$\{X_i, I(T_i \le C_i), I(X_i \le \tau, B_i = 1), I(X_i \le \tau, B_i = 2)\}$$

for $i = 1, \ldots, n$. The value of τ is set to be 2.5. The sample size n was set to be 100 or 300.

The parameters of interest are $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$. Besides the three proposed estimators $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*} = (\hat{\beta}_{\mathrm{w}^*,0}, \hat{\beta}_{\mathrm{w}^*,1})^T$, $\hat{\boldsymbol{\beta}}_{I1} = (\hat{\beta}_{I1,0}, \hat{\beta}_{I1,1})^T$ and $\hat{\boldsymbol{\beta}}_{I2} = (\hat{\beta}_{I2,0}, \hat{\beta}_{I2,1})^T$ and, for comparison, we also evaluated the estimator proposed by Fine (1999), denoted as $\hat{\boldsymbol{\beta}}_F = (\hat{\beta}_{F,0}, \hat{\beta}_{F,1})^T$, which solves $U_F(\boldsymbol{\beta}) = 0_{2\times 1}$. Recall that $U_F(\boldsymbol{\beta})$ has the form of $U_{\mathrm{w}1}(\boldsymbol{\beta})$ with $V_{1i} = \pi(\tilde{\mathbf{Z}}_i^T \boldsymbol{\beta})\bar{\pi}(\tilde{\mathbf{Z}}_i^T \boldsymbol{\beta})$. Based on 1000 replications, we report the average bias (BS), the standard error (SE) and the mean squared errors (MSE) for $\hat{\beta}_{\mathrm{w}^*,i}$, $\hat{\beta}_{I1,i}$, $\hat{\beta}_{I2,i}$ and $\hat{\beta}_{F,i}$ for $i = 0, 1$ respectively. We also report the relative efficiency (RE) which is defined as the ratio of the mean square errors of $\hat{\beta}_{F,i}$ to that for the other three estimators, $\hat{\beta}_{\mathrm{w}^*,i}$, $\hat{\beta}_{I1,i}$ and $\hat{\beta}_{I2,i}$ for $i = 0, 1$. The criteria mentioned above are used to assess the performance of different point estimators of β_0 and β_1. For each case, we also evaluate the accuracy of the proposed variance estimators. The criteria include the average of the square root of proposed variance estimates (SVE) and the corresponding empirical

coverage probabilities of nominal 95% confidence intervals for β_i (CP) for $i = 0, 1$ based on 1000 replications. For the reference of comparison, we also report SE, which is the standard error calculated based on 1000 replications. The variance estimates of $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$ and $\hat{\boldsymbol{\beta}}_F$ were computed using the formula given in equation (A.4) which can handle Z of both discrete and continuous types. Although equation (A.4) can be applied to estimating the variance of $\hat{\boldsymbol{\beta}}_{I1}$, it is complicated and becomes intractable analytically when Z is continuous. Thus we used the bootstrap re-sampling method for variance estimation for $\hat{\boldsymbol{\beta}}_{Ij}$ for $j = 1, 2$. Specifically, 1000 sub-samples were drawn with replacement from the original sample, and for the kth sub-sample, we obtained $\hat{\boldsymbol{\beta}}_{Ij}^{(k)} = (\hat{\beta}_{Ij,0}^{(k)}, \hat{\beta}_{Ij,1}^{(k)})^T$ by solving $U_{Ij}(\boldsymbol{\beta}) = 0_{2\times 1}$ for $j = 1, 2$ and $k = 1, \ldots, 1000$. Then the variance of $\hat{\beta}_{Ij,i}$ can be estimated by calculating the variance of $\hat{\beta}_{Ij,i}^{(k)}$ for $j = 1, 2$, $i = 0, 1$ and $k = 1, \ldots, 1000$.

Tables 5.1 lists the results when Z is binary and in Tables 5.2 and 5.3, we report the results when Z follows the standard normal and uniform distributions respectively. The results show that all the proposed estimators are more efficient than $\hat{\boldsymbol{\beta}}_F$ especially when Z is continuous. Furthermore, $\hat{\boldsymbol{\beta}}_{I1}$ and $\hat{\boldsymbol{\beta}}_{I2}$, obtained based on the imputation approach, perform better than $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$ and $\hat{\boldsymbol{\beta}}_F$ which utilize the weighting approach. We also observe larger bias of $\hat{\boldsymbol{\beta}}_F$ especially when the sample size is small. As in Table 5.2, $\hat{\boldsymbol{\beta}}_F$ still has large bias even when $n = 300$. We found that the IPCW technique, which utilizes $I(X_i \le \tau, B_i = 1)/\hat{G}(X_i)$ as a proxy of $I(T_i \le \tau, \tilde{B}_i = 1)$, would make an observation with larger X_i to be more influential in the estimation. Our proposal by setting $V_{1i} = \pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})(M_G - \pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta}))$ somewhat offset the influence of these observations. In contrast, Fine (1999) did not adjust the effect of censoring in his proposal of V_{1i} and hence $\hat{\boldsymbol{\beta}}_F$ was less stable.

Finally, we investigated whether the proposed methods remain robust when C actually depends on Z. We set $\log(C) = \gamma_{0,c} + \gamma_{1,c}Z + \sigma_c W_c$, where $\gamma_{0,c}, \gamma_{1,c}$ and σ_c are nuisance parameters and W_c has the pre-specified error distribution. In the computation of the proposed estimators, we evaluated two estimators of $G(t) = \Pr(C \ge t)$. One is the Kaplan-Meier estimator given in equation (4.8) and the other is a kernel-type smoothing estimator which simply replaces $B_i \ne 1$ by $B_i = 0$ in equation (4.23). Note that

the former is based on the wrong assumption that C does not depend on Z. Table 5.4 lists the results when Z is binary or follows the standard normal distribution. We only present the analysis for the estimation of β_1 since the results for β_0 are similar and hence omitted. It turns out that the results based on the Kaplan-Meier estimator of $G(t)$ are biased while the kernel approach yields less biased estimators. Generally speaking, it seems that the misspecification of $\hat{G}(t)$ has more influence on the bias term (BS) and less on the standard error (SE). All the proposed estimators are relatively more robust than $\hat{\boldsymbol{\beta}}_F$ under such a model mis-specification.

Table 5.1: *Finite-sample comparison for four estimators of* $(\beta_0, \beta_1) = (0.8, -1.24)$ *when the covariate* Z *is binary.*

Sample size	% censored	Estimators		Comparison criteria BS	SE	SVE	CP(%)	MSE	RE
100	30	$\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$	$\hat{\beta}_{\mathrm{w}^*,0}$	0.009	0.367	0.361	96.5	0.135	1.184
			$\hat{\beta}_{\mathrm{w}^*,1}$	-0.015	0.519	0.515	95.9	0.270	1.208
		$\hat{\boldsymbol{\beta}}_{I1}$	$\hat{\beta}_{I1,0}$	0.006	0.361	0.374	96.2	0.130	1.229
			$\hat{\beta}_{I1,1}$	0.000	0.507	0.530	97.2	0.257	1.267
		$\hat{\boldsymbol{\beta}}_{I2}$	$\hat{\beta}_{I2,0}$	0.006	0.360	0.377	96.6	0.130	1.232
			$\hat{\beta}_{I2,1}$	-0.001	0.507	0.523	97.0	0.257	1.267
		$\hat{\boldsymbol{\beta}}_{F}$	$\hat{\beta}_{F,0}$	0.017	0.399	0.401	97.6	0.160	1
			$\hat{\beta}_{F,1}$	-0.028	0.570	0.580	96.8	0.326	1
100	40	$\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$	$\hat{\beta}_{\mathrm{w}^*,0}$	0.049	0.464	0.440	94.5	0.218	1.405
			$\hat{\beta}_{\mathrm{w}^*,1}$	-0.083	0.598	0.581	94.4	0.365	1.498
		$\hat{\boldsymbol{\beta}}_{I1}$	$\hat{\beta}_{I1,0}$	0.044	0.456	0.499	97.2	0.210	1.457
			$\hat{\beta}_{I1,1}$	-0.061	0.579	0.613	96.7	0.339	1.610
		$\hat{\boldsymbol{\beta}}_{I2}$	$\hat{\beta}_{I2,0}$	0.042	0.456	0.478	96.8	0.210	1.460
			$\hat{\beta}_{I2,1}$	-0.060	0.579	0.630	96.3	0.339	1.613
		$\hat{\boldsymbol{\beta}}_{F}$	$\hat{\beta}_{F,0}$	0.061	0.550	0.561	97.0	0.306	1
			$\hat{\beta}_{F,1}$	-0.111	0.731	0.739	96.9	0.546	1
300	30	$\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$	$\hat{\beta}_{\mathrm{w}^*,0}$	0.007	0.230	0.230	94.8	0.053	1.154
			$\hat{\beta}_{\mathrm{w}^*,1}$	-0.011	0.306	0.297	95.1	0.094	1.208
		$\hat{\boldsymbol{\beta}}_{I1}$	$\hat{\beta}_{I1,0}$	0.007	0.227	0.234	95.9	0.052	1.188
			$\hat{\beta}_{I1,1}$	-0.010	0.301	0.300	95.3	0.091	1.248
		$\hat{\boldsymbol{\beta}}_{I2}$	$\hat{\beta}_{I2,0}$	0.001	0.225	0.231	95.7	0.051	1.213
			$\hat{\beta}_{I2,1}$	0.001	0.297	0.296	95.2	0.088	1.283
		$\hat{\boldsymbol{\beta}}_{F}$	$\hat{\beta}_{F,0}$	0.009	0.248	0.254	96.0	0.061	1
			$\hat{\beta}_{F,1}$	-0.013	0.336	0.334	96.0	0.113	1
300	40	$\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$	$\hat{\beta}_{\mathrm{w}^*,0}$	0.019	0.279	0.268	94.6	0.079	1.331
			$\hat{\beta}_{\mathrm{w}^*,1}$	-0.028	0.338	0.338	94.8	0.115	1.411
		$\hat{\boldsymbol{\beta}}_{I1}$	$\hat{\beta}_{I1,0}$	0.019	0.277	0.277	95.7	0.077	1.353
			$\hat{\beta}_{I1,1}$	-0.027	0.334	0.346	95.1	0.112	1.443
		$\hat{\boldsymbol{\beta}}_{I2}$	$\hat{\beta}_{I2,0}$	0.019	0.276	0.275	95.5	0.077	1.362
			$\hat{\beta}_{I2,1}$	-0.026	0.333	0.340	95.3	0.112	1.447
		$\hat{\boldsymbol{\beta}}_{F}$	$\hat{\beta}_{F,0}$	0.024	0.323	0.319	95.9	0.105	1
			$\hat{\beta}_{F,1}$	-0.033	0.401	0.411	96.4	0.162	1

Table 5.2: *Finite-sample comparison for four estimators of* $(\beta_0, \beta_1) = (0, 1.8)$ *when the covariate* Z *follows the standard normal distribution.*

Sample size	% censored	Estimators		Comparison criteria BS	SE	SVE	CP(%)	MSE	RE
100	30	$\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$	$\hat{\beta}_{\mathrm{w}^*,0}$	0.023	0.306	0.294	95.5	0.094	1.406
			$\hat{\beta}_{\mathrm{w}^*,1}$	0.134	0.498	0.446	94.1	0.266	2.317
		$\hat{\boldsymbol{\beta}}_{I1}$	$\hat{\beta}_{I1,0}$	-0.012	0.292	0.295	96.5	0.085	1.552
			$\hat{\beta}_{I1,1}$	-0.070	0.455	0.474	94.0	0.212	2.909
		$\hat{\boldsymbol{\beta}}_{I2}$	$\hat{\beta}_{I2,0}$	-0.020	0.288	0.294	95.8	0.084	1.581
			$\hat{\beta}_{I2,1}$	-0.073	0.455	0.444	93.4	0.212	2.906
		$\hat{\boldsymbol{\beta}}_{F}$	$\hat{\beta}_{F,0}$	-0.019	0.363	0.373	96.5	0.132	1
			$\hat{\beta}_{F,1}$	0.134	0.774	0.841	93.4	0.617	1
100	40	$\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$	$\hat{\beta}_{\mathrm{w}^*,0}$	0.009	0.338	0.325	94.8	0.114	1.455
			$\hat{\beta}_{\mathrm{w}^*,1}$	0.138	0.541	0.479	94.3	0.312	3.742
		$\hat{\boldsymbol{\beta}}_{I1}$	$\hat{\beta}_{I1,0}$	-0.003	0.325	0.341	97.0	0.106	1.577
			$\hat{\beta}_{I1,1}$	-0.079	0.475	0.476	95.3	0.231	5.044
		$\hat{\boldsymbol{\beta}}_{I2}$	$\hat{\beta}_{I2,0}$	-0.005	0.326	0.345	96.8	0.106	1.569
			$\hat{\beta}_{I2,1}$	-0.077	0.474	0.481	95.7	0.231	5.062
		$\hat{\boldsymbol{\beta}}_{F}$	$\hat{\beta}_{F,0}$	-0.013	0.408	0.426	96.2	0.166	1
			$\hat{\beta}_{F,1}$	0.157	1.069	1.549	96.1	1.167	1
300	30	$\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$	$\hat{\beta}_{\mathrm{w}^*,0}$	0.006	0.171	0.169	94.9	0.029	1.738
			$\hat{\beta}_{\mathrm{w}^*,1}$	0.038	0.257	0.252	94.4	0.067	4.689
		$\hat{\boldsymbol{\beta}}_{I1}$	$\hat{\beta}_{I1,0}$	-0.002	0.159	0.163	95.4	0.025	2.013
			$\hat{\beta}_{I1,1}$	-0.025	0.249	0.247	95.7	0.063	5.035
		$\hat{\boldsymbol{\beta}}_{I2}$	$\hat{\beta}_{I2,0}$	-0.003	0.158	0.165	96.2	0.025	2.027
			$\hat{\beta}_{I2,1}$	-0.025	0.248	0.251	96.0	0.062	5.098
		$\hat{\boldsymbol{\beta}}_{F}$	$\hat{\beta}_{F,0}$	0.030	0.223	0.219	96.9	0.051	1
			$\hat{\beta}_{F,1}$	0.097	0.553	0.566	93.8	0.316	1
300	40	$\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$	$\hat{\beta}_{\mathrm{w}^*,0}$	0.006	0.188	0.182	94.7	0.035	1.871
			$\hat{\beta}_{\mathrm{w}^*,1}$	0.048	0.280	0.270	95.1	0.081	6.940
		$\hat{\boldsymbol{\beta}}_{I1}$	$\hat{\beta}_{I1,0}$	-0.003	0.180	0.180	95.0	0.032	2.037
			$\hat{\beta}_{I1,1}$	-0.065	0.251	0.252	93.5	0.067	8.322
		$\hat{\boldsymbol{\beta}}_{I2}$	$\hat{\beta}_{I2,0}$	-0.003	0.180	0.182	95.5	0.032	2.045
			$\hat{\beta}_{I2,1}$	-0.064	0.253	0.258	94.0	0.068	8.212
		$\hat{\boldsymbol{\beta}}_{F}$	$\hat{\beta}_{F,0}$	0.022	0.256	0.276	97.2	0.066	1
			$\hat{\beta}_{F,1}$	0.101	0.742	0.792	96.2	0.560	1

Table 5.3: *Finite-sample comparison for four estimators of* $(\beta_0, \beta_1) = (1.23, 0.86)$ *when the covariate* Z *follows the uniform distribution.*

Sample size	% censored	Estimators		Comparison criteria BS	SE	SVE	CP(%)	MSE	RE
100	30	$\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$	$\hat{\beta}_{\mathrm{w}^*,0}$	0.063	0.371	0.362	93.8	0.142	3.388
			$\hat{\beta}_{\mathrm{w}^*,1}$	0.041	0.291	0.261	93.3	0.086	3.330
		$\hat{\boldsymbol{\beta}}_{I1}$	$\hat{\beta}_{I1,0}$	0.067	0.421	0.406	96.8	0.181	2.643
			$\hat{\beta}_{I1,1}$	0.037	0.259	0.264	97.0	0.068	4.211
		$\hat{\boldsymbol{\beta}}_{I2}$	$\hat{\beta}_{I2,0}$	0.060	0.394	0.406	97.0	0.159	3.023
			$\hat{\beta}_{I2,1}$	0.032	0.258	0.270	97.2	0.068	4.259
		$\hat{\boldsymbol{\beta}}_{F}$	$\hat{\beta}_{F,0}$	0.116	0.684	0.719	93.6	0.481	1
			$\hat{\beta}_{F,1}$	0.090	0.529	0.607	93.2	0.288	1
100	40	$\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$	$\hat{\beta}_{\mathrm{w}^*,0}$	0.111	0.494	0.432	92.8	0.256	3.759
			$\hat{\beta}_{\mathrm{w}^*,1}$	0.073	0.347	0.308	93.5	0.126	3.568
		$\hat{\boldsymbol{\beta}}_{I1}$	$\hat{\beta}_{I1,0}$	0.092	0.448	0.471	97.1	0.209	4.607
			$\hat{\beta}_{I1,1}$	0.049	0.311	0.317	96.4	0.099	4.521
		$\hat{\boldsymbol{\beta}}_{I2}$	$\hat{\beta}_{I2,0}$	0.091	0.445	0.466	97.5	0.206	4.675
			$\hat{\beta}_{I2,1}$	0.044	0.308	0.313	96.0	0.097	4.626
		$\hat{\boldsymbol{\beta}}_{F}$	$\hat{\beta}_{F,0}$	0.181	0.965	1.107	92.6	0.964	1
			$\hat{\beta}_{F,1}$	0.127	0.657	0.891	93.3	0.448	1
300	30	$\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$	$\hat{\beta}_{\mathrm{w}^*,0}$	0.009	0.198	0.203	96.2	0.039	3.312
			$\hat{\beta}_{\mathrm{w}^*,1}$	0.019	0.143	0.146	95.1	0.021	4.303
		$\hat{\boldsymbol{\beta}}_{I1}$	$\hat{\beta}_{I1,0}$	0.023	0.194	0.206	95.5	0.038	3.410
			$\hat{\beta}_{I1,1}$	-0.028	0.124	0.140	96.5	0.016	5.543
		$\hat{\boldsymbol{\beta}}_{I2}$	$\hat{\beta}_{I2,0}$	0.024	0.193	0.207	97.0	0.038	3.442
			$\hat{\beta}_{I2,1}$	-0.030	0.122	0.131	96.1	0.016	5.664
		$\hat{\boldsymbol{\beta}}_{F}$	$\hat{\beta}_{F,0}$	0.060	0.355	0.358	94.4	0.130	1
			$\hat{\beta}_{F,1}$	0.059	0.293	0.292	94.2	0.089	1
300	40	$\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$	$\hat{\beta}_{\mathrm{w}^*,0}$	0.043	0.252	0.243	95.3	0.065	5.619
			$\hat{\beta}_{\mathrm{w}^*,1}$	0.028	0.172	0.163	94.2	0.030	6.892
		$\hat{\boldsymbol{\beta}}_{I1}$	$\hat{\beta}_{I1,0}$	0.042	0.244	0.249	95.7	0.062	5.956
			$\hat{\beta}_{I1,1}$	0.015	0.167	0.168	95.6	0.028	7.443
		$\hat{\boldsymbol{\beta}}_{I2}$	$\hat{\beta}_{I2,0}$	0.041	0.242	0.250	96.0	0.060	6.062
			$\hat{\beta}_{I2,1}$	-0.015	0.163	0.170	95.5	0.027	7.777
		$\hat{\boldsymbol{\beta}}_{F}$	$\hat{\beta}_{F,0}$	0.139	0.589	0.611	93.9	0.367	1
			$\hat{\beta}_{F,1}$	0.099	0.447	0.495	94.6	0.209	1

Table 5.4: *Robustness analysis when the censoring variable depends on Z. When Z is binary,* $\beta_1 = -1.24$ *and when* $Z \sim N(0,1)$, $\beta_1 = 1.8$.

Sample size = 300, % censored = 30						
			Estimators			
type of covariate	$\hat{G}$	Criteria	$\hat{\beta}_{\mathrm{w}^*,1}$	$\hat{\beta}_{I1,1}$	$\hat{\beta}_{I2,1}$	$\hat{\beta}_{F,1}$
Binary	Kernel-type	BS	-0.029	-0.015	-0.001	-0.037
		SE	0.319	0.316	0.315	0.323
		MSE	0.103	0.100	0.099	0.105
	Kaplan-Meier	BS	0.092	0.083	0.081	-0.989
		SD	0.326	0.318	0.318	0.471
		MSE	0.114	0.108	0.108	1.199
Standard Normal	Kernel-type	BS	-0.073	-0.066	-0.064	0.098
		SD	0.254	0.245	0.244	0.260
		MSE	0.070	0.064	0.064	0.077
	Kaplan-Meier	BS	-0.191	-0.106	-0.103	2.775
		SD	0.258	0.248	0.246	1.058
		MSE	0.103	0.073	0.071	8.823

5.2 Analysis of Heart Transplant Data

The proposed inference procedures are applied to the Stanford Heart Transplant data (Crowley and Hu, 1977, pp.$\sim$ 28-29). Larson and Dinse (1985) also analyzed this data set in the context of model (2.5). Following Larson and Dinse, we consider only the subset of 65 patients who received a transplant and had complete data on the covariates of interest. Deaths were attributed to transplant rejection ($\tilde{B} = 1$) or other causes ($\tilde{B} = 2$). Among the 65 heart recipients, there were 29 rejected deaths; 12 deaths were from other causes and 24 patients were censored. The covariates include the waiting time from acceptance to surgery (w); the age at surgery (age) and a continuous mismatch score (m). Both m and age are transformed to have zero mean and unit variance, and w was recorded as a binary variable according to whether or not the waiting time exceeded 31 days. The survival time T (in days) was measured from the date of transplant surgery. The main objective is to explore the relationship between certain covariates and the cause of death due to transplant rejection.

To assess if the censoring time C depends on the selected covariates, the Cox proportional hazard model was fitted for C on each covariate separately. All p-values are larger than 0.1, hence we assume that the distribution of C does not depend on Z. The quantity of interest is $F_1(\tau) = \Pr(T \leq \tau, \tilde{B} = 1)$, the cumulative incidence probability of rejection by time τ. We set $\tau = 250, 500, 900, 1800$ (days). For each covariate, we ran simple logistic regression under the model:

$$\log\left[\frac{F_1(\tau)}{1 - F_1(\tau)}\right] = b_0(\tau) + b_1(\tau)z,$$

where z is one of the covariates. The waiting time w was not significant at all values of τ. The effect of the mismatch score m was insignificant for small values of τ and then became more obvious as τ increases. The covariate age is significant for all values of τ. Excluding w, we fitted a multiple logistic regression model which contained the covariates age and m. In Table 5.5, we see that age still played an important role for all values of τ, but the effect of mismatch score became insignificant when it is considered jointly with age. We conclude that age was the determining factor of $F_1(\tau)$. That is,

a patient with a younger age at the transplant surgery tended to have lower chance to develop the transplant rejection.

Larson and Dinse (1985) analyzed the same dataset under the framework of model (2.5). It was assumed that $\Pr(\tilde{B} = 1)$, the incidence rate of dying from transplant rejection, follows a logistic model and the latency distribution $1 - Q_j(t) = \Pr(T \leq t|\tilde{B} = j)$ follows a proportional hazard model for $j = 1, 2$ with

$$\Pr(T > t|\tilde{B} = j, \mathbf{Z}) = \exp\left[-\int_0^t h_j(u)\exp(\mathbf{Z}^T\gamma_j)du\right],$$

where $h_j(t)$ is specified as a piece-wise exponential function. Their analysis showed that no covariates have significant effect on $\Pr(\tilde{B} = 1)$ but both *age* and m were important for the latency distribution associated with transplant rejection. Our result coincides with that of Larson and Dinse (1985) in that *age* plays an important role for $F_1(t)$. However Larson and Dinse (1985) attributed the influence of *age* on $F_1(t)$ to the latency distribution $1 - Q_1(t)$. In contrast, our analysis showed that the effect of *age* on $F_1(\tau)$ persisted throughout all selected values of τ. It is reasonable to expect that such effect might continue to $\Pr(\tilde{B} = 1)$, which however, conflicts with the conclusion of Larson and Dinse (1985).

To investigate this contradiction, we divide the data set into three age groups such that *group* j represents the group with *age* ≤ 45, $\in (45, 51)$ and > 51 for $j = 1, 2, 3$ respectively. By the formula (3.6), nonparametric estimators of $F_1(t)$ for each age group can be obtained. For comparison, we also applied the mixed logistic/proportional hazard model of Larson and Dinse (1985) for the grouping age variable and the model-based estimators of $F_1(t)$ can be obtained by (2.5) with plugging in estimates of corresponding regression parameters. Figure 5.1 lists plots of both types of estimators of $F_1(t)$ in which the nonparametric estimators, $\hat{F}_1^{NP}(t)$ were obtained as stepped functions and the model-based ones, $\hat{F}_1^{LD}(t)$ were illustrated by thinner curves.

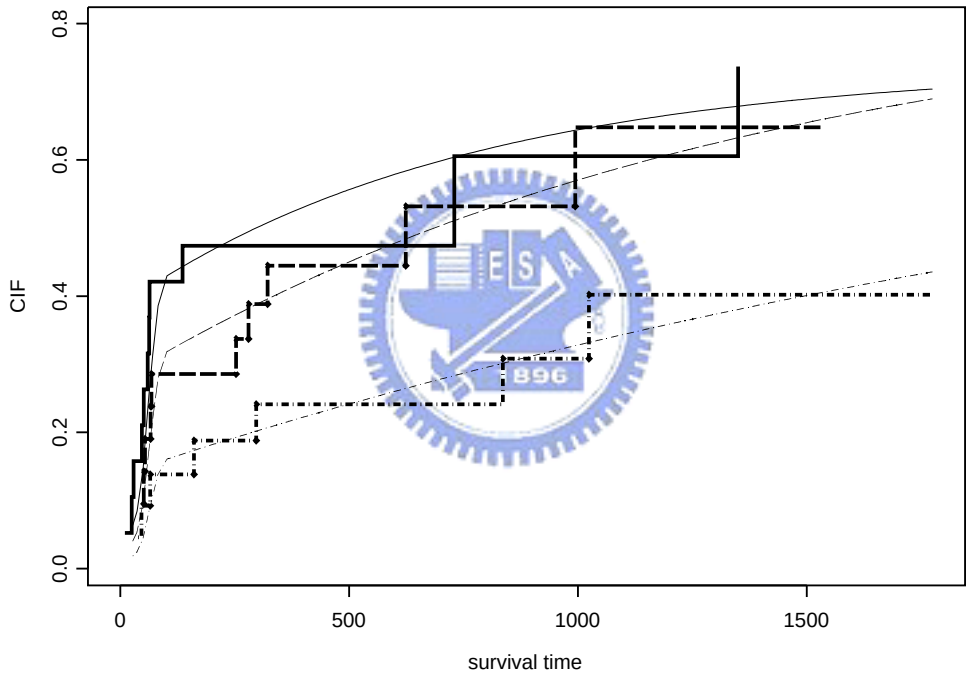
We look at the curves of $\hat{F}_1^{NP}(t)$ first. The curves of the two elder groups differed at the beginning but then became closer as the time passed by, suggesting that the two elder groups have similar rates of dying from rejection but different evolutions. On

the other hand, the youngest group had lower cumulative incidence probability of dying from rejection throughout the entire study period which illustrates an evidence of lower incidence rates of rejected death of the younger than that of the elder. Plots of $\hat{F}_1^{LD}(t)$ seem to agree with that of $\hat{F}_1^{NP}(t)$ within the range of study period.

Table 5.5: *Multiple Regression analysis for Heart Transplant data. In each cell, the estimated parameter and its standard error (in parenthesis) are given. Items with p-value* < 0.05 *are marked by* ^{a}.

	Covariate	$\tau = 1800$	$\tau = 900$	$\tau = 500$	$\tau = 250$
U_{w^*}	int	0.545 (0.463)	-0.037 (0.374)	-0.653 (0.311)	-1.016 (0.333)
	age	1.561 (0.542)^a	1.279 (0.382)^a	0.970 (0.310)^a	1.070 (0.351)^a
	m	0.727 (0.549)	0.786 (0.496)	0.691 (0.392)	0.672 (0.386)
U_{I1}	int	0.139 (0.470)	-0.136 (0.410)	-0.775 (0.336)	-1.087 (0.375)
	age	1.357 (0.569)^a	1.208 (0.518)^a	0.927 (0.370)^a	1.052 (0.442)^a
	m	0.665 (0.629)	0.790 (0.654)	0.563 (0.432)	0.601 (0.452)
U_{I2}	int	0.137 (0.464)	-0.152 (0.410)	-0.760 (0.333)	-1.076 (0.378)
	age	1.329 (0.527)^a	1.197 (0.458)^a	0.921 (0.380)^a	1.047 (0.438)^a
	m	0.598 (0.580)	0.696 (0.553)	0.543 (0.410)	0.588 (0.451)
U_F	int	0.420 (0.484)	-0.061 (0.370)	-0.657 (0.308)	-1.002 (0.330)
	age	1.624 (0.748)^a	1.265 (0.412)^a	0.949 (0.307)^a	1.080 (0.356)^a
	m	0.416 (0.603)	0.570 (0.502)	0.613 (0.395)	0.634 (0.394)

Figure 5.1: Plot of the cumulative incidence function of rejected death versus survival time for three groups with $age \leq 45$ ($-\cdot-\cdot-$), $45 < age \leq 51$ ($---$) and $51 < age$ (——). Step functions are nonparametric estimators and thinner curves comes from Larson and Dinse (1985)'s mixture model.



5.3 Analysis of SARS Data

5.3.1 Data Description

The Taiwan nationwide Laboratory-confirmed SARS database was kindly provided by Dr. Mei-Shiang Ho and her colleagues in the institute of Biomedical Sciences, Academia Sinica. Patients with SARS had to be isolated in the hospital until recovery or death. The process can be described using the framework of competing risks. Here we define $\tilde{B} = 1$ to indicate that a patient was cured from the disease (being discharged from the hospital and alive) and $\tilde{B} = 2$ to indicate that a patient was not cured (died during the hospitalization period). Because this infectious disease has been eventually under control in Taiwan, the database contains complete information about the two outcomes and the corresponding failure time. There are 258 infected patients in which 58 subjects were dead during the hospitalization period and 200 subjects were discharged from the hospital and alive.

Possible covariates include *age*, *gender*, *disease*, PCR, *viral load*, where *age* denotes a patient's age by years; *disease* is a binary variable indicating whether a patient had suffered from other diseases before getting infected of SARS (1: yes, 0: no); PCR is an indicator of whether the Polymerase Chain Reaction (PCR) test detected the SARS virus (1: yes, 0: no) and *viral load* measures the viral load detected by the PCR test. Note that if PCR equals 0, the individual had a negative virus titer, meaning that the patient has anti-body but zero viral load, and then the *viral load* is set as zero.

5.3.2 Analysis of the Original Complete Data

The function $F_1(t) = \Pr(\tilde{B} = 1, T \leq t)$ measures the probability of being discharged from the hospital (cured) by time t. We first present nonparametric analysis for each covariate group. Then we perform simple regression analysis for each covariate group using the LOGISTIC procedure in SAS.

Figures 5.2-5.6 depict the empirical estimators of $F_1(t)$ based on the covariate groups. The continuous variable *age* was first divided into three groups, $age < 30$, $30 \leq age \leq 50$ and $age > 50$. Figure 5.2 shows that the two younger groups ($age < 30$ and $30 \leq age \leq 50$) have similar patterns, while the older group ($age > 50$) has much lower chance of recovery at every t. At the end, the cure proportions of three age groups (from the youngest to the oldest) are 0.925, 0.879 and 0.443, respectively. The patterns of $F_1(t)$ for the *gender* groups and *disease* groups are similar such that the curves associated with different covariate values have no crossings. At the end, the female group (cure proportion = 0.842) had better recovery than the male group (cure proportion = 0.656). Individuals without previous diseases (cure proportion = 0.845) also revealed better recovery than those in presence of other disease (cure proportion = 0.444).

The curves based on different groups of PCR and *viral load* behave differently than the former covariates. Note that *viral load*, originally measured continuously, was stratified into four groups: no viral load detected, $\leq 10^3$, $\in (10^3, 10^5]$, and $> 10^5$. The first group includes those with PCR = 0 (cure proportion = 0.946) and the last three groups are those with PCR = 1 (cure proportions equal 0.763, 0.648 and 0.526, respectively). At the end, the larger the level of viral load, the lower chance of recovery. However the four empirical curves have intersections in some middle time points.

We conducted several simple regression analyses based on the model

$$\text{logit}\left[F_1(\tau_j)\right] = \beta_{0,j} + \beta_{1,j}\ \text{Z}, \tag{5.2}$$

where Z is a selected covariate and $\beta_{k,j}$ are simplifications of $\beta_k(\tau_j)$ $(k = 0, 1)$ for $j = 1, \ldots, 5$. In the analysis, *age* was divided into two groups (≤ 50 and > 50) and *viral load* was transformed into the scale of $\log_{10}$ to stabilize the effect caused by extreme large values. We set $\tau_1 = 14$, $\tau_2 = 21$, $\tau_3 = 28$, $\tau_4 = 35$ and τ_5 to be the maximum length of hospitalization for the cure satisfying $F_1(\tau_5) = \Pr(\tilde{B} = 1)$. The results are summarized in Table 5.6. Here we discuss the effect of *age* for illustration. Treating the younger group ($age \leq 50$) as the baseline, the odds ratios along the time $e^{\hat{\beta}_{1,j}}$'s are 0.466, 0.221, 0.158, 0.146 and 0.090. This implies that the effect of *age* on the odds of $F_1(\tau_j)$ tends to

be more influential as τ_j gets larger. Notice that *age* has substantial effect on $F_1(\tau_5)$, the final chance of recovery. For comparison, we analyze the data under model (2.2) which assumes $\beta_{1,1} = \ldots = \beta_{1,5} = \beta_1$. The overall odds ratio $e^{\hat{\beta}_1}$ is 0.147 which seems very different from the separate odds ratios reported above. To formally examine whether the effect of *age* is time independent, a score test for assessing the difference between the reduced and the full model (with four degree of freedom) was performed. The resulting p-value is 0.006 which implies that model (2.2) is not suitable for measuring the influence of *age* on $F_1(t)$. Table 5.6 also shows that each covariate has a significant effect on $F_1(\tau_j)$ for larger τ_j. In general, younger females, who did not have other diseases and had lower viral load, had the best chance of recovery from SARS. Note that the effect of *gender* remained the same along the time. In fact, Figure 5.3 shows that the curves for the male and the female do not intersect. However the two curves with different disease status have no crossing but the test of time homogeneity is rejected.

Although our paper does not study whether a covariate affects the latency distribution $Q_1(t|\tau)$, here we illustrate how to conduct further analysis if this is also of some interest. Let us use *age* again for illustration. We fit $Q_1(t|\tau_5)$ by the accelerated failure time model with a Weibull distribution, the estimated regression parameter for *age* is 0.276 (p-value = 0.004). The result implies that, for older patients ($age > 50$) who were eventually cured, they also needed longer time to get recovery (Figure 5.7).

Figure 5.2: The cumulative incidence function of cure for three age groups: $age < 30$ ($-\cdot-\cdot-$), $30 \leq age \leq 50$ (——) and $50 < age$ ($- - -$).

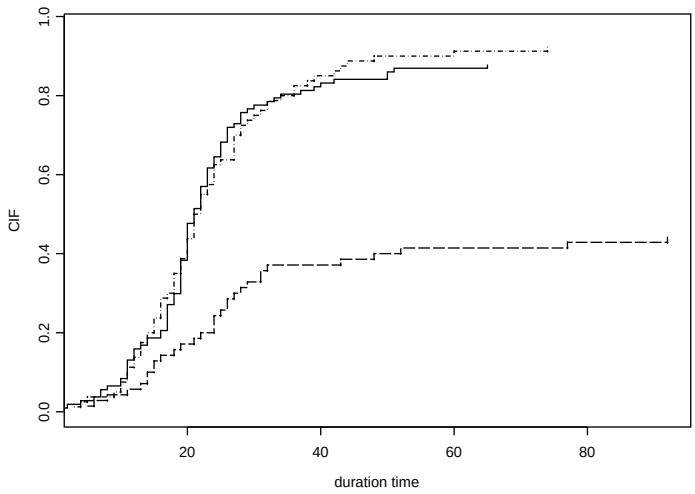


Figure 5.3: The cumulative incidence function of cure for two groups of different gender: $female$ ($-\cdot-\cdot-$) and $male$ (——).

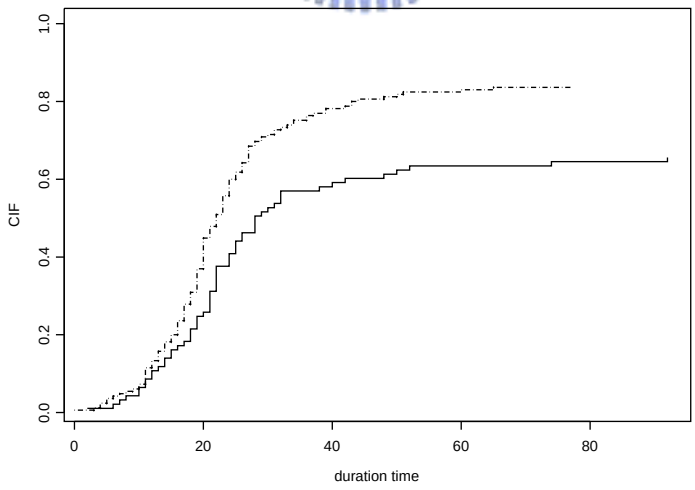


Figure 5.4: The cumulative incidence function of cure for two groups with/without other diseases: "without disease" (−·−·−) and "with disease" (——).

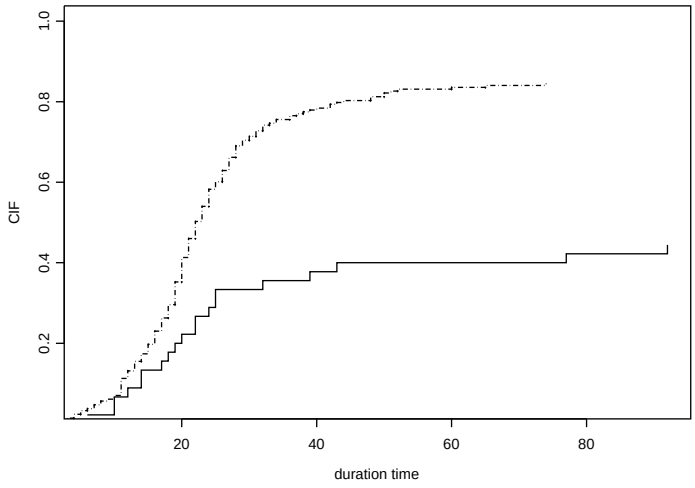


Figure 5.5: The cumulative incidence function of cure for groups with/without PCR: PCR = 1 (——) and PCR = 0 (−·−·−).

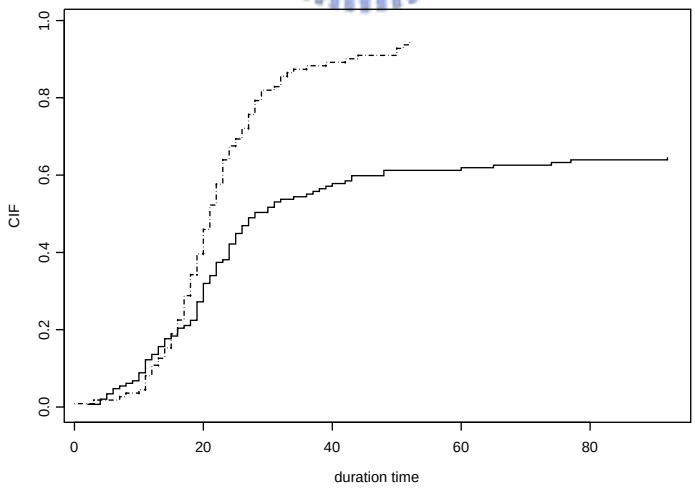


Figure 5.6: The cumulative incidence function of cure for groups with different level of virus load (vl): vl = 0 (−···−), $0 <$ vl $< 10^3$ (−−−), $10^3 <$ vl $< 10^5$ (−·−·−) and $10^5 <$ vl (——).

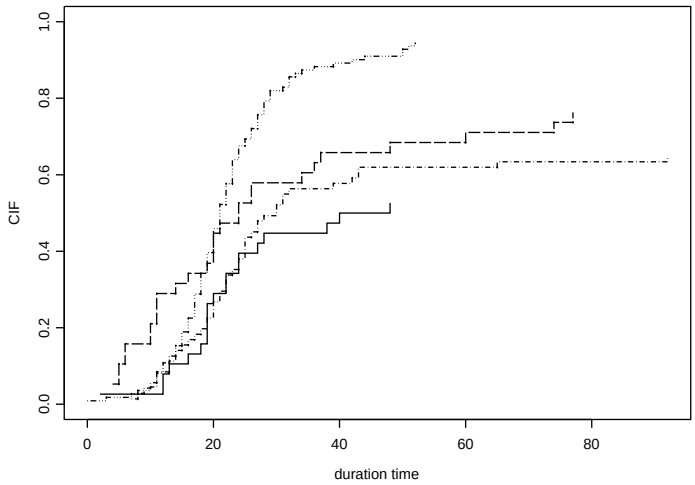


Figure 5.7: The latency survival function of cure for two age groups: $age \leq 50$ (−·−·−) and $age > 50$ (——).

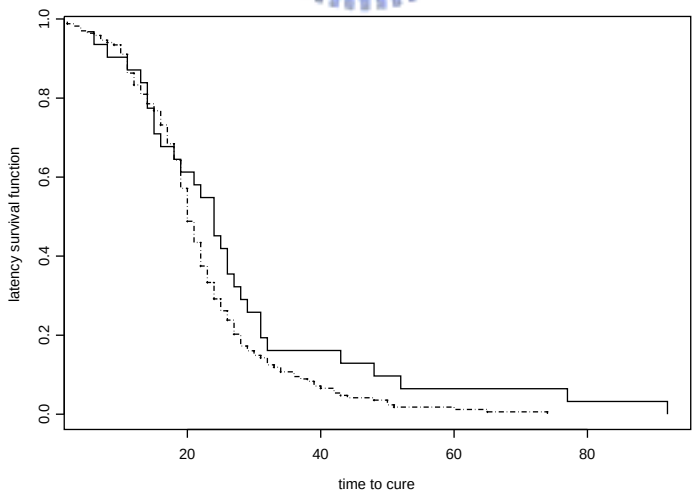


Table 5.6: Output for SARS analysis for different τ's, spaced by one week but staring at two weeks. Items with p-values < 0.05 are marked by a.

Covariate	$\hat{\beta}_{1,1}$	$\hat{\beta}_{1,2}$	$\hat{\beta}_{1,3}$	$\hat{\beta}_{1,4}$	$\hat{\beta}_{1,5}$	$\hat{\beta}_1$	p-value of testing homogeneity
age							
> 50	-0.763	-1.510^a	-1.843^a	-1.926^a	-2.409^a	-1.914^a	0.006
	(0.439)	(0.340)	(0.307)	(0.308)	(0.341)	(0.276)	
gender							
male	-0.313	-0.707^a	-0.811^a	-0.825^a	-1.031^a	-0.755^a	0.476
	(0.361)	(0.273)	(0.268)	(0.276)	(0.305)	(0.234)	
disease							
with disease	-0.312	-1.093^a	-1.494^a	-1.725^a	-1.920^a	-1.568^a	0.007
	(0.474)	(0.384)	(0.349)	(0.350)	(0.355)	(0.312)	
PCR							
positive	0.172	-0.753^a	-1.328^a	-1.758^a	-2.260^a	-0.934^a	< 0.001
	(0.341)	(0.258)	(0.286)	(0.330)	(0.454)	(0.229)	
$\log_{10}$viral load	-0.064	-0.207^a	-0.293^a	-0.370^a	-0.461^a	-0.264^a	< 0.001
	(0.075)	(0.059)	(0.061)	(0.066)	(0.078)	(0.051)	

5.3.3 Analysis of Censored SARS Data

In practice, interim analysis based on incomplete data provides timely information for decision making. Although the original SARS dataset contains complete information about the value of $(T, \tilde{B})$, it is worthy to investigate how the proposed methods behave if this dataset is subject to further censoring. Here we generated a censoring variable which has a uniform distribution taking values from 0 to 70 making the censoring proportion to be around 30%.

Based on a censored version of the SARS data, we applied the proposed methods to fit a simple logistic regression for each covariate group and found that each covariate was statistically significant since time τ_2. Then we included all the covariates in the multiple logistic regression model which showed that the covariates *gender* and PCR became insignificant at all values of τ. The final fitted model is

$$logit[F_1(\tau)] = \beta_0(\tau) + \beta_1(\tau)\ age + \beta_2(\tau)\ disease + \beta_3(\tau)\ \log_{10}(viral\ load).$$

Table 5.7 lists the detail results of the above analysis based on a single run using the artificial censoring scheme. Note that in the table we also report the previous results obtained from solving $\tilde{U}(\boldsymbol{\beta}) = 0_{4\times 1}$, the score equation based on the original complete data. With the additional censoring, the proposed methods yield similar point estimates but larger standard deviations, as expected. Table 5.8 list the average results by repeating the censoring scheme 300 times. The patterns are similar to that in a single run. Note that the proposed estimators also produce more precise results compared with the estimator of Fine (1999).

Table 5.7: *Multiple logistic regression analysis for SARS data subject to a single run of artificial censoring. In each cell, the estimated parameter and the estimated standard error (in parenthesis) are given. Items with p-value* < 0.05 *are marked by a and with p-value* < 0.1 *are marked by b.*

	Covariate	$\tau_2 = 21$	$\tau_3 = 28$	$\tau_4 = 35$	$\tau_5 = 92$
$\tilde{U}$	age > 50	-1.240 (0.370)^a	-1.478 (0.343)^a	-1.486 (0.355)^a	-2.066 (0.414)^a
	with disease	-0.400 (0.435)	-0.654 (0.417)	-0.904 (0.427)^a	-0.854 (0.470)^b
	$\log_{10}$(viral load)	-0.158 (0.061)^a	-0.251 (0.065)^a	-0.338 (0.072)^a	-0.464 (0.093)^a
U_{w^*}	age > 50	-1.237 (0.462)^a	-1.647 (0.430)^a	-1.566 (0.435)^a	-1.885 (0.464)^a
	with disease	-0.581 (0.521)	-0.633 (0.485)	-0.805 (0.460)^b	-0.837 (0.507)^b
	$\log_{10}$(viral load)	-0.181 (0.068)^a	-0.286 (0.073)^a	-0.396 (0.082)^a	-0.473 (0.090)^a
U_{I_1}	age > 50	-1.276 (0.435)^a	-1.646 (0.373)^a	-1.568 (0.414)^a	-1.925 (0.488)^a
	with disease	-0.500 (0.567)	-0.757 (0.462)	-1.110 (0.509)^a	-1.113 (0.530)^a
	$\log_{10}$(viral load)	-0.173 (0.078)^a	-0.251 (0.067)^a	-0.360 (0.090)^a	-0.398 (0.089)^a
U_{I_2}	age > 50	-1.277 (0.435)^a	-1.655 (0.376)^a	-1.569 (0.410)^a	-1.947 (0.479)^a
	with disease	-0.488 (0.556)	-0.767 (0.468)	-1.015 (0.502)^a	-1.050 (0.527)^a
	$\log_{10}$(viral load)	-0.179 (0.077)^a	-0.259 (0.064)^a	-0.361 (0.087)^a	-0.408 (0.091)^a
U_F	age > 50	-1.222 (0.459)^a	-1.364 (0.449)^a	-1.291 (0.521)^a	-1.714 (0.639)^a
	with disease	-0.399 (0.516)	-0.505 (0.517)	-0.869 (0.568)	-0.910 (0.675)
	$\log_{10}$(viral load)	-0.215 (0.075)^a	-0.327 (0.092)^a	-0.443 (0.123)^a	-0.497 (0.181)^a

Table 5.8: *Multiple logistic regression analysis for SARS data by repeating artificial censoring 300 times. In each cell, the average of the parameter estimates and the average of the standard-error estimates are reported.*

	Covariate	$\tau_2 = 21$	$\tau_3 = 28$	$\tau_4 = 35$	$\tau_5 = 92$
$\tilde{U}$	age > 50	-1.240 (0.370)	-1.478 (0.343)	-1.486 (0.355)	-2.066 (0.414)
	with disease	-0.400 (0.435)	-0.654 (0.417)	-0.904 (0.427)	-0.854 (0.470)
	$\log_{10}$(viral load)	-0.158 (0.061)	-0.251 (0.065)	-0.338 (0.072)	-0.464 (0.093)
U_{w^*}	age > 50	-1.279 (0.450)	-1.507 (0.417)	-1.530 (0.440)	-2.116 (0.506)
	with disease	-0.393 (0.487)	-0.672 (0.454)	-0.941 (0.478)	-1.170 (0.537)
	$\log_{10}$(viral load)	-0.158 (0.069)	-0.250 (0.075)	-0.339 (0.085)	-0.500 (0.102)
U_{I_1}	age > 50	-1.269 (0.486)	-1.502 (0.442)	-1.538 (0.462)	-2.115 (0.520)
	with disease	-0.402 (0.557)	-0.699 (0.509)	-0.993 (0.526)	-1.174 (0.585)
	$\log_{10}$(viral load)	-0.156 (0.071)	-0.247 (0.076)	-0.331 (0.088)	-0.442 (0.100)
U_{I_2}	age > 50	-1.270 (0.487)	-1.500 (0.438)	-1.540 (0.463)	-2.114 (0.515)
	with disease	-0.404 (0.550)	-0.685 (0.498)	-0.989 (0.521)	-1.173 (0.585)
	$\log_{10}$(viral load)	-0.157 (0.071)	-0.241 (0.076)	-0.333 (0.086)	-0.433 (0.099)
U_F	age > 50	-1.265 (0.469)	-1.483 (0.480)	-1.500 (0.553)	-2.365 (1.099)
	with disease	-0.410 (0.535)	-0.711 (0.550)	-0.990 (0.602)	-1.246 (0.955)
	$\log_{10}$(viral load)	-0.156 (0.075)	-0.250 (0.098)	-0.345 (0.126)	-0.587 (0.346)

Chapter 6

Concluding Remarks

In the thesis, we apply two useful techniques, namely inverse probability censoring weighting and imputation, to handle missing responses in analysis of a logistic regression model. The proposed estimating functions based on the weighting approach further consider efficiency improvement by taking the censoring effect into account and utilizing more data information. The imputation approach has better performance in the simulations but it also involves estimating more nuisance quantities. In the thesis, we demonstrate that these nuisance functions can be handled nonparametrically by applying the results of Wang (2003) to the current regression setting which however may need to use smoothing techniques and hence is quite technically involved. Furthermore if the dimension of the continuous covariates is high, kernel smoothing may not work well unless the sample size is substantially large. In such a case, one may try to reduce the dimension of $\mathbf{Z}$ based on preliminary analysis or impose additional model assumptions on the latency distributions to avoid the curse of dimensionality.

The logistic model assumption on $\pi(\mathbf{Z}_i^T\boldsymbol{\beta})$ is specified only for a chosen value of τ. However if the analysis will be implemented at several time points, we may want to let the form of $\pi(\cdot)$ to vary at different values of τ. For example, we can impose a parametric link family and then test the corresponding parameter value which also serves as a way of selecting the most fitted link at a given time. For example, the logistic link can be viewed as a member of the family $g(u;\rho) = \log\left[\{(1/(1-u))^{\rho} - 1\}/\rho\right]$ with $\rho = 1$. A

common way of testing $\rho = \rho_0$ is via the deviance of the likelihood function which is not suitable for our purpose. We may adopt the idea of Pregibon (1980) to test $\rho = \rho_0$ via a score test constructed based on the estimating function without specifying the likelihood.

Our proposed methods can be directly modified to incorporate the transformation model given in (2.2). We start with a grid of time points, $\tau_1, \ldots, \tau_L$. At these grid points, model (2.2) can be expressed as

$$F_1(\tau_\ell|\mathbf{Z}) = \pi\left(m(\tau_\ell) + \mathbf{Z}^T\boldsymbol{\theta}\right) = \pi\left(\alpha_\ell + \mathbf{Z}^T\boldsymbol{\theta}\right), \quad \ell = 1, \ldots, L \tag{6.1}$$

where $\pi = H^{-1}$. Let $\boldsymbol{\mu}^T = (\alpha_1, \ldots, \alpha_L, \boldsymbol{\theta}^T)$. We can construct estimating functions of $\boldsymbol{\mu}$ by the same steps in Sections 4.2.2 and 4.2.3. For example, we can reconstruct (4.6) and (4.7) as

$$H_{\ell,1i} = \frac{I(X_i \le \tau_\ell, B_i = 1)}{G(X_i)} - \pi(\alpha_\ell + \mathbf{Z}_i^T\boldsymbol{\theta}) \tag{6.2}$$

and

$$H_{\ell,2i} = \frac{I(X_i > \tau_\ell)}{G(\tau_\ell+)} + \frac{I(X_i \le \tau_\ell, B_i = 2)}{G(X_i)} - \bar{\pi}(\alpha_\ell + \mathbf{Z}_i^T\boldsymbol{\theta}) \tag{6.3}$$

for $i = 1, \ldots, n$. The resulting estimating function of $\boldsymbol{\mu}$ is

$$U_{\mathrm{w3}}(\mu) = \sum_{i=1}^{n} -E\left(\frac{\partial\tilde{\mathbf{H}}_i}{\partial\boldsymbol{\mu}}\right) V_i^{-1}(\boldsymbol{\mu})\ \tilde{\mathbf{H}}_i \tag{6.4}$$

where

$$\tilde{\mathbf{H}}_i = \begin{bmatrix} \hat{H}_{1,1i}, & \ldots, & \hat{H}_{L,1i} \\ \hat{H}_{2,1i}, & \ldots, & \hat{H}_{L,2i} \end{bmatrix}$$

in which $\hat{H}_{\ell,ji}$ are $H_{\ell,ji}$ $(j = 1, 2)$ with G being replaced by $\hat{G}$ defined in (4.8) and $V_i(\mu)$ is a working covariance matrix. We can select a reasonable covariance matrix for U_{w3} in the same way as did for U_{w^*}. Estimating functions based on the imputation approach can be also obtained in a similar way.

Now we briefly illustrate how to use model (4.1) to verify the "parallel-lines" assumption of model (2.2) or help choosing time-dependent covariates in that model. For

selected grid points, we reparameterize model (4.1) as

$$F_1(\tau_\ell|\mathbf{Z}) = \pi(\tilde{\mathbf{Z}}^T\boldsymbol{\beta}(\tau_\ell)) = \pi\left(\alpha_\ell + \mathbf{Z}^T\boldsymbol{\theta} + \mathbf{Z}^T\eta_1\phi_1 + \cdots + \mathbf{Z}^T\eta_{L-1}\phi_{L-1}\right), \quad \ell = 1, \ldots, L \tag{6.5}$$

where η_h is a dummy variable with value equal to 1 if $h = \ell$ or 0 otherwise and ϕ_h's are extra parameters measuring interactions between time and covariates. In this sense model (2.2) can be viewed as a reduced model of model (4.1) and hence the parallel lines assumption can be verified by testing the null hypothesis $\phi_1 = \cdots = \phi_{L-1} = 0$. We can construct a score test for testing this hypothesis. Let $\boldsymbol{\phi}^T = (\phi_1, \ldots, \phi_{L-1})$ and $\boldsymbol{\gamma}^T = (\boldsymbol{\mu}^T, \boldsymbol{\phi}^T)$ in which $\boldsymbol{\mu}^T = (\alpha_1, \ldots, \alpha_L, \boldsymbol{\theta}^T)$. Denote the estimating function of $\boldsymbol{\gamma}$ by $U_S(\boldsymbol{\gamma})$. For the IPCW approach, we can obtain $U_S(\boldsymbol{\gamma})$ in the same form of (6.4) with $H_{\ell,ji}$ $(j = 1, 2)$ are evaluated under model (6.5). Then $U_S(\boldsymbol{\gamma})$ may be shown to be asymptotically normally distributed with mean-zero and variance Γ, where Γ can be partitioned according to $\boldsymbol{\gamma}^T = (\boldsymbol{\mu}^T, \boldsymbol{\phi}^T)$. Denote this partition as $\Gamma = (\Gamma_{kl})$ for $k, l = 1, 2$. The score statistic for testing the above hypothesis is

$$Q = n^{-1}U_S^T(\hat{\boldsymbol{\gamma}})(\hat{\Gamma}_{22} - \hat{\Gamma}_{21}\hat{\Gamma}_{11}^{-1}\hat{\Gamma}_{12})^{-1}U_S(\hat{\boldsymbol{\gamma}}),$$

where $\hat{\boldsymbol{\gamma}}$ and $\hat{\Gamma}_{kl}$ are evaluated under the null hypothesis. Under the null hypothesis, Q is asymptotically chi-squared with $L-1$ degrees of freedom. One can refer to the papers by Chen (1983) and Li (1991) for details.

Appendix

Appendix A: *Asymptotic properties of* $U_{w^*}(\boldsymbol{\beta})$

Assume that the true value $\boldsymbol{\beta}_0$ is located in the interior of the parameter space, which is a bounded convex region and $\pi_\phi(\cdot)$ is bounded. The estimating function $U_{w^*}(\boldsymbol{\beta})$ in Section 4.2.2 can be written as

$$U_{w^*}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[(V_{2i} - V_{3i})H_{1i} - (V_{1i} - V_{3i})H_{2i}\right] \frac{\pi_\phi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})}{V_{1i}V_{2i} - V_{3i}^2}\tilde{\mathbf{Z}}_i + B_{2n}(\boldsymbol{\beta}),$$

where

$$\begin{aligned} B_{2n}(\boldsymbol{\beta}) \;=\; & \sum_{i=1}^{n} \left\{ \left[\frac{I(X_i \le \tau, B_i = 1)}{G(X_i)} \frac{V_{2i} - V_{3i}}{V_{1i}V_{2i} - V_{3i}^2} \pi_\phi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\tilde{\mathbf{Z}}_i \right] \frac{G(X_i) - \hat{G}(X_i)}{\hat{G}(X_i)} \right. \\ & - \left[\frac{I(X_i \le \tau, B_i = 2)}{G(X_i)} \frac{V_{1i} - V_{3i}}{V_{1i}V_{2i} - V_{3i}^2} \pi_\phi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\tilde{\mathbf{Z}}_i \right] \frac{G(X_i) - \hat{G}(X_i)}{\hat{G}(X_i)} \\ & \left. - \left[\frac{I(X_i > \tau)}{G(\tau+)} \frac{V_{1i} - V_{3i}}{V_{1i}V_{2i} - V_{3i}^2} \pi_\phi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta})\tilde{\mathbf{Z}}_i \right] \frac{G(\tau+) - \hat{G}(\tau+)}{\hat{G}(\tau+)} \right\}. \end{aligned}$$

To derive the asymptotic distribution of $n^{-1/2}U_{w^*}(\boldsymbol{\beta}_0)$, we first express the Kaplan-Meier estimator $\hat{G}(t)$ as the following integral form,

$$\frac{G(t) - \hat{G}(t)}{G(t)} = \sum_{i=1}^{n} \int_0^t \frac{\hat{G}(u-)}{G(u)} \frac{dM_{C,i}(u)}{\bar{Y}(u)},$$

where

$$M_{C,i}(u) = I(X_i \le u, B_i = 0) - \int_0^u I(X_i \ge s) d\Lambda_C(s),$$

$\bar{Y}(u) = \sum_{i=1}^{n} I(X_i \ge u)$ and $\Lambda_C(s)$ is the cumulative hazard function of C. By the uniform convergence of the Kaplan-Meier estimator, we can write $n^{-1/2}B_{2n}(\boldsymbol{\beta}_0)$ as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\infty \left[q_1(t;\boldsymbol{\beta}_0) - q_2(t;\boldsymbol{\beta}_0) - q_3(t;\boldsymbol{\beta}_0)\right] \left(\frac{\bar{Y}(t)}{n}\right)^{-1} dM_{C,i}(t) + o_p(1),$$

where

$$q_1(t;\boldsymbol{\beta}_0) = \frac{1}{n} \sum_{k=1}^{n} I(X_k \ge t) \left[\frac{I(X_k \le \tau, B_k = 1)}{G(X_k)} \right] \frac{\mathrm{v}_{2k} - \mathrm{v}_{3k}}{\mathrm{v}_{1k}\mathrm{v}_{2k} - \mathrm{v}_{3k}^2} \pi_\phi(\tilde{\mathbf{Z}}_k^T\boldsymbol{\beta}_0)\tilde{\mathbf{Z}}_k, \qquad \text{(A.1)}$$

$$q_2(t;\boldsymbol{\beta}_0) = \frac{1}{n}\sum_{k=1}^{n} I(X_k \geq t)\left[\frac{I(X_k \leq \tau, B_k = 2)}{G(X_k)}\right]\frac{\mathrm{v}_{1k} - \mathrm{v}_{3k}}{\mathrm{v}_{1k}\mathrm{v}_{2k} - \mathrm{v}_{3k}^2}\pi_\phi(\tilde{\mathbf{Z}}_k^T\boldsymbol{\beta}_0)\tilde{\mathbf{Z}}_k, \qquad \text{(A.2)}$$

$$q_3(t;\boldsymbol{\beta}_0) = \frac{1}{n}\sum_{k=1}^{n} I(\tau \geq t)\left[\frac{I(X_k > \tau)}{G(\tau+)}\right]\frac{\mathrm{v}_{1k} - \mathrm{v}_{3k}}{\mathrm{v}_{1k}\mathrm{v}_{2k} - \mathrm{v}_{3k}^2}\pi_\phi(\tilde{\mathbf{Z}}_k^T\boldsymbol{\beta}_0)\tilde{\mathbf{Z}}_k, \qquad \text{(A.3)}$$

$\mathrm{v}_{1k} = \pi(\tilde{\mathbf{Z}}_k^T\boldsymbol{\beta}_0)(\tilde{M} - \pi(\tilde{\mathbf{Z}}_k^T\boldsymbol{\beta}_0))$, $\mathrm{v}_{2k} = \bar{\pi}(\tilde{\mathbf{Z}}_k^T\boldsymbol{\beta}_0)(\tilde{M} - \bar{\pi}(\tilde{\mathbf{Z}}_k^T\boldsymbol{\beta}_0))$, $\mathrm{v}_{3k} = \bar{\pi}(\tilde{\mathbf{Z}}_k^T\boldsymbol{\beta}_0)\pi(\tilde{\mathbf{Z}}_k^T\boldsymbol{\beta}_0)$ and $\tilde{M}$ is the median of the random variable $1/G(X)$.

Therefore $n^{-1/2}U_{\mathrm{w}^*}(\boldsymbol{\beta}_0)$ can be expressed as $n^{-1/2}\sum_{i=1}^{n}\xi_i + o_p(1)$, where

$$\begin{aligned}
\xi_i &= \left\{\left[\frac{I(X_i \leq \tau, B_i = 1)}{G(X_i)} - \pi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta}_0)\right](\mathrm{v}_{2i} - \mathrm{v}_{3i})\right. \\
&- \left.\left[\frac{I(X_i > \tau)}{G(\tau+)} + \frac{I(X_i \leq \tau, B_i = 2)}{G(X_i)} - \bar{\pi}(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta}_0)\right](\mathrm{v}_{1i} - \mathrm{v}_{3i})\right\}\frac{\pi_\phi(\tilde{\mathbf{Z}}_i^T\boldsymbol{\beta}_0)}{\mathrm{v}_{1i}\mathrm{v}_{2i} - \mathrm{v}_{3i}^2}\tilde{\mathbf{Z}}_i \\
&+ \int_0^\infty \frac{q(t;\boldsymbol{\beta}_0)}{y(t)}dM_{C,i}(t),
\end{aligned}$$

$y(t) = \lim_{n\to\infty}\bar{Y}(t)/n$ and $q(t;\boldsymbol{\beta}_0) = \lim_{n\to\infty}[q_1(t;\boldsymbol{\beta}_0) - q_2(t;\boldsymbol{\beta}_0) - q_3(t;\boldsymbol{\beta}_0)]$. Since $\{\xi_i\ (i = 1, ..., n)\}$ are mean-zero independent random variables, by the multivariate central limit theorem, $n^{-1/2}U_{\mathrm{w}^*}(\boldsymbol{\beta}_0)$ has an asymptotic normal distribution with mean 0 and covariance matrix $\Gamma_{\mathrm{w}^*} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\xi_i\xi_i^T$.

Appendix B: *Asymptotic properties of* $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$

Recall that $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$ is the solution to $U_{\mathrm{w}^*}(\boldsymbol{\beta}) = 0$. Since $U_{\mathrm{w}^*}(\boldsymbol{\beta})$ is differentiable with respect to $\boldsymbol{\beta}$ and has a bounded derivative, consistency of $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$ follows. By a Taylor expansion of $n^{-1/2}U_{\mathrm{w}^*}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}_0$, we can write

$$0 = n^{-1/2}U_{\mathrm{w}^*}(\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}) = n^{-1/2}U_{\mathrm{w}^*}(\boldsymbol{\beta}_0) - A_{\mathrm{w}^*}(\boldsymbol{\beta}_0)\ n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{w}^*} - \boldsymbol{\beta}_0) + o_p(1),$$

where

$$A_{\mathrm{w}^*}(\boldsymbol{\beta}_0) = -\lim_{n\to\infty}\frac{1}{n}\frac{\partial U_{\mathrm{w}^*}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}.$$

It follows that

$$n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{w}^*} - \boldsymbol{\beta}_0) = [A_{\mathrm{w}^*}(\boldsymbol{\beta}_0)]^{-1}\, n^{-1/2}U_{\mathrm{w}^*}(\boldsymbol{\beta}_0) + o_p(1). \qquad \text{(A.4)}$$

Hence $n^{1/2}(\hat{\boldsymbol{\beta}}_{\mathrm{w}^*} - \boldsymbol{\beta}_0)$ has an asymptotically normal distribution with mean 0 and covariance matrix $V_{\mathrm{w}^*} = [A_{\mathrm{w}^*}(\boldsymbol{\beta}_0)]^{-1} \Gamma_{\mathrm{w}^*} [A_{\mathrm{w}^*}(\boldsymbol{\beta}_0)]^{-1}$.

Replacing $\boldsymbol{\beta}_0$, G, $y(t)$ and $d\Lambda_C(t)$ by the corresponding estimates, $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$, $\hat{G}$, $\bar{Y}(t)/n$ and $dN_C(t)/\bar{Y}(t)$, where $N_C(t) = \sum_k I(X_k \le t, B_k = 0)$, respectively, $\hat{\xi}_i$ equals

$$\left\{ \left[\frac{I(X_i \le \tau, B_i = 1)}{\hat{G}(X_i)} - \pi(\tilde{\mathbf{Z}}_i^T \hat{\boldsymbol{\beta}}_{\mathrm{w}^*}) \right] (\hat{\mathrm{v}}_{2i} - \hat{\mathrm{v}}_{3i}) \right.$$
$$\left. - \left[\frac{I(X_i > \tau)}{\hat{G}(\tau+)} + \frac{I(X_i \le \tau, B_i = 2)}{\hat{G}(X_i)} - \bar{\pi}(\tilde{\mathbf{Z}}_i^T \hat{\boldsymbol{\beta}}_{\mathrm{w}^*}) \right] (\hat{\mathrm{v}}_{1i} - \hat{\mathrm{v}}_{3i}) \right\} \frac{\pi_\phi(\tilde{\mathbf{Z}}_i^T \hat{\boldsymbol{\beta}}_{\mathrm{w}^*})}{\hat{\mathrm{v}}_{1i}\hat{\mathrm{v}}_{2i} - \hat{\mathrm{v}}_{3i}^2} \tilde{\mathbf{Z}}_i$$
$$+ \frac{nI(B_i = 0)\hat{q}(X_i; \hat{\boldsymbol{\beta}}_{\mathrm{w}^*})}{\sum_{k=1}^n I(X_k \ge X_i)} - \sum_{j=1}^n \frac{nI(B_j = 0, X_i \ge X_j)\hat{q}(X_j; \hat{\boldsymbol{\beta}}_{\mathrm{w}^*})}{\left(\sum_{k=1}^n I(X_k \ge X_j)\right)^2},$$

where $\hat{\mathrm{v}}_{1i} = \pi(\tilde{\mathbf{Z}}_i^T \hat{\boldsymbol{\beta}}_{\mathrm{w}^*})(M_G - \pi(\tilde{\mathbf{Z}}_i^T \hat{\boldsymbol{\beta}}_{\mathrm{w}^*}))$, $\hat{\mathrm{v}}_{2i} = \bar{\pi}(\tilde{\mathbf{Z}}_i^T \hat{\boldsymbol{\beta}}_{\mathrm{w}^*})(M_G - \bar{\pi}(\tilde{\mathbf{Z}}_i^T \hat{\boldsymbol{\beta}}_{\mathrm{w}^*}))$,

$$\hat{\mathrm{v}}_{3i} = \bar{\pi}(\tilde{\mathbf{Z}}_i^T \hat{\boldsymbol{\beta}}_{\mathrm{w}^*})\pi(\tilde{\mathbf{Z}}_i^T \hat{\boldsymbol{\beta}}_{\mathrm{w}^*}),\ \hat{q}(t; \hat{\boldsymbol{\beta}}_{\mathrm{w}^*}) = \hat{q}_1(t; \hat{\boldsymbol{\beta}}_{\mathrm{w}^*}) - \hat{q}_2(t; \hat{\boldsymbol{\beta}}_{\mathrm{w}^*}) - \hat{q}_3(t; \hat{\boldsymbol{\beta}}_{\mathrm{w}^*}),$$

and $\hat{q}_j(t; \hat{\boldsymbol{\beta}}_{\mathrm{w}^*})$ $(j = 1, 2, 3)$ are obtained by using $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$, $\hat{G}$ and $(\hat{\mathrm{v}}_{1k}, \hat{\mathrm{v}}_{2k}, \hat{\mathrm{v}}_{3k})$ instead of $\boldsymbol{\beta}_0$, G and $(\mathrm{v}_{1k}, \mathrm{v}_{2k}, \mathrm{v}_{3k})$ in (A.1)−(A.3). It follows that the covariance matrix Γ_{w^*} can be estimated by $\hat{\Gamma}_{\mathrm{w}^*} = n^{-1} \sum_{i=1}^n \hat{\xi}_i \hat{\xi}_i^T$ and then

$$\hat{V}_{\mathrm{w}^*} = \left[\hat{A}_{\mathrm{w}^*}(\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}) \right]^{-1} \hat{\Gamma}_{\mathrm{w}^*} \left[\hat{A}_{\mathrm{w}^*}(\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}) \right]^{-1}$$

where

$$\hat{A}_{\mathrm{w}^*}(\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}) = \sum_{i=1}^n \frac{1}{n} \left[\frac{\hat{\mathrm{v}}_{1i} + \hat{\mathrm{v}}_{2i} - 2\hat{\mathrm{v}}_{3i}}{\hat{\mathrm{v}}_{1i}\hat{\mathrm{v}}_{2i} - \hat{\mathrm{v}}_{3i}^2} \pi_\phi^2(\tilde{\mathbf{Z}}_i^T \hat{\boldsymbol{\beta}}_{\mathrm{w}^*}) \tilde{\mathbf{Z}}_i \tilde{\mathbf{Z}}_i^T \right].$$

Appendix C: *Previous nonparametric results of Wang (2003)*

Modifying the idea of Wang (2003), we can estimate $p_j(x) = \Pr(T \le \tau, \tilde{B} = j | T > x)$ by

$$\hat{p}_j(x) = \frac{1}{n\hat{S}(x)} \sum_{i=1}^n \frac{I(x < X_i \le \tau, B_i = j)}{\hat{G}(X_i)},$$

where $\hat{S}(x)$ is the Kaplan-Meier estimator of $S(x)$ which, according to Satten and Datta (2001), can be re-expressed as an average of inverse probability of censoring given by

$$\frac{1}{n} \sum_{i=1}^n \left[\frac{I(X_i > x, B_i \ne 0)}{\hat{G}(X_i)} + \frac{I(X_i > X_{(m)})}{\hat{G}(X_{(m)}+)} \right],$$

where $X_{(m)}$ denotes the largest observed failure time. Based on Wang's idea, $Q_j(t|\tau)$ can be estimated by

$$\prod_{u\le t}\left\{1-\frac{\sum_{i=1}^{n} I(u=X_i\le\tau, B_i=j)}{\sum_{i=1}^{n}\left[I(u\le X_i\le\tau, B_i=j)+I(u\le X_i\le\tau, B_i=0)\hat{p}_j(X_i)\right]}\right\}.$$

Appendix D: *Asymptotic properties of* $U_{I1}(\boldsymbol{\beta})$

Suppose that $\tilde{\mathbf{Z}}$ takes K distinct values, $z_1,\ldots,z_K$. Original data are partitioned into K mutually exclusive subsets, $\left\{\left(\Delta_{1k}^j, X_k^j, B_k^j\right)\ (k=1,\ldots,n_j)\right\}$, which corresponds to the set of $\{i:(\Delta_{1i}, X_i, B_i, \tilde{\mathbf{Z}}_i=z_j)\ (i=1,\ldots,n)\}$ and $n_j=\sum_{i=1}^{n} I(\tilde{\mathbf{Z}}_i=z_j)$. We have $p_{z_j}(X_k^j)=E(\Delta_{1k}^j\big|\, X_k^j, B_k^j=0, \tilde{\mathbf{Z}}=z_j)$, which can be estimated by

$$\hat{p}_{z_j}(X_k^j)=\frac{1}{n_j\hat{S}_{z_j}(X_k^j)}\sum_{h=1}^{n_j}\frac{I(X_k^j<X_h^j\le\tau, B_h^j=1)}{\hat{G}_{z_j}(X_h^j)},$$

where $\hat{S}_{z_j}(t)$ and $\hat{G}_{z_j}(t)$ are Kaplan-Meier estimators of $S_{z_j}(t)=\Pr(T>t|\tilde{\mathbf{Z}}=z_j)$ and $G_{z_j}(t)=\Pr(C\ge t|\tilde{\mathbf{Z}}=z_j)$. The estimating equation $U_{I1}(\boldsymbol{\beta})$ can be re-expressed as

$$U_{I1}(\boldsymbol{\beta})=\sum_{j=1}^{K}\left\{\sum_{k=1}^{n_j}\left[\hat{\Delta}_{1k}^j-\pi(z_j^T\boldsymbol{\beta})\right]\frac{\pi_\phi(z_j^T\boldsymbol{\beta})}{\pi(z_j^T\boldsymbol{\beta})\bar{\pi}(z_j^T\boldsymbol{\beta})}z_j\right\},$$

where $\hat{\Delta}_{1k}^j=I(B_k^j=1, X_k^j\le\tau)+I(B_k^j=0, X_k^j\le\tau)\hat{p}_{z_j}(X_k^j)$.

To derive asymptotic distribution of $n^{-1/2}U_{I1}(\boldsymbol{\beta}_0)$, we first express it as sum of the following two terms,

$$\begin{aligned}\frac{1}{\sqrt{n}}U_{I1}(\boldsymbol{\beta}_0) &= \sum_{j=1}^{K}\sqrt{\frac{n_j}{n}}\left\{\frac{1}{\sqrt{n_j}}\sum_{k=1}^{n_j}\left[E_k^j-\pi(z_j^T\boldsymbol{\beta}_0)\right]\Psi_{z_j}(\boldsymbol{\beta}_0)\right\}\\ &+ \sum_{j=1}^{K}\sqrt{\frac{n_j}{n}}\left\{\frac{1}{\sqrt{n_j}}\sum_{k=1}^{n_j} I(B_k^j=0)\left[\hat{p}_{z_j}(X_k^j)-p_{z_j}(X_k^j)\right]\Psi_{z_j}(\boldsymbol{\beta}_0)\right\} \quad \text{(A.5)}\end{aligned}$$

where

$$E_k^j=E\left(\Delta_{1k}^j\big|\, X_k^j, B_k^j; \tilde{\mathbf{Z}}=z_j\right)=I(B_k^j=1, X_k^j\le\tau)+I(B_k^j=0, X_k^j\le\tau)p_{z_j}(X_k^j)$$

and

$$\Psi_{z_j}(\boldsymbol{\beta}_0)=\frac{\pi_\phi(z_j^T\boldsymbol{\beta}_0)}{\pi(z_j^T\boldsymbol{\beta}_0)\bar{\pi}(z_j^T\boldsymbol{\beta}_0)}z_j.$$

Denote the last part of (A.5) by $C_2(\boldsymbol{\beta}_0)$, by the strong consistency of Kaplan-Meier estimators, we have

$$C_2(\boldsymbol{\beta}_0) \;=\; \sum_{j=1}^{K}\left\{\sqrt{\frac{n_j}{n}}\Psi_{z_j}(\boldsymbol{\beta}_0)\left[C_{2.1}^j + C_{2.2}^j\right]\right\} + o_p(1),$$

where

$$C_{2.1}^j = \frac{1}{\sqrt{n_j}}\sum_{k=1}^{n_j}\left[\frac{I(B_k^j=0)}{n_j S_{z_j}(X_k^j)}\sum_{h=1}^{n_j} I(X_k^j < X_h^j \le \tau, B_h^j = 1)\left(\frac{1}{\hat{G}_{z_j}(X_h^j)} - \frac{1}{G_{z_j}(X_h^j)}\right)\right],$$

$$C_{2.2}^j = \frac{1}{\sqrt{n_j}}\sum_{k=1}^{n_j}\left[I(B_k^j=0)\left(\frac{1}{\hat{S}_{z_j}(X_k^j)} - \frac{1}{S_{z_j}(X_k^j)}\right)\frac{1}{n_j}\sum_{h=1}^{n_j}\left(\frac{I(X_k^j < X_h^j \le \tau, B_h^j = 1)}{G_{z_j}(X_h^j)}\right)\right].$$

Interchanging the summations in $C_{2.1}^j$, we get

$$C_{2.1}^j = \frac{1}{\sqrt{n_j}}\sum_{h=1}^{n_j}\left[D(X_h^j)\frac{I(X_h^j \le \tau, B_h^j = 1)}{G_{z_j}(X_h^j)}\frac{\hat{G}_{z_j}(X_h^j) - G_{z_j}(X_h^j)}{G_{z_j}(X_h^j)}\right] + o_p(1)$$

where

$$D(X_h^j) \;=\; \lim_{n_j\to\infty}\frac{1}{n_j}\sum_{k=1}^{n_j}\frac{I(B_k^j = 0, X_k^j < X_h^j)}{S_{z_j}(X_k^j)}.$$

One can write

$$\frac{\hat{G}_{z_j}(t) - G_{z_j}(t)}{G_{z_j}(t)} = \sum_{l=1}^{n_j}\int_0^t \frac{\hat{G}_{z_j}(u-)}{G_{z_j}(u)}\frac{dM_{C,l}^j(u)}{\bar{Y}^j(u)}$$

where

$$\bar{Y}^j(u) = \sum_{i=1}^{n_j} I(X_i^j \ge u), \quad M_{C,l}^j(u) = I(X_l^j \le u, B_l^j = 0) - \int_0^u I(X_l^j \ge s)d\Lambda_C^j(s),$$

and $\Lambda_C^j(s)$ is the cumulative hazard function of C given $\tilde{\mathbf{Z}} = z_j$. It follows that

$$C_{2.1}^j = \frac{1}{\sqrt{n_j}}\sum_{l=1}^{n_j}\int_0^\infty \frac{q^j(u)}{p^j(u)}dM_{C,l}^j(u) + o_p(1),$$

where

$$q^j(u) = \lim_{n_j\to\infty}\frac{1}{n_j}\sum_{h=1}^{n_j} D(X_h^j)\frac{I(u \le X_h^j \le \tau, B_h^j = 1)}{G_{z_j}(X_h^j)} \quad \text{and} \quad p^j(u) = \lim_{n_j\to\infty}\frac{\bar{Y}^j(u)}{n_j}.$$

Similarly, one can write

$$C_{2.2}^{j} = \frac{1}{\sqrt{n_j}} \sum_{l=1}^{n_j} \int_0^{\infty} \frac{r^j(u)}{p^j(u)} \, dM_{T,l}^{j}(u) + o_p(1),$$

where

$$r^j(u) = \lim_{n_j \to \infty} \frac{1}{n_j} \sum_{k=1}^{n_j} \frac{I(B_k^j = 0, X_k^j \geq u) P_{z_j}(X_k^j)}{S_{z_j}(X_k^j)},$$

$$P_{z_j}(X_k^j) = \lim_{n_j \to \infty} \frac{1}{n_j} \sum_{h=1}^{n_j} \frac{I(B_h^j = 1, X_k^j < X_h^j \leq \tau)}{G_{z_j}(X_h^j)},$$

and

$$M_{T,l}^{j}(u) = I(X_l^j \leq u, B_l^j \neq 0) - \int_0^u I(X_l^j \geq s) d\Lambda_T^j(s),$$

$\Lambda_T^j(s)$ is the cumulative hazard function of T given $\tilde{\mathbf{Z}} = z_j$.

In summary, we have

$$\frac{1}{\sqrt{n}} U_{I1}(\boldsymbol{\beta}_0) = \sum_{j=1}^{K} \sqrt{\frac{n_j}{n}} \left(\frac{1}{\sqrt{n_j}} \sum_{k=1}^{n_j} \zeta_k^j \right) \Psi_{z_j}(\boldsymbol{\beta}_0) + o_p(1)$$

where

$$\zeta_k^j = E_k^j - \pi(z_j^T \boldsymbol{\beta}_0) + \int_0^{\infty} \frac{q^j(u)}{p^j(u)} dM_{C,k}^{j}(u) + \int_0^{\infty} \frac{r^j(u)}{p^j(u)} dM_{T,k}^{j}(u).$$

Notice that $(\zeta_1^j, \ldots, \zeta_{n_j}^j)$ are zero-mean independent random variables for each j where $j = 1, \ldots, K$. By the multivariate central limit theorem, $\frac{1}{\sqrt{n}} U_{I1}(\boldsymbol{\beta}_0)$ has an asymptotical normal distribution with mean 0 and covariance matrix

$$\Gamma_{I1} = \lim_{n \to \infty} n^{-1} \sum_{j=1}^{K} \sum_{k=1}^{n_j} (\zeta_k^j)^2 \; \Psi_{z_j}(\boldsymbol{\beta}_0) \Psi_{z_j}^T(\boldsymbol{\beta}_0).$$

Let $\hat{\boldsymbol{\beta}}_{I1}$ be the solution of $U_{I1}(\boldsymbol{\beta}) = 0$. Asymptotic properties of $\hat{\boldsymbol{\beta}}_{I1}$ can be obtained as of $\hat{\boldsymbol{\beta}}_{\mathrm{w}^*}$ stated in Appendix B. According to (A.4), $n^{1/2}(\hat{\boldsymbol{\beta}}_{I1} - \boldsymbol{\beta}_0)$ has an asymptotically normal distribution with mean 0 and covariance matrix $V_{I1} = [A_{I1}(\boldsymbol{\beta}_0)]^{-1} \Gamma_{I1} [A_{I1}(\boldsymbol{\beta}_0)]^{-1}$ where

$$A_{I1}(\boldsymbol{\beta}_0) = E \left[\frac{\pi_\phi^2(\tilde{\mathbf{Z}}^T \boldsymbol{\beta}_0)}{\pi(\tilde{\mathbf{Z}}^T \boldsymbol{\beta}_0) \bar{\pi}(\tilde{\mathbf{Z}}^T \boldsymbol{\beta}_0)} \tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^T \right].$$

References

Betensky, R. A. and Schoenfeld, D. A. (2001) Nonparametric estimation in a cure model with random cure times. *Biometrics*, **57**, 282–286.

Buckley, J. and James, I. (1979) Linear regression with censored data. *Biometrika*, **66**, 429–436.

Chen, C. F. (1983) Score tests for regression models. *Journal of the American Statistical Association*, **78**, 158–161.

Cheng, S. C., Wei, L. J. and Ying, Z (1995). Analysis of transformation models with censored data. *Biometrika*, **82**, 835–845.

Cheng, S. C., Fine, J. P. and Wei, L. J. (1998) Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*, **54**, 219–228.

Chen, Y. Q., Jewell, N. P., Lei, X. and Cheng, S. C. (2005) Semiparametric Estimation of Proportional Mean Residual Life Model in Presence of Censoring. *Biometrics*, **61**, 170–178.

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

Crowley, J. and Hu, M. (1977) Covariance analysis of heart transplant survival data. *J. Amer. Statist. Ass.*, **72**, 27–36

Dabrowska, D. M. (1987) Non-parametric regression with censored survival time data. *Scand. J. Statist.*, **14**, 181–197.

Dempster, A., Laird, N. and Rubin, D. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc. B*, **39**, 1–38.

Fine, J. P. (1999) Analyzing competing risks data with transformation models. *J. R. Statist. Soc.* B, **61**, 817–830.

Fine, J. P. and Gray, R. J. (1999) A proportional hazards model for the subdistribution of a competing risk. *J. Am. Statist. Ass.*, **94**, 496–509.

Fine, J. P., Ying, Z. and Wei, L. J. (1998) On the linear transformation model for censored data. *Biometrika*, **85**, 980–986.

Fine, J. P. (2001) Regression modeling of competing crude failure probabilities. *Biostatistics*, **2**, 85–97.

Ghitany, M. E., Maller, R. A. and Zhou, X. (1994) Exponential mixture models with long term survivors and covariates. *J. Multivariate Anal.*, **49**, 218–241.

Heyde, C. C. (1997) *Quasi-Likelihood and its Application : A General Approach to Optimal Parameter Estimation*. New York: Springer.

Jeong, J. H. and Fine, J. P. (2006) Parametric regression on cumulative incidence function. *Biostatistics*, **0**, 1–13

Jung, S. H. (1996) Regression analysis for long-term survival rate. *Biometrika*, **83**, 227–232.

Klein, J. P. and Andersen, P. K. (2005) Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, **61**, 223–229.

Klien, J. P. and Moeschberger, M. L. (2003) *Survival Analysis, Techniques for Censored and Truncated Data*, 2nd Ed. New York: Springer.

Kuk, A. Y. C. (1992) A semiparametric mixture model for the analysis of competing risks data. *Austral. J. Statist.*, **34**(2), 169–180.

Larson, M. G. and Dinse, G. E. (1985) A mixture model for the regression analysis of competing risks data. *Applied Statistics*, **34**, 201–211.

Li, K. C., Wang, J. L. and Chen, C. H. (1999) Dimension Reduction for Censored Regression Data. *Annals of Statistics*, **27**, 1–23.

Li, W. A. (1991) Testing model adequacy for some Markov regression models for time series. *Biometrika*, **78**, 83–89.

Louis, T. (1982) Finding the observed information matrix using the EM algorithm. *J. R. Statist. Soc. B*, **44**, 226–233.

Maller, R. A. and Zhou, X. (1996) Survival Analysis with Long-term Survivors. New York: Wiley.

Maller, R. A. and Zhou, X. (2002) Analysis of parametric models for competing risks. *Statistica Sinica*, **12**, 725–750.

Ng, S. K. and McLachlan, G. J. (2003) An EM-based semi-parametric mixture model approach to the regression analysis of competing-risks data. *Statistics in Medicine*, **22**, 1097–1111.

Pepe, M. S. and Mori, M. (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine*, **12**, 737–751.

Pregibon, D. (1980) Goodness of Link Tests for Generalized Linear Models. *Applied Statistics*, **29**, 15–24.

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

Satten, G. A. and Datta, S. (2001) The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *The American Statistician*, **55**, 207–210.

Subramanian, S. (2001) Parameter estimation in regression for long-term survival rate from censored data. *Journal of Statistical Planning and Inference*, **99**, 211–222.

Vu, H. T. V., Maller, R. A. and Zhou, X. (1998) Mixture models for failure data. *Ann. Inst. Statist. Math.*, **50**, 627–653.

Wang, W. (2003) Nonparametric estimation of the sojourn time distributions for a multipath model. *J. R. Statist. Soc. B*, **65**, 921–935.

Wu, C. (1983) On the convergence properties of the EM algorithm. *Ann. Statist.*, **11**, 95–103.