Abstract

GAO, GUOZHI. Semiparametric Estimators for the Regression Coefficients in the Linear Transformation Competing Risks Models with Missing Cause of Failure. (Under the direction of Dr. Anastasios A. Tsiatis.)

In many clinical studies, researchers are mainly interested in studying the effects of some prognostic factors on the hazard of failure from a specific cause while individuals may failure from multiple causes. This leads to a competing risks problem. Often, due to various reasons such as finite study duration, loss to follow-up, or withdrawal from the study, the time-to-failure is right-censored for some individuals. Although the proportional hazards model has been commonly used in analyzing survival data, there are circumstances where other models are more appropriate. Here we consider the class of linear transformation models that contains the proportional hazards model and the proportional odds model as special cases. Sometimes, patients are known to die but the cause of death is unavailable. It is well known that when cause of failure is missing, ignoring the observations with missing cause or treating them as censored may result in erroneous inferences. Under the Missing At Random assumption, we propose two methods to estimate the regression coefficients in the linear transformation models. The augmented inverse probability weighting method is highly efficient and doubly robust. In addition, it allows the possibility of using auxiliary covariates to model the missing mechanism. The multiple imputation method is very efficient, is straightforward and easy to implement and also allows for the use of auxiliary covariates. The asymptotic properties of these estimators are developed using theory of counting processes and semiparametric theory for missing data problems. Simulation studies demonstrate the relevance of the theory in finite samples. These methods are also illustrated using data from a breast cancer stage II clinical trial.

KEY WORDS: Cause-specific hazard; Competing risks; Double Robustness; Influence function; Inverse probability weighted; Linear transformation model; Multiple Imputation; Missing at random; Semiparametric estimator.

Semiparametric Estimators for the Regression Coefficients in the Linear Transformation Competing Risks Models with Missing Cause of Failure

by

Guozhi Gao

A Dissertation

submitted to the advisory committee on graduate studies of

North Carolina State University

in partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

DEPARTMENT OF STATISTICS

Raleigh, NC

August, 2005

APPROVED BY:

Anastasios A. Tsiatis Chair of Advisory Committee	Marie Davidian
Zhang Daowen	Lu Wenbin

To my parents, my brother, my sister, and Yunyun

Biography

Guozhi Gao was born in Guangzhou, GuangDong, China to parents Jintai Gao and Ruirong Chen on July 17, 1977. He received his B.S. in Applied Mathematics from South China University of Technology in June, 2000. He then spent two years in Texas Tech University and received a M.S. in Statistics in May, 2002. Thereafter, he joined the Ph.D program in Statistics at North Carolina State University. Upon completion of his doctoral degree, he will start working at Amgen as a senior biostatistician and begin his career in the biotechnology industry.

Acknowledgements

I would like to express my deepest gratitude and appreciation to my thesis advisor, Dr. Anastasios A. Tsiatis, for his constant help, insightful inspiration and careless encouragement. Not only being one of the greatest researchers in Biostatistics, he is also one of the most respectable people one could ever associate with. It would be almost impossible for me to complete my doctorate degree without his guidance.

I would also like to thank Drs. Bill Swallow and Leonard A. Stefanski for their continuous support of my graduate study, Dr. Marie Davidian for her useful tips on LaTeX and helpful advice on job opportunities, Dr. Wenbin Lu for his numerous help on linear transformation models and the theory of empirical processes, Dr. Daowen Zhang for his insightful explanation of model diagnostics using smoothing splines. I am also grateful to Terry Byron for his assistance on computing, Adrian Blue for his assistance toward graduation. My thanks also go to other faculty, stuff and students in our department. I would also thank Dr. Steven Rebach for his kindly agreement to represent the graduate school on my committee.

Special thanks go to Yunyun. Finally and foremost, I would like to thank my parents, my brother and sister and their families for their constant support and love that will support all the way in my whole life.

Contents

List of Tables

List of Abbreviations

BCC	Biased Complete Case
CP	Coverage Probability
DR	Doubly Robust
GR	Goetghebeur and Ryan
IPW	Inverse Probability Weighted
IPWCC	Inverse Probability Weighted Complete-Case
LT	Lu and Tsiatis (2001), Multiple Imputation
LTM	Linear Transformation Models
MAR	Missing At Random
MI	Multiple Imputation
PH	Proportional Hazards model
PO	Proportional Odds model
SEE	Standard Error Estimate
SI	Single Imputation
SSE	Sampling Standard error Estimate

Chapter 1

Inverse Probability Weighting Approaches

1.1 Introduction

In clinical trials, patients may fail from one of multiple causes but often interest focuses on one of the causes. For instance, in a breast cancer study, interest may focus on death from breast cancer although patients may die from other causes. In these situations, it is natural to consider models for the relationship of the cause-specific hazard as a function of covariates. From competing risks theory we also know that if we cast the problem through a set of competing risks, each of which has a potential time to failure for which we only get to observe the minimum and failure type of the minimum, then there is a one-to-one correspondence between the cause-specific hazard function and the hazard function for the potential failure time if it is assumed that the potential failure times (including censoring) are conditionally independent given the covariates. Thus, when we are considering models for the cause-specific hazard for a particular cause, as a function of the covariates, then this is equivalent to considering models for the relationship of the potential failure time of that cause given the covariates. The most common model is the proportional hazards model (Cox, 1972, 1975), which has been extensively studied and commonly used on survival data. However, as shown by many authors, in some survival studies, the proportional hazards model may not be suitable for modeling survival times, and alternative models may be more appropriate. For instance, the proportional odds model (Pettitt 1982, 1984; Bennett, 1983; Murphy et al., 1997) is preferable to the proportional hazards model, if the hazard functions for the two treatment groups converge to the same limit. The linear transformation model is a class of flexible models that have been considered recently in the literature which contain both the proportional hazards and proportional odds models as special cases. See Cheng et al., 1995, 1997; Fine et al., 1998 for detail. Because of the correspondence between the cause-specific hazard functions and the hazard functions of the potential failure times in a competing risks model, one can estimate the relationship of the cause-specific hazard function for the cause of interest, as a function of covariates, by using censored data methods where all other causes of failure and censoring time are combined into a single censoring variable. Cheng et al. (1995, 1997), Fine et al. (1998) and Cai et al. (2000) proposed and further developed a general estimation method for linear transformation models with censored data. However, these methods are based on the assumption that censoring is independent of covariates. In the competing risks setting this would mean that not only would censoring have to be independent of the covariates but also all cause-specific hazard functions for other causes of failure would also have to be independent of covariates. Such a restrictive assumption is not likely to hold in practice. Chen et al. (2002) gave a general estimation procedure that did not require the independence assumption above. Their procedure, which reduces to the Cox partial likelihood estimator (Cox, 1975) for the proportional hazards model, is the solution to estimating equations that take advantage of the martingale structure of the model and the resulting estimator was proved to be consistent and asymptotically normal. In addition, their algorithm is relatively easy to implement. Consequently, if cause of failure information were known for all patients that were uncensored, then, as mentioned above, the methods of Chen et al.

(2002) could be used to derive estimates for the regression parameters in a model where we assumed that the cause-specific hazard function for a particular cause of interest, as a function of covariates, followed the linear transformation model by simply combining all other causes of failure together with censoring into a single censoring variable.

One difficulty that occurs in many clinical trials is that the time to failure might be recorded but cause of failure may not be easy to determine and is therefore missing, e.g., whether failure is attributable to the cause of interest or other causes may require documentation with information that is not collected or lost or cause may be difficult for investigators to determine for certain patients (Andersen, Goetghebeur, and Ryan, 1996). It has been shown that excluding the observations whose causes are missing may result in biased estimates. Hence the method proposed by Chen et al. (2002) can not be directly applied in the presence of missing cause of failure. Our goal is to develop methods for estimating the regression parameters in arbitrary linear transformation models for the cause-specific hazard when some of the causes of failure are missing. Lu and Tsiatis (2001) used parametric models to model the probability that the missing cause is the cause of interest and then estimated the regression parameters by using a multiple imputation method (Rubin, 1987, 1996). In this paper, we derive inverse probability weighted complete-case estimators, including augmented inverse probability weighted complete-case estimators that are doubly-robust (in a manner which we will describe later), for the general linear transformation model.

In the next section, we state notation and assumptions. In §1.3 we develop the estimating equations for the inverse probability weighted and augmented inverse probability weighted complete-case estimators. A computational algorithm is presented in §1.4. We establish the properties of the proposed estimators in §1.5. Some simulation studies are presented in §1.6. In §1.7, the method is illustrated using data from a breast cancer study, followed by a brief discussion in §1.8. Finally, a proof of some asymptotic properties of the estimator is given in the Appendix.

1.2 Notation and assumptions

For simplicity, and without loss of generality, we consider only two particular causes that will lead to failure of individuals. We label the cause-of-interest as cause 2 and the other cause as cause 1. If there was no censoring, then the data could be summarised as (U^*, Δ^*), where U^* denotes the time to failure and Δ^* denotes the cause of failure, taking values 1 or 2. We also define Z to be the q-dimensional vector of regression covariates for modeling the cause-specific hazard for the cause-of-interest to these covariates; namely

$$\lambda_2^*(u|z) = \lim_{h\to 0} h^{-1}\mathrm{pr}(u \le U^* < u+h, \Delta^* = 2|U^* \ge u, Z = z). \tag{1.1}$$

If we take the point of view that there are potential failure times T_1 and T_2 associated with causes 1 and 2, respectively, and that $U^* = \min(T_1, T_2)$, $\Delta^* = 1$ if $U^* = T_1$ and $\Delta^* = 2$ if $U^* = T_2$, then it is well known that if T_1 is assumed to be conditionally independent of T_2 given Z, then the hazard function of T_2 given Z is the same as the cause-specific hazard function $\lambda_2^*(u|z)$ given in (1.1). This representation allows us to consider models that have been developed for the hazard function of the survival time T_2 as a function of Z to be used for modeling the cause-specific hazard function in a competing risks model.

Recently, a popular class of models for this relationship is the linear transformation model that is defined as

$$H(T_2) = -\beta^T Z + \varepsilon, \tag{1.2}$$

where H is some unknown monotone increasing function, β is a q-dimensional regression parameter of interest, and ε is a random error term with a known continuous distribution that is assumed to be independent of Z. The models given by (1.2) include many common statistical models in survival analysis. For instance, (1.2) reduces to the proportional hazards model if ε follows the extreme value distribution; and (1.2) becomes the proportional odds model when ε has the standard logistic distribution. For a more detailed description of the class of linear transformation models, see Cheng et al. (1995). The linear transformation model induces a relationship for the hazard function of T_2 given Z as a function of parameters β, which, by the correspondence above, will be the relationship for the cause-specific hazard function $\lambda_2^*(u|z)$ in our competing risks model.

Because of incomplete follow-up, cause of failure data are often censored by a variable C; in which case, the data we observe can be summarised as (T, Δ) where $T = \min(U^*, C)$ and Δ is the failure-censoring indicator that is equal to 0, 1, and 2. Specifically, $\Delta = \Delta^*$ if $U^* \leq C$, and $\Delta = 0$ otherwise. To avoid non-identifiability problems, C is assumed to be conditionally independent of (U^*, Δ^*) given Z. Under this assumption, the observable cause-specific hazard functions for cause 1 and 2, in the presence of censoring, defined as

$$\lambda_j(u|z) = \lim_{h \to 0} h^{-1} \text{pr}(u \leq T < u + h, \Delta = j | T \geq u, Z = z), \quad j = 1, 2,$$

are the same as the cause-specific hazards of interest. In particular $\lambda_2(u|z) = \lambda_2^*(u|z)$. With a sample of data (T_i, Δ_i, Z_i), $i = 1, \cdots, n$, the parameter β in the linear transformation model can be estimated using censored methods like the one proposed by Chen et al. (2002).

If a patient dies and cause of failure is unknown, classification is uncertain, and in the literature there is currently no suitable method for dealing with missing cause of failure information for the linear transformation models. We begin by defining R to be the complete case indicator; that is, $R = 1$ if Δ is known and $R = 0$ otherwise. We wish to note that because the issue of missing cause of death doesn't apply when an individual is censored, we always define $R = 1$ whenever $\Delta = 0$.

In addition to the covariates Z which are used in the linear transformation model, we also introduce auxiliary covariates A that are not used to model the hazards, but may be used to describe the missing mechanism. Therefore, with the possibility of missing cause of death, the data that are observed can be summarised as $O_i = \{R_i, T_i, Z_i, A_i, I(\Delta_i = 0), R_i I(\Delta_i = 1), R_i I(\Delta_i = 2)\}$, which are independent across i.

We assume that the cause of failure is missing at random (Rubin, 1976). That is, given $\Delta_i > 0$ and $W_i = (T_i, Z_i, A_i)$, the probability that the cause of failure is missing for the ith individual depends only on the observed W_i, but not on the unobserved Δ_i. In other words,

$$\text{pr}(R_i = 1|\Delta_i, \Delta_i > 0, W_i) = \text{pr}(R_i = 1|\Delta_i > 0, W_i) = \pi(W_i). \tag{1.3}$$

Since $\Delta_i = 0$ implies $R_i = 1$, we have equivalently,

$$\text{pr}(R_i = 1|\Delta_i, W_i) = \pi(W_i)I(\Delta_i > 0) + I(\Delta_i = 0) = \pi(Q_i), \tag{1.4}$$

where $Q_i = \{W_i, I(\Delta_i > 0)\}$. Although the model of interest only involves the covariates Z_i, the auxiliary variables A_i are also introduced as they may be necessary to make the missing at random assumption tenable.

1.3 Estimating Equations

If there is no missing cause of death information, we define the counting process $N_i(t) = I(\Delta_i = 2)I(T_i \leq t)$ and the at-risk indicator process $Y_i(t) = I(T_i \geq t)$. By so doing, one will obtain the corresponding martingale process $M_i(t) = N_i(t) - \int_0^t Y(u) d\Lambda\{H_0(u) + \beta_0^T Z_i\}$, where Λ is the cumulative hazard function of ε in (1.2), and H_0 and β_0 denote true values of H and β, respectively. This martingale structure coupled with the counting process theory (Fleming and Harrington, 1991) led Chen et al. (2002) to the following estimating equations for jointly estimating (β, H):

$$\sum_{i=1}^{n} \int_0^\infty Z_i[dN_i(t) - Y_i(t)d\Lambda\{H(t) + \beta^T Z_i\}] \equiv \sum_{i=1}^{n} \int_0^\infty Z_i dM_i(t, Z_i, \beta, H) = 0, \tag{1.5}$$

$$\sum_{i=1}^{n}[dN_i(t) - Y_i(t)d\Lambda\{H(t) + \beta^T Z_i\}] \equiv \sum_{i=1}^{n} dM_i(t, Z_i, \beta, H) = 0, \quad t \geq 0. \tag{1.6}$$

By taking advantage of the martingale representation of (1.5) and (1.6), Chen et. al. (2002) were able to prove that their estimator was asymptotically normal with limiting variance-covariance matrix that was expressed explicitly.

In the presence of missing cause of failure, complete data are not observed, and consequently (1.5) and (1.6) can not be applied directly. In the following we propose two methods to deal with this problem.

1.3.1 Inverse Probability Weighted Complete-Case Estimator

We first propose the inverse probability weighted complete-case estimator that has been commonly used in missing data problems. Toward that end, we need to derive an estimator for the probability of a complete-case, i.e. $R_i = 1$ as a function of the data that are observed on everyone; i.e. $\{I(\Delta_i > 0), W_i\}$. Therefore, for the probability of a complete-case $\pi(W_i) = \text{pr}(R_i = 1|\Delta_i > 0, W_i)$, we posit a parametric model $\pi(W, \psi)$ in terms of a finite dimensional parameter ψ, which satisfies $\pi(W) = \pi(W, \psi_0)$. Since R_i is binary, a logistic regression model is often used, but other parametric models can also be easily accommodated. By (1.3), the maximum likelihood estimator $\hat{\psi}_n$ of ψ is obtained by maximizing the observed-data likelihood

$$\prod_{i=1}^{n}\{\pi(W_i, \psi)\}^{R_i I(\Delta_i > 0)}\{1 - \pi(W_i, \psi)\}^{1-R_i}. \tag{1.7}$$

Under suitable regularity conditions, $\hat{\psi}_n$ converges to some limit, denoted by ψ^*. If the model $\pi(Q, \psi) = \pi(W_i, \psi)I(\Delta_i > 0) + I(\Delta_i = 0)$ is correctly specified, then we denote this by taking $\psi^* = \psi_0$, where $\pi(Q, \psi_0) = \text{pr}(R = 1|Q)$.

To avoid problems associated with the tails of the distributions, we will truncate time at some finite value τ, where $\text{pr}(T > \tau|Z) > 0$ and $\text{pr}(C > \tau|Z) > 0$ with probability one. With missing cause of failure data, the usual inverse probability weighted complete-case theory naturally leads us to modify the estimating equations (1.5) and (1.6) by considering the following inverse probability weighted complete-case estimating equations for $(\beta,\, H)$,

$$\sum_{i=1}^{n} \int_0^{\tau} \frac{R_i}{\pi(Q_i, \hat{\psi}_n)} Z_i \left[dN_i(t) - Y_i(t) d\Lambda\{\beta^T Z_i + H(t)\} \right] = 0, \tag{1.8}$$

$$\sum_{i=1}^{n} \frac{R_i}{\pi(Q_i, \hat{\psi}_n)} \left[dN_i(t) - Y_i(t) d\Lambda\{\beta^T Z_i + H(t)\} \right] = 0,\ 0 \le t \le \tau, \tag{1.9}$$

where $\hat{\psi}_n$ is the fixed maximum likelihood estimator derived by maximizing (1.2). In the case of no missing data, $R_i \equiv 1$, $\pi(Q_i) \equiv 1$ and consequently (1.8) and (1.9) become (1.5) and (1.6) that were proposed by Chen et al. (2002).

From now on we denote the inverse probability weighted complete-case estimator by $\hat{\beta}_n^{IPWCC}$. Since this estimator only uses complete cases, $\hat{\beta}_n^{IPWCC}$ is inefficient, and its asymptotic consistency relies on correctly modeling the probability of missingness $\pi(Q, \psi)$. Consequently, we now consider methods to improve the efficiency of the estimator for β.

1.3.2 Augmented Inverse Probability Weighted Complete-Case Estimator

To obtain efficiency gains over $\hat{\beta}_n^{IPWCC}$, as well as to get asymptotic consistency even if $\pi(Q, \psi)$ is misspecified; i.e the so-called double-robustness property, we now propose using the augmented inverse probability weighted complete-case estimator derived by Robins et al. (1994) for our problem. This estimator uses the complete cases similar to $\hat{\beta}_n^{IPWCC}$, but also augments the estimator by using the data from those with missing cause of death to gain efficiency and to obtain the double-robustness property that we will describe in greater detail in §5. We start by estimating the probability that cause of failure is the cause of interest; i.e.

$$\rho(W_i) = \text{pr}(\Delta_i = 2|\Delta_i > 0, W_i).$$

This probability can also be deduced through the relationship between the hazard function of T_1 and T_2; namely,

$$\frac{\lambda_2(u|z,a)}{\lambda_1(u|z,a)} = \frac{\rho(w)}{1-\rho(w)}, \tag{1.10}$$

where $\lambda_1(u|z,a)$ and $\lambda_1(u|z,a)$ are respectively, hazard functions of T_1 and T_2. Goetghebeur and Ryan (1995) used (1.10) to obtain a likelihood based estimator among others. Lu and Tsiatis (2001) used (1.10) for their multiple imputation estimator. We can posit a parametric model $\rho(W,\gamma)$ with unknown parameters γ. Because of its popularity and versatility, it is convenient to use a logistic regression model, where $\text{logit}\{\rho(W,\gamma)\} = W^T\gamma$, but again, other parametric models can also be easily accommodated. There is, however, the issue of how to obtain estimates for γ in the presence of missingness. It is easy to see that the missing at random assumption, stated in §2, also implies that, given Q_i, R_i is independent of Δ_i, for all i. i.e.,

$$\rho(W_i) = \text{pr}(\Delta_i = 2|\Delta_i > 0, W_i) = \text{pr}(\Delta_i = 2|R_i = 1, \Delta_i > 0, W_i). \tag{1.11}$$

This suggests that $\rho(W_i)$ can be deduced from the complete-cases for whom $(R_i = 1, \Delta_i > 0)$. Therefore, γ can be estimated using maximum likelihood; that is, by maximizing the following likelihood function with respect to γ

$$\prod_{i=1}^{n} \{\rho(W_i,\gamma)\}^{I(\Delta_i=2,R_i=1)}\{1-\rho(W_i,\gamma)\}^{I(\Delta_i=1,R_i=1)}, \tag{1.12}$$

we obtain the maximum likelihood estimator $\hat{\gamma}_n$ of γ. Similar to $\hat{\psi}_n$, under suitable regularity conditions, $\hat{\gamma}_n$ converges to the limit γ^*. If the model $\rho(W,\gamma)$ is correctly specified, we denote this by taking $\gamma^* = \gamma_0$, where $\rho(W,\gamma_0) = \text{pr}(\Delta = 2|\Delta > 0, W)$.

When there was no missing cause of death data, we already defined the counting process $N_i(t)$ and the corresponding martingale process $M_i(t)$ in §3.1. In addition, we also define the counting process $N_i^*(t) = I(T_i \le t)$, in which case, $N_i(t) = I(\Delta_i = 2)N_i^*(t)$. We note that, with missing cause of death data, when $R_i = 1$, we are able to observe the counting process $N_i(t)$, but are only able to observe the counting process $N_i^*(t)$ when $R_i = 0$. Using the inverse probability weighting theory for missing data developed by Robins, et al. (1994), for $0 \le t \le \tau$, we consider the following augmented inverse probability weighted estimating equations for $(\beta,\, H)$,

$$\sum_{i=1}^{n} \int_0^{\tau} Z_i \left[\frac{R_i dN_i(t)}{\pi(Q_i, \hat{\psi}_n)} - \frac{R_i - \pi(Q_i, \hat{\psi}_n)}{\pi(Q_i, \hat{\psi}_n)} \rho(W_i, \hat{\gamma}_n) dN_i^*(t) - Y_i(t) d\Lambda\{\beta^T Z_i + H(t)\} \right] = 0, \tag{1.13}$$

$$\sum_{i=1}^{n} \left[\frac{R_i dN_i(t)}{\pi(Q_i, \hat{\psi}_n)} - \frac{R_i - \pi(Q_i, \hat{\psi}_n)}{\pi(Q_i, \hat{\psi}_n)} \rho(W_i, \hat{\gamma}_n) dN_i^*(t) - Y_i(t) d\Lambda\{\beta^T Z_i + H(t)\} \right] = 0, \tag{1.14}$$

where $\hat{\psi}_n$ and $\hat{\gamma}_n$ are fixed maximum likelihood estimates. In the case of no missing data, $R_i \equiv 1$, $\pi(Q_i) \equiv 1$ and consequently (1.13) and (1.14) become (1.5) and (1.6).

The estimators for β and H which are the solution to the estimating equations (1.13) and (1.14) will be denoted by $\hat{\beta}_n^{DR}$ and $\hat{H}_n^{DR}$ respectively. Before deriving the statistical properties of the inverse probability weighted complete case estimator $\hat{\beta}_n^{IPW}$ and the augmented estimator $\hat{\beta}_n^{DR}$, we first describe how these estimators are actually computed.

1.4 A Computational Algorithm

A convenient computational algorithm for obtaining the augmented inverse probability weighted complete-case estimator, which can be easily modified for the inverse probability weighted complete-case estimator, is given as follows: Suppose there are K observed failure times, denoted by $0 < t_1 < t_2 < \cdots < t_K$. It is easy to show that, for any fixed value of β, the solution of (1.14) is unique. Therefore, we here propose the following iterative algorithm for computing $(\hat{\beta}_n^{DR}, \hat{H}_n^{DR})$. This algorithm (without *Step* 1) was also used by Chen et al. (2002).

Step 1. Get the maximum likelihood estimates $\hat{\gamma}_n$ and $\hat{\psi}_n$.

Step 2. Fix an initial value of β, denoted by $\beta^{(0)}$.

Step 3. Get $H^{(0)}$ as follows. First solve

$$\sum_{i=1}^{n} Y_i(t_1)\Lambda\{\beta^T Z_i + H(t_1, \beta, \hat{\psi}_n, \hat{\gamma}_n)\} = \sum_{i=1}^{n}\left\{\frac{R_i}{\pi(Q_i, \hat{\psi}_n)} dN_i(t_1) - \frac{R_i - \pi(Q_i, \hat{\psi}_n)}{\pi(Q_i, \hat{\psi}_n)}\rho(W_i, \hat{\gamma}_n)dN_i^*(t_1)\right\}$$

for $H^{(0)}(t_1)$ with $\beta = \beta^{(0)}$. Next, we solve the following equations one-by-one

$$\sum_{i=1}^{n} Y_i(t_k)\left[\Lambda\{\beta^T Z_i + H(t_k, \beta, \hat{\psi}_n, \hat{\gamma}_n)\} - \Lambda\{\beta^T Z_i + H(t_{k-1}, \beta, \hat{\psi}_n, \hat{\gamma}_n)\}\right]$$
$$= \sum_{i=1}^{n}\left\{\frac{R_i}{\pi(Q_i, \hat{\psi}_n)} dN_i(t_k) - \frac{R_i - \pi(Q_i, \hat{\psi}_n)}{\pi(Q_i, \hat{\psi}_n)}\rho(W_i, \hat{\gamma}_n)dN_i^*(t_k)\right\}$$

for $H^{(0)}(t_k)$ with $\beta = \beta^{(0)}$, $k = 2, 3, \cdots, K$.

Step 4. With the $H = H^{(0)}$ obtained in the previous step, solve (1.13) for a new estimate of β.

Step 5. Set $\beta^{(0)}$ to be the estimate obtained in step 4, repeat steps 3 and 4 until convergence.

The resulting estimators for β and H are denoted by $\hat{\beta}_n^{DR}$ and $\hat{H}_n^{DR}$.

1.5 Properties of the estimator

1.5.1 Double Robustness

We now show that the estimators $\hat{\beta}_n^{DR}$ and $\hat{H}_n^{DR}$ of (β, H) are doubly-robust; that is, the estimators $\hat{\beta}_n^{DR}$ and $\hat{H}_n^{DR}$ are consistent if either $\pi(Q, \psi)$ or $\rho(W, \gamma)$ are correctly specified. We first note that the estimating equation (1.14) being set equal to zero for $0 \le t \le \tau$, is equivalent to setting the following stochastic integrals equal to zero for $0 \le t \le \tau$, namely,

$$\sum_{i=1}^{n}\int_0^t \left[\frac{R_i dN_i(u)}{\pi(Q_i, \hat{\psi}_n)} - \frac{R_i - \pi(Q_i, \hat{\psi}_n)}{\pi(Q_i, \hat{\psi}_n)}\rho(W_i, \hat{\gamma}_n)dN_i^*(u) - Y_i(u)d\Lambda\{\beta^T Z_i + H(u)\}\right] = 0. \tag{1.15}$$

Consequently, to prove consistency, it suffices to show that the expected value of any summand of (1.13) and (1.15), evaluated at $(\beta, H) = (\beta_0, H_0)$ and substituting (ψ^*, γ^*) for $(\hat{\psi}_n, \hat{\gamma}_n)$, is equal to zero. In order to prove the property of double-robustness we need to show that the expected value of the summands are equal to zero if either $\psi^* = \psi_0$ or $\gamma^* = \gamma_0$. After adding and subtracting common terms and a little algebra, we can show that the summand of equation (1.13) and equation (1.15), evaluated at $(\beta_0, H_0, \psi^*, \gamma^*)$, are given by

$$\int_0^\tau Z_i dM_i(t, Z_i) + \int Z_i \frac{R_i - \pi(Q_i, \psi^*)}{\pi(Q_i, \psi^*)} \{I(\Delta_i = 2) - \rho(W_i, \gamma^*)\} dN_i^*(t). \tag{1.16}$$

and

$$\int_0^t dM_i(u, Z_i) + \frac{R_i - \pi(Q_i, \psi^*)}{\pi(Q_i, \psi^*)} \{I(\Delta_i = 2) - \rho(W_i, \gamma^*)\} dN_i^*(u), \quad 0 \le t \le \tau, \tag{1.17}$$

respectively.

The expected value of the first term in (1.16) is automatically zero by standard martingale theory for counting process (Fleming and Harrington, 1991). We now show that the expected value of the second term in (1.16) equals zero if either $\psi^* = \psi_0$ or $\gamma^* = \gamma_0$. If we have $\psi^* = \psi_0$, then by conditioning on (Δ_i, Q_i), the conditional expectation of the second term is $[\{E(R_i|\Delta_i, Q_i) - \pi(Q_i, \psi_0)\}/\pi(Q_i, \psi_0)]\{I(\Delta_i = 2) - \rho(W_i, \gamma^*)\} Z_i \int dN_i^*(t)$. Since $E(R_i|\Delta_i, Q_i) = \pi(Q_i, \psi_0)$ follows by the missing-at-random assumption, the conditional expectation is equal to zero, and consequently so is the unconditional expectation. If $\gamma^* = \gamma_0$, now by conditioning on (R_i, Q_i), the conditional expectation of the second term is equal to $[E\{I(\Delta_i = 2)|R_i, Q_i\} - \rho(W_i, \gamma_0)] Z_i [\{R_i - \pi(Q_i, \psi^*)\}/\pi(Q_i, \psi^*)] \int dN_i^*(t)$. Since the missing-at-random assumption implies $E\{I(\Delta_i = 2)|R_i, Q_i\} = \rho(W_i, \gamma_0) I(\Delta_i > 0)$, the conditional expectation equals zero, and consequently so is the unconditional expectation. In a similar fashion we can prove that the expectation of (1.17) equals zero.

It is important here to note that the proof of double robustness, given above, implicitly assumes that the estimator $\hat{\gamma}_n$ converges in probability to γ_0 whenever the model

$\rho(W, \gamma)$ is correctly specified regardless of whether the model $\pi(Q, \psi)$ for the probability of missingness is specified correctly or not. The maximum likelihood estimator $\hat{\gamma}_n$ that maximises (1.12) satisfies this property.

Although here we prove the double robustness relationship explicitly for our problem, it is important to point out that this result also follows from the theorem and proof on page 1144 of Scharfstein et al. (1999), which gives a general double robustness result for all augmented inverse probability weighted complete-case estimators.

1.5.2 Asymptotic Properties

We now establish the asymptotic properties of the augmented inverse probability weighted complete-case estimator $\hat{\beta}_n^{DR}$. At the end of this section we will briefly state the results for the inverse probability weighted complete-case estimator $\hat{\beta}_n^{IPWCC}$, which can be derived in a similar way. In all the asymptotic arguments we will always assume that at least one of the models $\pi(W_i, \psi)$ and $\rho(W_i, \gamma)$ are correctly specified, either (ψ^*, γ_0) or (ψ_0, γ^*) or (ψ_0, γ_0). We need to use similar notation and regularity conditions discussed in Chen et al. (2002). Define $\xi(t) = (\partial/\partial t) \log \lambda(t) = \dot{\lambda}(t)/\lambda(t)$, where $\dot{\lambda}(t)$ is the derivative of $\lambda(t)$. Also define $\pi_\psi(W_i, \psi) = (\partial/\partial\psi)\pi(W_i, \psi)$ and $\rho_\gamma(W_i, \gamma) = (\partial/\partial\gamma)\rho(W_i, \gamma)$ to be the partial derivatives. We assume that $\lambda(t)$ is positive, $\xi(t)$ is continuous and $\lim_{u\to-\infty} \lambda(u) = 0 = \lim_{u\to-\infty} \xi(u)$. We also assume that Z is bounded by some constant m almost surely, and that the derivatives of H_0 are continuous and positive. Then for any $t, s \in (0, \tau]$, we define

$$B(t, s) = \exp\left(\int_s^t \frac{E[\dot{\lambda}\{\beta_0^T Z + H_0(x)\}Y(x)]}{E[\lambda\{\beta_0^T Z + H_0(x)\}Y(x)]} dH_0(x)\right),$$

$$\mu_z(t) = \frac{E[Z\lambda\{\beta_0^T Z + H_0(T)\}Y(t)B(t, T)]}{E[\lambda\{\beta_0^T Z + H_0(t)\}Y(t)]},$$

$$A = \int_0^\tau E[(Z_1 - \mu_z(t))Z_1^T \dot{\lambda}\{\beta_0^T Z_1 + H_0(t)\}Y_1(t)]dH_0(t). \tag{1.18}$$

Proposition *Under suitable regularity conditions,*

$$n^{1/2}(\hat{\beta}_n^{DR} - \beta_0) = -A^{-1}\left(n^{-1/2}\left[\sum_{i=1}^{n}\int_0^{\tau}\{Z_i - \mu_z(t)\}dM_i^*(t,\cdot) - P_\psi I_\psi^{-1}S_{\psi i} - P_\gamma I_\gamma^{-1}S_{\gamma i}\right]\right) + o_p(1), \tag{1.19}$$

where $o_p(1)$ denotes a term that converges in probability to zero, I_ψ and $S_{\psi i}$ are, respectively, the information matrix and score vector for $\hat{\psi}_n$; I_γ and $S_{\gamma i}$ have similar meanings; namely,

$$I_\psi = E\left[\frac{I(\Delta_i > 0)\pi_\psi(W_i,\psi^*)\pi_\psi^T(W_i,\psi^*)}{\pi(W_i,\psi^*)\{1-\pi(W_i,\psi^*)\}}\right],$$

$$I_\gamma = E\left[\frac{I(\Delta_i > 0)\rho_\gamma(W_i,\gamma^*)\rho_\gamma^T(W_i,\gamma^*)}{\rho(W_i,\gamma^*)\{1-\rho(W_i,\gamma^*)\}}\right],$$

$$S_{\psi i} = \frac{I(\Delta_i > 0)\{R_i - \pi(W_i,\psi^*)\}\pi_\psi(W_i,\psi^*)}{\pi(W_i,\psi^*)\{1-\pi(W_i,\psi^*)\}},$$

$$S_{\gamma i} = \frac{R_i I(\Delta_i > 0)\{I(\Delta_i = 2) - \rho(W_i,\gamma^*)\}\rho_\gamma(W_i,\gamma^*)}{\rho(W_i,\gamma^*)\{1-\rho(W_i,\gamma^*)\}};$$

$$dM_i^*(t,\dot{}) = \frac{R_i dN_i(t)}{\pi(Q_i,\psi^*)} - \frac{R_i - \pi(Q_i,\psi^*)}{\pi(Q_i,\psi^*)}\rho(W_i,\gamma^*)dN_i^*(t) - Y_i(t)d\Lambda\{\beta_0^T Z_i + H_0(t)\}, \tag{1.20}$$

while P_ψ and P_γ are quantities defined in the Appendix.

In order to prove this proposition, we must first prove the consistency of $\hat{\beta}_n^{DR}$ and $\hat{H}_n^{DR}$ to β_0 and H_0, respectively. Having demonstrated that the expectation in (1.16) and (1.17) are equal to zero, the proof of consistency follows almost identical to that of Lu and Ying (2004). The remainder of the proof for the proposition is given in the Appendix.

Equation (1.19) shows that $n^{1/2}(\hat{\beta}_n^{DR} - \beta_0)$ is asymptotically a sum of independent and identically distributed random vectors that have zero mean, and thus, by the central limit theorem, converges to a normal random vector with mean zero and covariance matrix

$$A^{-1}\text{Var}\left[\int_0^{\tau}\{Z - \mu_z(t)\}dM^*(t,\cdot) - P_\psi I_\psi^{-1}S_\psi - P_\gamma I_\gamma^{-1}S_\gamma\right](A^{-1})^T. \tag{1.21}$$

This asymptotic covariance can be estimated by

$$\hat{A_n}^{-1}\left(n^{-1}\sum_{i=1}^{n} v_i v_i^T\right)(\hat{A_n}^{-1})^T, \tag{1.22}$$

where

$$v_i = \int_0^\tau \{Z_i - \hat{\mu}_z(t)\} d\hat{M}_i^*(t,\cdot) - \hat{P}_\psi \hat{I_\psi}^{-1}\hat{S_{\psi i}} - \hat{P}_\gamma \hat{I_\gamma}^{-1}\hat{S_{\gamma i}},$$

$$\hat{A}_n = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \{Z_i - \hat{\mu}_z(t)\} Z_i^T \dot{\lambda}\{\hat{\beta}_n^{DR^T} Z_i + \hat{H}_n^{DR}(t, \hat{\beta}_n^{DR}, \hat{\psi}_n, \hat{\gamma}_n)\} Y_i(t) d\hat{H}_n^{DR}(t, \hat{\beta}_n^{DR}, \hat{\psi}_n, \hat{\gamma}_n),$$

$$\begin{aligned} d\hat{M}_i^*(t,\cdot) &= \frac{R_i dN_i(t)}{\pi(Q_i,\hat{\psi}_n)} - \frac{R_i - \pi(Q_i,\hat{\psi}_n)}{\pi(Q_i,\hat{\psi}_n)}\rho(W_i,\hat{\gamma}_n) dN_i^*(t) \\ &\quad - Y_i(t) d\Lambda\{\hat{\beta}_n^{DR^T} Z_i + \hat{H}_n^{DR}(t, \hat{\beta}_n^{DR}, \hat{\psi}_n, \hat{\gamma}_n)\}, \end{aligned}$$

$$\hat{\mu}_z = \frac{\sum_{i=1}^n Z_i \lambda\{\hat{\beta}_n^{DR^T} Z_i + \hat{H}_n(T_i, \hat{\beta}_n^{DR}, \hat{\psi}_n, \hat{\gamma}_n)\} Y_i(t) \hat{B}_n(t, T_i)}{\sum_{i=1}^n \lambda\{\hat{\beta}_n^{DR^T} Z_i + \hat{H}_n(t, \hat{\beta}_n^{DR}, \hat{\psi}_n, \hat{\gamma}_n)\} Y_i(t)},$$

$$\hat{B}_n(t,s) = \exp\left(\int_s^t \frac{\sum_{i=1}^n \dot{\lambda}\{\hat{\beta}_n^{DR^T} Z_i + \hat{H}_n(x, \hat{\beta}_n^{DR}, \hat{\psi}_n, \hat{\gamma}_n)\} Y_i(x)}{\sum_{i=1}^n \lambda\{\hat{\beta}_n^{DR^T} Z_i + \hat{H}_n(x, \hat{\beta}_n^{DR}, \hat{\psi}_n, \hat{\gamma}_n)\} Y_i(x)} d\hat{H}_n(x, \hat{\beta}_n^{DR}, \hat{\psi}_n, \hat{\gamma}_n)\right),$$

for $t, s \in [0,\tau]$. Also $\hat{P}_\psi$, $\hat{P}_\gamma$ are the estimates by substituting $(\hat{\beta}_n^{DR}, \hat{H}_n^{DR}, \hat{\psi}_n, \hat{\gamma}_n)$ for $(\beta_0, H_0, \gamma^*, \psi^*)$, $\hat{I}_\psi$ and $\hat{I}_\gamma$ are usual observed information matrix, and $\hat{S}_{\psi i}$ and $\hat{S}_{\gamma i}$ are sample score vectors for the ith observation, with (γ^*, ψ^*) substituted by $(\hat{\psi}_n, \hat{\gamma}_n)$.

Using similar arguments, we now state the asymptotic properties for $\hat{\beta}_n^{IPWCC}$, when $\pi(W_i, \psi)$ is correctly specified. That is, $n^{1/2}(\hat{\beta}_n^{IPWCC} - \beta_0)$ is asymptotically normal with mean zero and covariance matrix

$$(A)^{-1}\mathrm{Var}\left[\int_0^\tau \frac{R}{\pi(Q)}\{Z - \mu_z(t)\} dM(t,\cdot) - P_\psi^{IPWCC} I_\psi^{-1} S_\psi(Q,\psi_0)\right]\{(A)^{-1}\}^T, \tag{1.23}$$

where P_ψ^{IPWCC} is defined in the Appendix, in a way similar to P_ψ.

We also give a theoretical justification of the superiority of the doubly-robust estimator $\hat{\beta}_n^{DR}$ to the inverse probability weighted complete-case estimator $\hat{\beta}_n^{IPWCC}$, in the situation where both $\pi(W_i, \psi)$ and $\rho(W_i, \gamma)$ are correctly specified.

As we showed in the Proposition, $n^{1/2}(\hat{\beta}_n^{DR} - \beta_0)$ is asymptotically equivalent to a sum of independent and identically distributed mean zero random vectors, where the ith element in the sum is referred to as the ith influence function of $\hat{\beta}_n^{DR}$. Specifically, using (1.19), we deduce that the ith influence function for $\hat{\beta}_n^{DR}$ is given by

$$\text{Inf}^{DR}(O_i) = A^{-1}\left[\int_0^\tau \{Z_i - \mu_z(t)\}dM_i^*(t) - P_\psi I_\psi^{-1} S_{\psi i} - P_\gamma I_\gamma^{-1} S_{\gamma i}\right], \tag{1.24}$$

and the asymptotic variance of $\hat{\beta}_n^{DR}$ is the variance of it's influence function; namely $\text{var}\{\text{Inf}^{DR}(O_i)\} = E\{\text{Inf}^{DR}(O_i)\text{Inf}^{DR^T}(O_i)\}$. Similarly, the ith influence function of $\hat{\beta}_n^{IPWCC}$ is

$$\text{Inf}^{IPWCC}(O_i) = A^{-1}\left[\int_0^\tau \frac{R_i}{\pi(Q_i)}\{Z_i - \mu_z(t)\}dM_i(t) - P_\psi^{IPWCC} I_\psi^{-1} S_{\psi i}\right], \tag{1.25}$$

and the asymptotic variance of $\hat{\beta}_n^{IPWCC}$ is $\text{var}\{\text{Inf}^{IPWCC}(O_i)\}$. We make a direct comparison of the variance of the influence function of these two estimators, which, in turn, gives us the desired result.

In the Appendix we prove that, in (1.24), $P_\psi = 0$ when $\rho(W_i, \gamma)$ is correct, and $P_\gamma = 0$ when $\pi(W_i, \psi)$ is correct. Therefore, when both models are correct, the ith influence function of $\hat{\beta}_n^{DR}$ given in (1.24) reduces to

$$\text{Inf}^{DR}(O_i) = A^{-1}\int_0^\tau \{Z_i - \mu_z(t)\}dM_i^*(t), \tag{1.26}$$

and it suffices to show that the variance of (1.26) is less than that of (1.25). It helps to note that when both models are correct, by the definition of $\rho(W)$, (1.20) can be written as

$$\begin{aligned} dM_i^*(t) &= \frac{R_i dM_i(t)}{\pi(Q_i)} - \frac{R_i - \pi(Q_i)}{\pi(Q_i)}\left[\rho(W_i)dN_i^*(t) - Y_i(t)d\Lambda\{\beta_0^T Z_i + H_0(t)\}\right] \\ &= \frac{R_i dM_i(t)}{\pi(Q_i)} - \frac{R_i - \pi(Q_i)}{\pi(Q_i)}\text{E}\{dM_i(t)|Q_i\}, \end{aligned}$$

hence (1.26) can be written as

$$\begin{aligned}\text{Inf}^{DR}(O_i) &= A^{-1}\left[\frac{R_i}{\pi(Q_i)}\int_0^\tau \{Z_i-\mu_z(t)\}dM_i(t)\right.\\ &\left.-\frac{R_i-\pi(Q_i)}{\pi(Q_i)}\int_0^\tau \{Z_i-\mu_z(t)\}\text{E}\{dM_i(t)|Q_i\}\right].\end{aligned}$$

Since by definition $P_\psi^{IPWCC}I_\psi^{-1}S_{\psi i}$ can be written as $\{R_i - \pi(Q_i)\}J(Q_i)$, for some function J, (1.25) can be written as

$$\begin{aligned}\text{Inf}^{IPWCC}(O_i) &\equiv A^{-1}\left[\frac{R_i}{\pi(Q_i)}\int_0^\tau \{Z_i-\mu_z(t)\}dM_i(t) - \{R_i-\pi(Q_i)\}J(Q_i)\right]\\ &= \text{Inf}^{DR}(O_i) + A^{-1}[\{R_i-\pi(Q_i)\}J^*(Q_i)],\end{aligned}$$

where $J^*(Q_i) = \int_0^\tau \{Z_i - \mu_z(t)\}/\pi(Q_i)\text{E}\{dM_i(t)|Q_i\} - J(Q_i)$. Notice that the Missing At Random assumption implies R_i is conditionally independent of $I(\Delta_i = 2)$ given Q_i, hence R_i is conditionally independent of $dM_i(t)$ given Q_i. Therefore the covariance of $A^{-1}\{R_i - \pi(Q_i)\}J^*(Q_i)$ and $\text{Inf}^{DR}(O_i)$ is given by

$$\begin{aligned}&\text{cov}\left[A^{-1}\{R_i-\pi(Q_i)\}J^*(Q_i), \text{Inf}^{DR}(O_i)\right]\\ =\ &\text{E}\left[A^{-1}\{R_i-\pi(Q_i)\}J^*(Q_i)\{\text{Inf}^{DR}(O_i)\}^T\right]\\ &-\text{E}\left[A^{-1}\{R_i-\pi(Q_i)\}J^*(Q_i)\right]\text{E}\left[\{\text{Inf}^{DR}(O_i)\}^T\right].\end{aligned}$$

The second term equals zero as $\text{E}\left[\{\text{Inf}^{DR}(O_i)\}^T\right] = 0$. The first term equals

$$\begin{aligned}&\text{E}\left(\text{E}\left[A^{-1}\{R_i-\pi(Q_i)\}J^*(Q_i)\{\text{Inf}^{DR}(O_i)\}^T \mid Q_i\right]\right)\\ =\ &\text{E}\left(A^{-1}\text{E}\left[\frac{R_i\{R_i-\pi(Q_i)\}}{\pi(Q_i)}|Q_i\right]J^*(Q_i)\text{E}\left[\int_0^\tau\{Z_i-\mu_z(t)\}dM_i(t)|Q_i\right]^T(A^{-1})^T\right)\\ &-\text{E}\left(A^{-1}\text{E}\left[\frac{\{R_i-\pi(Q_i)\}^2}{\pi(Q_i)}|Q_i\right]J^*(Q_i)\left[\int_0^\tau\{Z_i-\mu_z(t)\}\text{E}\{dM_i(t)|Q_i\}\right]^T(A^{-1})^T\right)\\ =\ &\mathbf{0}^{q\times q},\end{aligned}$$

where $\mathbf{0}^{q\times q}$ is a $q \times q$ matrix of zeros. Thus $A^{-1}\{R_i - \pi(Q_i)\}J^*(Q_i)$ is uncorrelated with $\text{Inf}^{DR}(O_i)$. Consequently

$$\text{var}\left\{\text{Inf}^{IPWCC}(O_i)\right\} = \text{var}\left\{\text{Inf}^{DR}(O_i)\right\} + \text{var}\left[A^{-1}\{R_i-\pi(Q_i)\}J^*(Q_i)\right]$$

$$> \quad \text{var}\left\{\text{Inf}^{DR}(O_i)\right\},$$

where for any two matrices A_1 and A_2, '$A_1 > A_2$' means $A_1 - A_2$ is a positive definite matrix.

1.6 Simulations

1.6.1 Simulation I: Proportional Hazards Assumption

To show the properties of the doubly-robust estimator, denoted by $\hat{\beta}_n^{DR}$, as well as to compare it with the inverse probability weighted complete-case estimator $\hat{\beta}_n^{IPWCC}$ and the naive complete-case estimator $\hat{\beta}_n^{BCC}$ that only uses the observations with complete information, under the proportional hazards assumption, we first conducted a simulation study that was designed as follows. We chose $H_0(t) = \log(t)$ and ε with hazard function of the form $\lambda(t) = \exp(t)$. The univariate covariate Z_i was chosen to take values 1 or 0 with equal probability 0.5. Given Z_i, the cause of interest T_{2i} followed an exponential distribution with hazard function $\lambda_2(t|z) = \lambda e^{\beta z}$. The other cause T_{1i} followed a Gompertz distribution with hazard $\lambda_1(t|z) = \lambda e^{\xi+\eta t}$. Finally, the censoring variable C_i was generated independent of Z_i following a uniform distribution $U(\tau_f, \tau_a + \tau_f)$. This setup results in a logistic regression model $\text{logit}\{\rho(W_i)\} = -\xi - \eta T_i + \beta Z_i$ for $\rho(W_i)$, as well as another logistic regression model $\text{logit}\{\pi(W_i)\} = \psi_0 + \psi_1 T_i + \psi_2 Z_i$ for $\pi(W_i)$. Since we wanted to study the behavior of our estimators in the situations where either one or both models were misspecified, when we misspecified $\pi(W)$ or $\rho(W)$ or both, we did so by assuming these probabilities are constants independent of the W.

The parameters were chosen to be as follows, $\beta = -1$, $\lambda = 1$, $\xi = -1$, $\eta = 0.5$, $\tau_a = 0.5$, $\tau_f = 1.5$, $\psi_0 = 1$, $\psi_1 = -1$ and $\psi_2 = -1$. Under this setting, we had, on average, 48% failures from the cause of interest, 39% failures from other cause, and the remaining 13% from censored observations. Moreover, 52% failures had missing cause of failure information. The size of the Monte-Carlo simulation was $MB = 1,000$ and the sample size was $n = 200$. We summarise the results in Table 1 where we present the bias, the Monte Carlo standard deviation, the Monte Carlo average of the estimated standard deviations and the empirical coverage probability of the 95% confidence intervals of the various estimators under the various scenarios.

The simulation results were consistent with the theoretical results. That is, the complete-case estimator $\hat{\beta}_n^{BCC}$ was biased in all four scenarios. The inverse probability weighted estimator $\hat{\beta}_n^{IPWCC}$ was unbiased only when the model for $\pi(W_i)$ was correctly specified. However, $\hat{\beta}_n^{DR}$ was unbiased when either one or both of the two models were correctly specified. When both models were misspecified, $\hat{\beta}_n^{DR}$ had the smallest bias among all estimators. In addition, the simulation results showed that $\hat{\beta}_n^{DR}$ was clearly more efficient than other estimators.

1.6.2 Simulation II: Proportional Odds Assumption

We also illustrate how our estimators perform with the proportional odds model. In this simulation, we chose $H_0(t) = \log(t)$ and ε with hazard function of the form $\lambda(t) = \exp(t)/\{1 + \exp(t)\}$. The covariate Z_i was again chosen to be equal to 1 or 0 with equal probability 0.5. The cause of interest T_{2i} had an hazard function $\lambda_2(t|z) = \lambda \exp(\beta z)/\{1+ t\exp(\beta z)\}$, and the other cause T_{1i} with hazard function $\lambda_1(t|z) = \lambda e^{\xi+\eta t}$. The censoring variable C_i was generated from a uniform distribution $U(\tau_f, \tau_a + \tau_f)$, independent of the covariate Z_i. Under this scenario the true models are $\rho(W_i) = \{1 + t\exp(\beta z) + \exp(-\xi - \eta t + \beta z)\}^{-1} \exp(-\xi - \eta t + \beta z)$ and logit $\pi(W_i) = \psi_0 + \psi_1 T_i + \psi_2 Z_i$. Again we purposely misspecified either one model or both models by taking them to be constants. We considered three estimators: $\hat{\beta}_n^{BCC}$, $\hat{\beta}_n^{IPWCC}$ and $\hat{\beta}_n^{DR}$. The parameter values were chosen to be $\beta = -1$, $\lambda = 1$, $\xi = -1.5$, $\eta = 0.5$, $\tau_a = 1$, $\tau_f = 2$, $\psi_0 = 1.5$, $\psi_1 = -1$ and $\psi_2 = -1$. This gave, on average, 46% failures from the cause of interest, 41%

failures from the other cause, and 13% from censored observations. And 49% failures had missing cause information. In each experiment, $1,000$ replicates were taken with sample size $n = 200$. The results were summarised in Table 2.

The simulation results in Table 2 lead us to the same conclusions as those in simulation 1. That is, the complete-case estimator $\hat{\beta}_n^{BCC}$ always had severe bias. The inverse probability weighted estimator $\hat{\beta}_n^{IPWCC}$ was biased when the $\pi(W_i)$ model was misspecified. $\hat{\beta}_n^{DR}$ was biased only if both models were misspecified. It can also be seen that $\hat{\beta}_n^{DR}$ was more efficient than other estimators.

1.7 Breast Cancer Example

We here illustrate the method using the data from a clinical trial in stage II breast cancer. There were a total of 169 patients enrolled in the study, among which 90 were censored. Of the 79 patients who were dead: 18 deaths had unknown cause, 44 deaths in breast cancer and the rest 17 in other causes. We were interested in measuring the covariate effects on the death rate of breast cancer. In Cummings et al. (1986), the two covariates, being ER-negative and presence of over four positive nodes, were observed to be significantly related to overall survival. For each failure type, Goetghebeur and Ryan (1995) conducted a cause-specific survival analysis using a proportional hazards model. Also assuming the hazard functions follow the proportional hazards model, Lu and Tsiatis (2001) posited a logistic regression model for $\rho(W)$ and then used a multiple imputation procedure. Their results together with the doubly robust estimator were presented in Table 1.3. To derive our doubly robust estimators, we start by positing logistic regression models for $\rho(W)$ and $\pi(W)$, as functions of the covariates W. The covariates that we considered including ER status, tumor size, number of positive notes and time of relapse. As mentioned in the last paper, among the 6 patients who were

ER-negative, 5 of them died and they all died of breast cancer. This suggests that including ER status as a covariate for fitting logistic regression models for either $\rho(W)$ or $\pi(W)$ is not appropriate as the maximum likelihood estimator would not exist. Thus the logistic regression model for $\rho(W)$ was derived using the subset of patients who died with know cause and were ER-positive, and $\pi(W)$ was fitted with a logistic regression model using the subset of patients who had positive ER status. Doubly robust estimates and their variance estimates were then computed, for the proportional hazards model and the proportional odds model. The results in Table 1.3 clearly show that the complete-case estimates and the corresponding estimated standard errors were significantly different from the others, while the doubly robust estimator with the proportional hazards model were quite close to those given by Goetghebeur and Ryan (GR, 1995) and Lu and Tsiatis (LT, 2001).

1.8 Discussion

We have proposed an inverse probability weighted complete-case estimator $\hat{\beta}_n^{IPW}$, as well as an augmented inverse probability weighted complete-case estimator $\hat{\beta}_n^{DR}$ for the regression coefficients of linear transformation models with missing cause of failure for some individuals. It was shown that $\hat{\beta}_n^{IPW}$ was consistent and asymptotically normal only when the model for the probability of missingness was truly specified, while $\hat{\beta}_n^{DR}$ was consistent and asymptotically normal when either the model for the probability of missingness or the model for the probability of cause of failure were correctly specified; i,e. the so-called double-robustness property of this estimator. A consistent estimator for the asymptotic variance of each estimator was also derived. Simulation results demonstrated the adequacy of the asymptotic theory of these two estimators, and the superiority (less bias and higher efficiency) of $\hat{\beta}_n^{DR}$ to other estimators for this model. Since the class of linear transformation models consist of a broad range of models, including the commonly used proportional hazards model and proportional odds model, further research on model diagnostic techniques for detecting the most appropriate transformation model in the presence of missing cause of failure is necessary and part of ongoing research.

Table 1.1: Simulation Results for Proportional Hazards Model

	Case 1.			Case 2.		
	BCC	IPW	DRT	BCC	IPW	DRT
Bias	-0.347	-0.029	-0.031	-0.347	-0.029	-0.027
SSE	0.376	0.351	0.319	0.376	0.351	0.331
SEE	0.354	0.375	0.308	0.354	0.375	0.320
CP	0.864	0.957	0.948	0.864	0.957	0.947
	Case 3.			Case 4.		
	BCC	IPW	DRT	BCC	IPW	DRT
Bias	-0.347	-0.118	-0.028	-0.347	-0.118	0.086
SSE	0.376	0.338	0.306	0.376	0.338	0.244
SEE	0.354	0.326	0.297	0.354	0.326	0.232
CP	0.864	0.878	0.947	0.864	0.878	0.922

NOTE: Case 1 is where both $\pi(\cdot)$ and $\rho(\cdot)$ were correctly specified, case 2 is where $\pi(\cdot)$ was correctly specified but $\rho(\cdot)$ was misspecified, case 3 is where $\pi(\cdot)$ was misspecified but $\rho(\cdot)$ was correctly specified, and finally case 4 is where both $\pi(\cdot)$ and $\rho(\cdot)$ were misspecified. Bias denotes the empirical bias of $\hat{\beta}_n$. SSE denotes the sample deviation of $\hat{\beta}_n$, SEE denotes the Monte-Carlo average of the standard error estimates of $\hat{\beta}_n$. CP denotes the empirical coverage probability of the 95% confidence intervals.

Table 1.2: Simulation Results for Proportional Odds Model

	Case 1.			Case 2.		
	BCC	IPW	DRT	BCC	IPW	DRT
Bias	-0.214	0.041	-0.024	-0.214	0.041	-0.018
SSE	0.403	0.429	0.389	0.403	0.429	0.417
SEE	0.411	0.449	0.378	0.411	0.449	0.392
CP	0.935	0.955	0.947	0.935	0.955	0.953
	Case 3.			Case 4.		
	BCC	IPW	DRT	BCC	IPW	DRT
Bias	-0.214	-0.099	-0.022	-0.214	-0.099	0.094
SSE	0.403	0.395	0.361	0.403	0.395	0.321
SEE	0.411	0.363	0.348	0.411	0.363	0.333
CP	0.935	0.929	0.945	0.935	0.929	0.912

NOTE: Case 1 is where both $\pi(\cdot)$ and $\rho(\cdot)$ were correctly specified, case 2 is where $\pi(\cdot)$ was correctly specified but $\rho(\cdot)$ was misspecified, case 3 is where $\pi(\cdot)$ was misspecified but $\rho(\cdot)$ was correctly specified, and finally case 4 is where both $\pi(\cdot)$ and $\rho(\cdot)$ were misspecified. Bias denotes the empirical bias of $\hat{\beta}_n$. SSE denotes the sample deviation of $\hat{\beta}_n$, SEE denotes the Monte-Carlo average of the standard error estimates of $\hat{\beta}_n$. CP denotes the empirical coverage probability of the 95% confidence intervals.

Table 1.3: Comparison of complete-cases, Goetghebeur and Ryan, multiple imputation by Lu and Tsiatis, and doubly robust estimator for both the proportional hazards model and the proportional odds model, using the breast cancer data

	Effective Covariates	
Methods	ER Status	Great Than 4 Nodes
BCC (PH Model)	$0.71^a(0.3065^b)$	1.70(0.4861)
GR^c (PH Model)	0.57(0.2803)	1.59(0.4822)
LT^d (PH Model)	0.60(0.2618)	1.61(0.4794)
DR^e (PH Model)	0.53(0.2865)	1.60(0.5418)
DR (PO Model)	0.60(0.3706)	2.10(0.6747)

[a] Estimate
[b] Standard Error
[c] Goetghebeur and Ryan
[d] Lu and Tsiatis
[c] Doubly robust
PH denotes Proportional Hazards Model
PO denotes Proportional Odds Model

Chapter 2

Multiple imputation method

2.1 Introduction

In many clinical studies, interest is often on comparing time-to-failure from a specific cause between two treatments. In these studies, a group of patients are observed from their entry into the study until the occurrence of some particular event such as death. Often the observation of time-to-failure is right-censored for a subset of patients for various reasons such as finite study duration, loss to follow-up, or withdrawal from the study. It is also very common that patients may fail from one of multiple causes, but only one of which is of primary interest. In these situations, the theory of competing risks can be easily applied to assess the effects of certain covariates, such as treatments, on cause-specific hazards. From this competing risks point of view, we can then think of the problem as the case where there is a potential failure time associated with each of a set of competing risks, for which we can only observe the minimum of the potential failure times along with the cause of this minimum. The theory also tells us, that under the assumption that all the potential failure times as well as the censoring are conditionally independent given the covariates, the problem is identifiable (Tsiatis, 1975) in the sense that each cause-specific hazard function corresponds to the hazard function for the potential failure time and vice versa. Hence our interest in estimating the cause-specific hazard function for the cause of interest as a function of covariates, is equivalent to the interest in estimating the hazard function for potential failure time of that cause given the covariates. A popular strategy is to use the proportional hazards model to model the cause-specific hazard of potential failure time of the cause of interest, and by treating other failure types as censoring (Cox and Oakes, 1984).

Sometimes the patients are observed to die but the cause of failure may not be available. For instance, whether or not the failure is attributable to the cause of interest may require documentation with information that is unavailable either because the information is lost or not collected, or it may be difficult to determine the cause for some patients (Anderson, Goetghebeur, and Ryan, 1996). In such cases, it has been shown that ignoring the observations with missing cause of failure from the analysis may yield biased inferences on covariate effects, unless the missingness mechanism is missing completely at random (Rubin, 1976), which is very unlikely to hold in actuality. Under the assumption that the probability of a missing cause of failure occurring may depend on time-to-failure but not on the covariates, and that baseline cause-specific hazards of the competing risks are proportional, Goetghebeur and Ryan (1995) proposed a method that uses a full partial likelihood (Holt, 1978; Kalbfleisch and Prentice, 1980) and a modified partial likelihood. Lu and Tsiatis (2005) extended their results to the more general setting where the probability of a missing cause of failure occurring may depend on time-to-failure as well as the covariates, and where the ratio of two baseline cause-specific hazards may be dependent of time-to-failure. Under these assumptions and by using the full likelihood approach above, Lu and Tsiatis (2005) deduced an estimator that is consistent, asymptotically normal, semiparametric efficient and is applicable under more general missingness assumptions. It is worthwhile noting that all these methods assume that the cause-specific hazards of the cause of interest follow the proportional hazards model.

In this article we take a different approach. By making direct assumptions on the probability that the missing cause is the cause of interest, we use multiple imputation procedures (Rubin, 1987, 1996) to impute the missing causes that in turn give us completed data sets. Since in a competing risks model each cause-specific hazard function uniquely corresponds to the hazard function of that potential failure time and vice versa, on each of several completed data sets, one can then estimate the cause-specific hazard function for the cause of interest as a function of covariates, using usual censored data methods by combining censoring time and all other cause types into one single censoring variable. The resulting estimators for the regression coefficients are then averaged across the completed data sets to yield a consistent and asymptotically normal estimator. A popular approach is to assume that the cause-specific hazard of interest follows a proportional hazards model, and then estimate the regression coefficients using maximum partial likelihood method (Cox, 1972, 1975). Lu and Tsiatis (2001) posited parametric models for the probability that the missing cause is that of interest, and then used a multiple imputation procedure to estimate the regression coefficients, assuming a proportional hazards model. However, it has been shown by many authors that sometimes proportional hazards model is not appropriate and alternative models may be more reasonable, e.g. the proportional odds model (Pettitt 1982, 1984; Bennett, 1983; Murphy et al., 1997) is more suitable in the situations where the hazard functions for the two treatment groups have the same limit.

Here we consider the class of linear transformation models that is defined in §2.2. A major advantage of working with the class of linear transformation models is that it is very flexible and it includes the proportional hazards model and the proportional odds model as special cases, which will be elaborated in great detail later. Under the rather restrictive assumption that the censoring is independent of covariates, Cheng et al. (1995, 1997), Fine et al. (1998) and Cai et al. (2000) proposed and further developed a unified estimation approach for arbitrary linear transformation models in the presence of censored data, with no missing cause of failures. However, in a competing risks model, their assumption implies that all other causes of failure would have to be independent of covariates, which is not practical in most applications. Chen et al. (2002) gave a general estimation procedure without making the independence assumption above. Their approach generalized the idea of partial maximum likelihood approach for the proportional hazards model, to the class of linear transformation models. By taking advantage of the martingale structure of the model, the solution to their estimating equations is proved to be consistent and asymptotically normal with variance that can be consistently estimated using sandwich estimators. However, in the presence of missing cause of failure, this method cannot be directly applied. Some careful modification is necessary. In this paper we propose a general single and multiple imputation procedure for the class of linear transformation models in the presence of both censored data and missing cause of failure.

We state the notation and necessary assumptions in §2.2. In §2.3 we describe a censored data method proposed by Chen et al. (2002). The imputation procedure is described in §2.4. A computational algorithm for obtaining the estimators is given in §2.5. Asymptotic properties of the estimators are stated in §2.6. Simulation studies demonstrate the relevance of the theory in finite samples in §2.7. In §2.8 the procedure is applied to the data from a clinical trial breast cancer stage II study. Some conclusions and remarks are given in §2.9. And finally we prove the asymptotic normality of the estimators in the Appendix.

2.2 Notation and Assumptions

Without loss of generality, we only consider the situations where the patients are associated with two specific causes that may lead to failure. We refer to these as cause 2 and cause 1. This can be easily generalized to the cases where there are more than two causes of failure. In those cases one can refer to the cause of interest as cause 2, and the combination of all other causes as cause 1. First consider the situation with the absence of censoring, where the data we get to observe can be summarized as (U^*, Δ^*), where U^* denotes the time-to -failure and Δ^* denotes the cause of failure, taking on value 1 if the cause of failure is cause 1 and value 2 otherwise. Suppose we are interested in estimating the effects of some covariates defined as Z, which characterize the cause-specific hazard function for the cause of interest (i.e. cause 2). Z is q dimensional. The cause-specific hazard function is defined as

$$\lambda_2^*(u|z) = \lim_{h \to 0} h^{-1} \text{pr}(u \leq U^* < u + h, \Delta^* = 2 | U^* \geq u, Z = z). \tag{2.1}$$

From the competing risks point of view, one can think of the problem as the case where there is potential failure time T_1 corresponding to cause 1 and potential failure time T_2 corresponding to cause 2, and U^* is the minimum of T_1 and T_2, while Δ^* is the cause of the minimum, i.e. $\Delta^* = 1$ if $U^* = T_1$ and $\Delta^* = 2$ otherwise. In the competing risks theory, under the assumption that the causes of failure are conditionally independent given the covariates Z, the problem is identifiable (Tsiatis, 1975) and (2.1) is the same as the hazard function of T_2 given Z. This relationship enables us to use the existing models for the hazard function of T_2 as a function of the covariate Z, to model the cause-specific hazard function of T_2 given Z in the competing risks settings. Often this relationship is modeled via the proportional hazards model, defined as

$$\lambda_2^*(u|z) = \lambda_2(u) \exp(\beta_0^T z), \tag{2.2}$$

where $\lambda_2(u)$ is an arbitrary baseline hazard function for cause 2, and β_0 denotes the true value of the $q \times 1$ regression coefficients β that measure the effect of covariates Z. However, as we have mentioned in §2.1, due to the limitation of the proportional hazards model, recently much attention has been drawn to the class of linear transformation models that is defined as

$$H(T_2) = -\beta_0^T Z + \varepsilon, \tag{2.3}$$

where H is an unknown monotone increasing function, β_0 has the same meaning as that in (2.2), and ε is a random error term that is independent of Z with a known continuous distribution. (2.3) reduces to the proportional hazards model if the error term ε follows the extreme value distribution; and (2.3) turns out to be the proportional odds model when the distribution of ε is the standard logistic distribution. More detailed description about the class of linear transformation models can be found in Cheng et al. (1995, 1997). Just like the proportional hazards model, the class of linear transformation models describes the hazard function of T_2 as a function of Z; yet (2.3) offers more flexibility than (2.2). Due to the correspondence between the hazard functions and the cause-specific hazard functions in a competing risks model, (2.3) will be used to model the cause-specific hazard functions $\lambda_2^*(u|z)$.

Incomplete follow-up is very common in clinical trials, and because of which, the time-to-failure data are often censored by a variable that is defined as C. In the presence of censoring, the data that are observed can be summarized as (T, Δ) where T is the minimum of U^* and C, and Δ is the failure-censoring indicator, i.e. $\Delta = 0$ if $T = C$, $\Delta = 1$ if $T = T_1$, and $\Delta = 2$ if $T = T_2$. By the theory of competing risks, the problem is identifiable under the assumption that the censoring C is conditionally independent of (U^*, Δ^*) given Z. Under this assumption, the cause-specific hazards functions of interest are equivalent to the 'observable' cause-specific hazard functions for cause 1 and 2, with censoring in present, defined as

$$\lambda_j(u|z) = \lim_{h\to 0} h^{-1}\text{pr}(u \le T < u+h, \Delta = j|T \ge u, Z = z), \quad j = 1, 2.$$

Particularly in our problem $\lambda_2(u|z) = \lambda_2^*(u|z)$. This relationship allows one to use the censored data methods such as the one proposed by Chen et al. (2002), to estimate the regression coefficients β and the monotone increasing function H in a linear transformation model, after treating the observed T as censored times where Δ is equal to 1 or 0.

Sometimes a patient dies and cause of failure is unknown, then the classification is uncertain. In the literature this problem has not yet been studied for the class of linear transformation models. We here define R to be the missing data indicator, i.e. $R = 1$ if Δ is observed and $R = 0$ otherwise. Since there is no missing cause of failure associated with a patient when it is censored, R is always set to be 1 whenever $\Delta = 0$. The imputation procedure relies on the assumption on the missing data mechanism. As pointed out by Lu and Tsiatis (2001), the assumption on the missing data mechanism will be more tenable if auxiliary covariates denoted by A are considered. A may not be of interest in modeling the hazards function, but it may be related to the missing data mechanism. For instance, A can be some post treatment variables that lead to the missing causes of failure but A was not used in modeling the hazards function because it may affect the causal interpretation of the parameters for the effects of treatment. To summarize, the data that we get to observe are $O_i = \{R_i, T_i, Z_i, A_i, I(\Delta_i = 0), R_i I(\Delta_i = 1), R_i I(\Delta_i = 2)\}$, which are independent across i.

In this paper, we assume that the causes of failures are missing at random (MAR, Rubin, 1976) in the sense that the missing causes do not depend on the causes themselves. Equivalently the probability that the cause of failure of the ith patient is missing, conditioned on $(\Delta_i, \Delta_i > 0)$ and $W_i = (T_i, Z_i, A_i)$, depends only on the always observed

W_i but not on Δ_i that is not always observed,

$$\text{pr}(R_i = 1|\Delta_i, \Delta_i > 0, W_i) = \text{pr}(R_i = 1|\Delta_i > 0, W_i), \qquad 1 \leq i \leq n. \tag{2.4}$$

This relationship also implies that R_i is conditionally independent of Δ_i given $\{W_i, \Delta_i > 0\}$, i.e.,

$$\text{pr}(\Delta_i = 2|R_i = 0, \Delta_i > 0, W_i) = \text{pr}(\Delta_i = 2|R_i = 1, \Delta_i > 0, W_i), \tag{2.5}$$

hence this probability can be consistently estimated using the observed data. In the following we first state the results in Chen et al. (2002), without missing causes of failure. The we propose a multiple imputation procedure to accommodate the missingness.

2.3 Censored Data Method

In the presence of censoring when there is no missing cause of death information, T_1 and C can be combined into a single censoring variable and, as usual, the counting process for T_2 is defined as $N_i(t) = I(\Delta_i = 2)I(T_i \leq t)$ and the at-risk indicator process as $Y_i(t) = I(T_i \geq t)$. Then by the counting process theory (Fleming and Harrington, 1991), the corresponding martingale process is defined as $M_i(t) = N_i(t) - \int_0^t Y(u) d\Lambda\{H_0(u) + \beta_0^T Z_i\}$, where Λ is the cumulative hazard function of ε in (2.3), and H_0 and β_0 denote the true values of H and β, respectively. This martingale structure allowed Chen et al. (2002) to estimate (β, H) jointly with the following estimating equations,

$$\sum_{i=1}^{n} \int_0^\infty Z_i[dN_i(t) - Y_i(t) d\Lambda\{H(t) + \beta^T Z_i\}] \equiv \sum_{i=1}^{n} \int_0^\infty Z_i dM_i(t, Z_i, \beta, H) = 0, \tag{2.6}$$

$$\sum_{i=1}^{n} [dN_i(t) - Y_i(t) d\Lambda\{H(t) + \beta^T Z_i\}] \equiv \sum_{i=1}^{n} dM_i(t, Z_i, \beta, H) = 0, \quad t \geq 0. \tag{2.7}$$

The resulting estimator of β was proved to be consistent and asymptotically normal, and the explicit expression of the estimator for the limiting variance-covariance matrix follows from the martingale representation of (2.6) and (2.7). As a special case, it is easy to see that (2.6) and (2.7) reduce to the partial likelihood score when the proportional hazards model is assumed.

However, with missing cause of failure, (2.6) and (2.7) can not be directly applied because complete data are not available.

2.4 Imputation Procedure

Here we propose an imputation procedure to deal with the issue of missing cause of failure. This procedure begins by forming a completed data set. Toward that end, we impute the missing cause of failure $D_i = I(\Delta_i == 2)$ values from the conditional distribution of D_i given the observed data. Since D_i is either zero or one, its conditional distribution is Bernoulli with success probability $\text{pr}(\Delta_i = 2|R_i = 0, \Delta_i > 0, W_i)$, which, from (2.5), equals $\text{pr}(\Delta_i = 2|R_i = 1, \Delta_i > 0, W_i) = \rho(W_i)$, say. This suggests that the imputation probability $\rho(W_i)$ may be deduced from the complete cases with $(R_i = 1, \Delta_i > 0)$. We posit a parametric model for $\rho(W_i)$ in terms of some unknown finite dimensional parameters γ, with $\rho(W_i) = \rho(W_i, \gamma_0)$ with γ_0 being the true value of γ. Due to the Bernoulli nature of D_i, it is convenient to use a logistic regression model, i.e., $\text{logit}\{\rho(W, \gamma)\} = W^T\gamma$, but other parametric models can also be easily accommodated. For instance, to make the model more flexible, we may include higher order polynomials and interaction terms for the logistic regression model $\rho(W, \gamma)$. It is also straight forward to show that, by the definition of $\rho(W)$, we have the following relationship

$$\frac{\lambda_2(t|z,a)}{\lambda_1(t|z,a)} = \left\{\frac{\rho(W)}{1-\rho(W)}\right\}, \tag{2.8}$$

where $w = (t, z, a)$ is any realization of W. This relationship suggests that $\rho(W)$ can be deduced from the ratio of the cause-specific hazard functions. Notice that the hazards in (2.8) depend on both Z and the auxiliary covariate A, and hence they may not be the same as the ones in (2.3) that do not depend on A. In any case, (2.8) can not be used to model $\rho(W)$ unless both hazards are known to us, which is not likely to be the case in applications. Nonetheless, we posit parametric models that closely describe this relationship. Under a chosen parametric model, γ can be estimated using maximum likelihood methods on the completed cases with $(R_i = 1, \Delta_i > 0)$. For instance, for the logistic regression model, by maximizing the following likelihood function with respect to γ

$$\prod_{i=1}^{n} \{\rho(W_i, \gamma)\}^{I(\Delta_i=2, R_i=1)} \{1 - \rho(W_i, \gamma)\}^{I(\Delta_i=1, R_i=1)}, \tag{2.9}$$

we obtain the maximum likelihood estimator $\hat{\gamma}_n$ for γ, which provides $\rho(W_i, \hat{\gamma}_n)$ as an estimate for the success probability $\mathrm{pr}(\Delta_i = 2 | R_i = 0, \Delta_i > 0, W_i)$. If the model for $\rho(W_i, \gamma)$ is correctly specified, then $\hat{\gamma}_n$ is consistent and so is $\rho(W_i, \hat{\gamma}_n)$.

For any fixed γ, let $D_{ij}(R_i, \gamma)$ denote the imputation of D_i from the jth imputed data set, $j = 1, \cdots, m$. If cause of failure is known to us, i.e. R_i=1, then there is no need to impute and hence we take $D_{ij}(R_i, \gamma)$ to be D_i. Otherwise we randomly choose $D_{ij}(R_i, \gamma)$ to be either one or zero with probability $\rho(W_i, \gamma)$ and $\{1 - \rho(W_i, \gamma)\}$, respectively.

It is helpful to see that the joint distribution of (D_i, W_i) and $\{D_{ij}(R_i, \gamma_0), W_i\}$ are one and the same. When cause of failure is not missing $(R_i = 1)$, as stated above, $D_{ij}(R_i, \gamma_0) = D_i$; and when $R_i = 0$, $\{I(\Delta_i > 0), W_i\}$ follows the distribution of the observed data and $D_{ij}(R_i, \gamma_0)$ follows the distribution of the unobserved D_i given the observed data, hence $\{D_{ij}(R_i, \gamma_0), W_i\}$ is just a draw from the joint distribution of the full data (D_i, W_i). Thus, for the correctly specified model $\rho(W, \gamma)$ and $\gamma = \gamma_0$, a single imputation of the missing causes is probabilistically the same as the experiment without any missingness.

For a correctly specified model $\rho(W, \gamma)$, as mentioned above, $\hat{\gamma}_n$ is consistent. This suggests that the data $D_{ij}(R_i, \hat{\gamma}_n)$ that we generate are asymptotically as good as that from the original experiment. Therefore, we can fit the linear transformation model to each completed data set and obtain consistent estimators using Chen et al.(2002), with the only modification being replacing $N_i(t)$ in (2.6) and (2.7) with $N_{ij}(D_{ij}, t) = I(T_i \leq t, D_{ij} = 1)$, $j = 1, \cdots, m$. I.e., we consider the following estimating equations,

$$\sum_{i=1}^{n} \int_0^{\infty} Z_i[dN_{ij}(t) - Y_i(t)d\Lambda\{H(t) + \beta^T Z_i\}] \equiv \sum_{i=1}^{n} \int_0^{\infty} Z_i dM_{ij}(t) = 0, \tag{2.10}$$

$$\sum_{i=1}^{n} [dN_{ij}(t) - Y_i(t)d\Lambda\{H(t) + \beta^T Z_i\}] \equiv \sum_{i=1}^{n} Z_i dM_{ij}(t) = 0, \tag{2.11}$$

where $dM_{ij}(t) = dN_{ij}(t) - Y_i(t)d\Lambda\{H(t) + \beta^T Z_i\}$ is the martingale process for the completed data. The resulting estimator is called single imputation estimator, denoted by $(\hat{\beta}_j, \hat{H}_j)$. For a more efficient estimator, we may carry out the imputation procedure multiple times and average the estimators. This results in the multiple imputation estimator $(\hat{\beta}, \hat{H})$. $\hat{\beta} = \sum_{j=1}^{m} \hat{\beta}_j$.

2.5 Algorithm

We here propose an iterative algorithm for computing the single and multiple imputation estimators for (β, H). This algorithm modifies the one used by Chen et al. (2002) to accommodate the missingness. Suppose there exists K observed failure times, denoted by $0 < t_1 < t_2 < \cdots < t_K$. Since for any fixed value of β the solution of (2.11) is unique, the algorithm for the single imputation estimators $(\hat{\beta}_j, \hat{H}_j)$, for a typical completed data set, is as follows,

Step 1. Get the maximum likelihood estimate $\hat{\gamma}_n$.

Step 2. Obtain a completed data set by imputing D_{ij} with the probability $\rho(W_i, \hat{\gamma}_n)$ for the observations with missing cause of failure.

Step 3. Fix an initial value of β. denoted by $\beta^{(0)}$.

Step 4. Get $H^{(0)}$ as follows. First solve

$$\sum_{i=1}^{n} Y_i(t_1)\Lambda\{\beta^T Z_i + H(t_1, \beta, \hat{\gamma}_n)\} = 1,$$

for $H^{(0)}(t_1)$ with $\beta = \beta^{(0)}$. Next, we solve the following equations one-by-one

$$\sum_{i=1}^{n} Y_i(t_k)\left[\Lambda\{\beta^T Z_i + H(t_k, \beta, \hat{\gamma}_n)\} - \Lambda\{\beta^T Z_i + H(t_{k-1}, \beta, \hat{\gamma}_n)\}\right] = 1, \qquad (2.12)$$

for $H^{(0)}(t_k)$ with $\beta = \beta^{(0)}$, $k = 2, 3, \cdots, K$.

Step 4. With the $H = H^{(0)}$ obtained in the previous step, solve (2.11) for a new estimate of β.

Step 5. Set $\beta^{(0)}$ to be the estimate obtained in step 4, repeat steps 3 and 4 until convergence.

The above algorithm results in a single imputation estimator $(\hat{\beta}_j, \hat{H}_j)$. A multiple imputation estimator $(\hat{\beta}, \hat{H})$ can be easily obtained by repeating steps 2 to 5 multiple (m) times and then take the average.

2.6 Asymptotic Properties

Here we establish the asymptotic properties of the imputation estimators, with the assumption that we have correctly specified both the linear transformation model (2.3) and the model $\rho(W, \gamma)$ for the probability that the missing cause is the cause of interest. Similar notation and regularity conditions used in Chen et al. (2002) are also needed here. To avoid technical difficulties resulting from sparsity of data in the tails of the distributions, both T and C are truncated by some finite τ, i.e. with probability one $\mathrm{pr}(T > \tau | Z) > 0$ and $\mathrm{pr}(C > \tau | Z) > 0$. Define $\xi(t) = (\partial/\partial t)\log\lambda(t) = \dot{\lambda}(t)/\lambda(t)$, where $\dot{\lambda}(t) = (\partial/\partial t)\lambda(t)$. We assume that $\lambda(t) > 0\ \forall\ t$, $\xi(t)$ is continuous and $\lim_{u\to-\infty}\lambda(u) = 0 = \lim_{u\to-\infty}\xi(u)$. Without loss of generality, We also assume that $Z \leq M$ almost surely, for some constant M, and that H_0 have continuous and positive derivatives. Then for any $t, s \in (0, \tau]$, define,

$$B(t, s) = \exp\left(\int_s^t \frac{E[\dot{\lambda}\{\beta_0^T Z + H_0(x)\}Y(x)]}{E[\lambda\{\beta_0^T Z + H_0(x)\}Y(x)]} dH_0(x)\right),$$

$$\mu_z(t) = \frac{E[Z\lambda\{\beta_0^T Z + H_0(T)\}Y(t)B(t,T)]}{E[\lambda\{\beta_0^T Z + H_0(t)\}Y(t)]},$$

$$\Sigma_* = \int_0^\tau E[\{Z - \mu_z(t)\}Z^T \dot{\lambda}\{\beta_0^T Z + H_0(t)\}Y(t)]dH_0(t), \tag{2.13}$$

$$\Sigma^* = \int_0^\tau E[\{Z - \mu_z(t)\}^{\otimes 2}\lambda\{\beta_0^T Z + H_0(t)\}Y(t)]dH_0(t), \tag{2.14}$$

where for any vector a, $a^{\otimes 2} = aa^T$. Realize that here both Σ_* and Σ^* are identically defined as the ones in Chen et al.(2002), where no missingness was considered. Also define

$$\begin{aligned} F &= E[\{Z - \mu_z(t)\}\mathrm{pr}(R = 0|W)\rho_\gamma^T(W, \gamma_0)] - \\ &\quad E\left(\int \{Z - \mu_z(t)\}Y(t)\left[\dot{\lambda}\{\beta_0^T Z + H_0(t)\}\frac{\partial H_0(t)}{\partial \gamma}dH_0(t) \right.\right. \\ &\quad \left.\left. + \lambda\{\beta_0^T Z + H_0(t)\}d\frac{\partial H_0(t)}{\partial \gamma}\right]\right) \end{aligned}$$

Let $\rho_\gamma(W_i, \gamma) = (\partial/\partial\gamma)\rho(W_i, \gamma)$ be the partial derivative. Also let

$$I_\gamma = E\left[\frac{I(\Delta_i > 0, R_i = 1)\rho_\gamma(W_i, \gamma_0)\rho_\gamma^T(W_i, \gamma_0)}{\rho(W_i, \gamma_0)\{1 - \rho(W_i, \gamma_0)\}}\right]$$

and

$$S_{\gamma i} = \frac{R_i I(\Delta_i > 0)\{I(\Delta_i = 2) - \rho(W_i, \gamma_0)\}\rho_\gamma(W_i, \gamma_0)}{\rho(W_i, \gamma_0)\{1 - \rho(W_i, \gamma_0)\}}$$

be respectively, the fisher's information matrix and the score vector of $\hat{\gamma}_n$.

THEOREM 1: *Each single imputation estimator, $\hat{\beta}_j$, $(1 \le j \le m)$, is consistent and*

$$n^{1/2}(\hat{\beta}_j - \beta_0) \to N\{0, \Sigma_*^{-1}V_{SI}(\Sigma_*^{-1})^T\},$$

where

$$V_{SI} = \Sigma^* + F^T I_{\gamma_0}^{-1} F + 2F^T I_{\gamma_0}^{-1} E[\rho_\gamma^T(W, \gamma_0)pr(R = 1, \Delta > 0|W)\{Z - \mu_z(T)\}]. \tag{2.15}$$

The second and third terms in V_{SI} represent the extra variation due to the estimation of γ with $\hat{\gamma}_n$ using logistic regression models. As it is shown in Appendix, without missing cause of failure, F equals zero, consequently both the second and third terms equal zero, leaving $V_{SI} = \Sigma^*$ and consequently $n^{1/2}(\hat{\beta}_j - \beta_0) \rightarrow N\{0, \Sigma_*^{-1}\Sigma^*(\Sigma_*^{-1})^T\}$, which is identical to the results given in Chen et al.(2002). It is easy to check that, in the special case of the proportional hazards model,

$$\Sigma^* = \Sigma_* = \text{var}\left[\int \{Z - \mu(t) dM(t)\right],$$

which is exactly Fisher's information matrix for the Cox partial likelihood estimator.

We now discuss the asymptotic properties of the multiple imputation estimators. The main difficulty we come across is the problem of estimating the asymptotic variance of the multiple imputation estimator. The method suggested by Rubin (1987) for estimating the asymptotic variance of the average of any quantity from m completed data sets does not work here; because the imputations are generated from the conditional distribution of D_i given the observed data and evaluated at the fixed value $\hat{\gamma}_n$ across j. Hence our imputation is not proper in the sense of Rubin (1987). However, according to Wang and Robins (1998), under these conditions, they refer this to as type B multiple imputation, and Rubin's variance expression yields estimators that are inconsistent for the sampling variance. In the next theorem, we propose a variance estimator that takes all sources of variability into account and hence it is consistent.

THEOREM 2: *The multiple imputation estimator $\hat{\beta} = \sum_{j=1}^{m} \hat{\beta}_j$, based on m completed data sets, is consistent and*

$$n^{1/2}(\hat{\beta} - \beta_0) \rightarrow N\{0, \Sigma_*^{-1}V_{MI}(\Sigma_*^{-1})^T\},$$

where

$$V_{MI} = V_{SI} - (1 - m^{-1})E[\{Z - \mu_z(T)\}^{\otimes 2} pr(R = 0|W)\rho(W, \gamma_0)\{1 - \rho(W, \gamma_0)\}]. \quad (2.16)$$

It is clear that V_{MI} reduces to V_{SI} when $m = 1$, i.e. single imputation. Thus the second term in V_{MI} measures the reduction of variability of the multiple imputation estimator over the single imputation estimators due to the extra effort for imputing the missing causes multiple times. Although it is evident that more imputations lead to more efficient estimators, imputations however increase the computational intensity of the procedure. Practical number of imputations depends on the magnitude of V_{SI} and the second term.

The consistency of the single imputation estimators can be proved using arguments that are almost identical to that of Lu and Ying (2004). Consequently, the multiple imputation estimators are also consistent, as they are averages of consistent estimators. The proof of the asymptotic normality of the estimators is given in the Appendix.

Consistent estimates of the quantities above can be easily obtained using the usual plug-in methods on every single completed data set, and averaging them across the m imputations. For instance, for jth completed data set, $\hat{\Sigma}_*$ and $\hat{\Sigma}^*$ are respectively,

$$\hat{\Sigma}_* = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{Z_i - \hat{\mu_z}(t)\}Z_i^T\dot{\lambda}\{\hat{\beta}_j^T Z_i + \hat{H}_j(t,\hat{\beta}_j,\hat{\gamma}_n)\}Y(t)d\hat{H}_j(t,\hat{\beta}_j,\hat{\gamma}_n),$$

$$\hat{\Sigma}^* = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{Z_i - \hat{\mu_z}(t)\}^{\otimes 2}\lambda\{\hat{\beta}_j^T Z_i + \hat{H}_j(t,\hat{\beta}_j,\hat{\gamma}_n)\}Y(t)d\hat{H}_j(t,\hat{\beta}_j,\hat{\gamma}_n).$$

Then the multiple imputation version of these two quantities are just the average of the single imputation estimates across the m imputations. Some quantities can even be obtained using software output, e.g. $\hat{I} = \sigma(\hat{\gamma}_n)/n$, where σ is just the variance estimate of $\hat{\gamma}_n$ that can be found in the SAS output.

2.7 Simulations

Two simulation studies were conducted to evaluate the performance of the single and multiple imputation estimators. The first simulation considered the proportional hazards model as a special case. The design of the study was as follows. $H_0(t)$ was chosen to be $\log(t)$ and the random error term ε was chosen with hazard function $\lambda(t) = \exp(t)$. For simplicity, we considered a single binary covariate Z to be equal to 0 or 1 each with probability 0.5. For example, Z may be the treatment indicator in a randomized study. The cause of interest, T_2, conditioned on Z, followed the distribution with hazard function $\lambda_2(t|z) = \lambda e^{\beta z}$, while the other cause, T_1, conditioned on Z, followed a Gompertz distribution with hazard $\lambda_1(t|z) = \lambda e^{\xi+\eta t}$. Since λ will be canceled out in the computation, we took it to be 1 for simplicity. The censoring variable C was generated independent of Z. C followed a uniform distribution $U(\tau_f, \tau_a + \tau_f)$. Although auxiliary covariates A could be used to generate $\rho(W, \gamma)$, for simplicity here we did not consider any auxiliary covariates. Thus by (2.8), logit $\{\rho(W)\} = -\xi - \eta T + \beta Z$. We chose the parameter values to be $\beta = -1$, $\xi = -1$, $\eta = 0.5$, $\tau_a = 0.5$, and $\tau_f = 1.5$. This, on average, gave 48% failures from the cause of interest, 39% failures from the other cause, and the rest 13% from censored observations. There were, on average, 52% failures with missing cause of failure, where the missingness were generated by the model $\pi(W, \psi) = \text{logit } \text{pr}(R = 0|\Delta > 0, W, \psi) = \psi_0 + \psi_1 T + \psi_2 Z$, with parameter values $\psi_0 = 1$, $\psi_1 = -1$ and $\psi_2 = -1$.

In the second simulation study, the proportional odds model was considered. Here $H_0(t) = \log(t)$ and ε had hazard function $\lambda(t) = \exp(t)/\{1 + \exp(t)\}$. Both Z and C were the same as that in simulation 1. T_2 followed a distribution with hazard function $\lambda_2(t|z) = \lambda \exp(\beta z)/\{1 + t \exp(\beta z)\}$, and T_1 had hazard function $\lambda_1(t|z) = \lambda e^{\xi+\eta t}$ where $\lambda = 1$. Again, by (2.8), $\rho(W) = \{1 + t\exp(\beta z) + \exp(-\xi - \eta t + \beta z)\}^{-1} \exp(-\xi - \eta t + \beta z)$. The missingness mechanism was also determined by $\pi(W, \psi)$ defined in simulation 1. We chose $\beta = -1$, $\xi = -1.5$, $\eta = 0.5$, $\tau_a = 1$, $\tau_f = 2$, $\psi_0 = 1.5$, $\psi_1 = -1$ and $\psi_2 = -1$. This setting leaded to, on average, 46% failures from the cause of interest, 41% failures from the other cause, 13% from censored observations, and 49% failures had missing cause information. Again, we purposely misspecified $\pi(W)$ or $\rho(W)$ or both to be constants independent of the W.

In both simulations, for sample size $n = 200$, 1000 Monte-Carlo simulations were conducted to demonstrate the properties of the single imputation (SI) estimator and the multiple ($m = 10$) imputation (MI) estimator, as well as to compare them with the Biased Complete-case (BCC) estimator that solves (2.6) and (2.7), ignoring the observations with missing information, the inverse probability weighted approaches IPWCC and DR proposed by Gao and Tsiatis (2005). The results were summarized in Tables 2.1 and 2.2 that includes the bias, the Monte Carlo standard error (SSE), the Monte Carlo average of the estimated standard deviations (SEE) and the empirical coverage probability (CP) of the 95% confidence intervals of the estimators.

The simulation results were consistent with the theoretical results. In both studies, BCC was severely biased and suffered from poor empirical coverage probability. While both SI and MI were virtually unbiased, with smaller standard errors and empirical coverage probabilities attaining the 95% nominal level, as long as $\rho(W)$ is correctly specified. The performance of SI was on par with that of MI, although MI had slightly smaller standard errors. We observed that the computational intensity of MI with $m = 10$ was at least 3 times as much as that of SI. The trade off between the efficiency gain and the cost of computation seemed to be in favor of the single imputation method. MI seemed to be as efficient as DR. It was also clear that for both SI and MI, SEE was very closed to SSE, which justifies the accuracy of the estimator of the asymptotic variance.

2.8 Breast Cancer Study

We here illustrate the multiple imputation procedure using the data from a clinical trial in stage II breast cancer. Among the total 169 patients enrolled in the study, 90 were censored; and among the rest 79 death: 18 patients died with unknown cause of death, 44 died of breast cancer and the rest 17 died of other causes. The objective of this study was to identify covariates that are significantly related to death of breast cancer.

With a complete-case analysis, Cummings et al. (1986) found that two covariates, being ER-negative and presence of more than four positive nodes, were significantly associated with overall survival. For each failure type, Goetghebeur and Ryan (1995) conducted a cause-specific survival analysis with proportional hazards assumption. Also using the proportional hazards model, Lu and Tsiatis (2001) assumed a logistic regression model for $\rho(W)$ and then applied their multiple imputation method. Gao and Tsiatis (2005) derived doubly robust (DR) estimators using linear transformation competing risks models. We summarize their results in Table 2.3 along with our multiple imputation estimators that was derived in the following way. First, we assumed a logistic regression model for $\rho(W)$ as functions of observed covariates W, which included ER status, tumor size, number of positive notes and time of death. As that was reported in Lu and Tsiatis (2001), since all of the five patients with ER-negative status died from breast cancer, our logistic regression model was derived using the subset of patients who had positive ER status and they died with known cause. We then computed the multiple imputation estimates and their variance estimates, for the proportional hazards model and the proportional odds model. We find that the results given by our multiple imputation method are very similar to those using the DR method proposed by Gao and Tsiatis (2005), and they are both closed to the results reported in the literature.

2.9 Discussion

We have proposed an imputation procedure for estimating the regression parameters in the class of linear transformation competing risks models when the cause of failure was missing for a subset of patients. Simulation studies confirmed the consistency and asymptotic normality of the imputation estimators as well as the consistency of the estimator for the asymptotic variance, for two important special cases: the proportional hazards model and the proportional odds model. The authors also noticed that the flexibility of the class of linear transformation models has raised the problem of selecting appropriate models in applications. Research on model diagnostic or model selection techniques are critical and part of ongoing research.

Table 2.1: Simulation study comparing estimators using complete-case analysis, inverse probability weighted approaches and imputation methods, for the proportional hazards model, based on 1,000 replications, sample size=200

	Correct $\pi(.)$ & Correct $\rho(.)$				
	BCC	IPWCC	DR	SI	MI
Bias	-0.347	-0.029	-0.031	-0.035	-0.035
SSE	0.376	0.351	0.319	0.335	0.310
SEE	0.354	0.375	0.308	0.338	0.313
CP	0.864	0.957	0.948	0.960	0.955
	Correct $\pi(.)$ & Incorrect $\rho(.)$				
	BCC	IPWCC	DR	SI	MI
Bias	-0.347	-0.029	-0.027	0.279	0.276
SSE	0.376	0.351	0.331	0.203	0.178
SEE	0.354	0.375	0.320	0.204	0.179
CP	0.864	0.957	0.947	0.647	0.709
	Incorrect $\pi(.)$ & Correct $\rho(.)$				
	BCC	IPWCC	DR	SI	MI
Bias	-0.347	-0.118	-0.028	-0.035	-0.035
SSE	0.376	0.338	0.306	0.335	0.310
SEE	0.354	0.326	0.297	0.338	0.313
CP	0.864	0.878	0.947	0.960	0.955
	Incorrect $\pi(.)$ & Incorrect $\rho(.)$				
	BCC	IPWCC	DR	SI	MI
Bias	-0.347	-0.118	0.086	0.279	0.276
SSE	0.376	0.338	0.244	0.203	0.178
SEE	0.354	0.326	0.232	0.204	0.179
CP	0.864	0.878	0.922	0.647	0.709

Table 2.2: Simulation study comparing estimators using complete-case analysis, inverse probability weighted approaches and imputation methods, for the proportional odds model, based on 1,000 replications, sample size=200

	Correct $\pi(.)$ & Correct $\rho(.)$				
	BCC	IPWCC	DR	SI	MI
Bias	-0.214	0.041	-0.024	-0.018	-0.020
SSE	0.403	0.429	0.389	0.369	0.352
SEE	0.411	0.449	0.378	0.385	0.360
CP	0.935	0.955	0.947	0.957	0.959
	Correct $\pi(.)$ & Incorrect $\rho(.)$				
	BCC	IPWCC	DR	SI	MI
Bias	-0.214	0.041	-0.018	0.232	0.231
SSE	0.403	0.429	0.417	0.290	0.252
SEE	0.411	0.449	0.392	0.294	0.260
CP	0.935	0.955	0.953	0.879	0.872
	Incorrect $\pi(.)$ & Correct $\rho(.)$				
	BCC	IPWCC	DR	SI	MI
Bias	-0.214	-0.099	-0.022	-0.018	-0.020
SSE	0.403	0.395	0.361	0.369	0.352
SEE	0.411	0.363	0.348	0.385	0.360
CP	0.935	0.929	0.945	0.957	0.959
	Incorrect $\pi(.)$ & Incorrect $\rho(.)$				
	BCC	IPWCC	DR	SI	MI
Bias	-0.214	-0.099	0.094	0.232	0.231
SSE	0.403	0.395	0.321	0.290	0.252
SEE	0.411	0.363	0.333	0.294	0.260
CP	0.935	0.929	0.912	0.879	0.872

Table 2.3: Comparison of complete-cases, Goetghebeur and Ryan, Lu and Tsiatis, and multiple imputation for both the proportional hazards model and the proportional odds model, using the breast cancer data

	Effective Covariates	
Methods	ER Status	Great Than 4 Nodes
BCC (PH Model)	$0.71^a(0.3065^b)$	1.70(0.4861)
GR (PH Model)	0.57(0.2803)	1.59(0.4822)
LT (PH Model)	0.60(0.2618)	1.61(0.4794)
DR (PH Model)	0.53(0.2865)	1.60(0.5418)
MI^c (PH Model)	0.54(0.2988)	1.60(0.5763)
DR (PO Model)	0.60(0.3706)	2.10(0.6747)
MI^c (PO Model)	0.62(0.3425)	2.10(0.6982)

^{a} Estimate
^{b} Standard Error
$^c m = 10$
PH denotes Proportional Hazards Model
PO denotes Proportional Odds Model

Chapter 3

Summary

For the problem of estimating regression coefficients in semiparametric models for survival data with missing cause of failure on a subset of individuals, so far we have briefly reviewed a number of approaches in the literature: the biased complete case (BCC) method, the method by Goetghebeur and Ryan (1995)(GR), the multiple imputation procedure by Lu and Tsiatis (2001)(LT); we have also proposed two new methods: the multiple imputation (MI) procedure, and the inverse probability weighting (IPW) method that includes the inverse probability weighted complete-case (IPWCC) method and the doubly robust (DR) approach, for linear transformation competing risks models. In this chapter, we first compare these methods in terms of their flexibility, efficiency and robustness against mis-specification of the models $\pi(W)$ and $\rho(W)$. These comparisons naturally lead to suggestions and instructions for using these methods in applications. Lastly, we will discuss some directions for future research.

3.1 Comparisons

3.1.1 Flexibility

A major advantage of the two approaches that we have proposed over the existing methods is, that our approaches work for arbitrary linear transformation models, which include the proportional hazards model and the proportional odds model as special cases; while the existing methods only work for the proportional hazards model, except for the biased complete case method. This suggests that the proposed methods are used for a class of models that offers more flexibility in model building. The results are summarized in Table 3.1.

Next, we would like to find out which approaches allow using auxiliary covariates that make the missing at random assumption more tenable. Since the BCC approach assumes missing completely at random (MCAR), it does not use any auxiliary covariates. Neither does the GR approach as each covariate is used only to model either of the two cause specific hazard functions. On the other hand, the LT approach, MI method and DR approach fully exploit the information contained in the auxiliary covariates and include them in modeling $\rho(W)$. Moreover, the IPW approaches allow using auxiliary covariates in modeling the missing mechanism, i.e. $\pi(W)$. These results are summarized in Table 3.2.

Finally, we want to discuss the corresponding missing data mechanism required to validate the inference of each approach. In order to use the BCC method, one has to assume that the missing data are MCAR, which is not very practical in applications. The GR approach assumes MAR but the missingness probability could not depend on covariates, which is a rather restrictive assumption, and it is denoted by RMAR. The LT approach and our approaches MI and IPW all assume MAR and the missingness probability may depend on covariates. If the missingness is Non-Ignorable Non-Response (NINR, Rubin, 1987), none of these approaches work. We summarize the results in Table 3.3.

3.1.2 Efficiency

In order to compare the efficiency, we have to make the comparison based on the assumption that all approaches are valid. For instance, to compare the LT approach and the IPWCC approach, we have to assume that the cause-specific hazards follow the proportional hazards model, the missingness is MAR, and the models for $\pi(W)$ and $\rho(W)$ are both correctly specified for both approaches. Because BCC does not use any information contained in the missing data, it is the least efficient. Since LT, MI, IPWCC and DR make use of additional information from the competing cause, they are more efficient. The GR estimator is very efficient under the specific assumptions on the hazards. LT is expected to be as efficient as MI. The DR estimator is more efficient than IPWCC estimator as it was shown theoretically in §1.5.2. Finally, we want to point out that it is very difficult to compare the efficiency of the MI estimator and the DR estimator, either theoretically or intuitively. We summarize the results in Table 3.4.

The above theoretical results about the efficiency comparison of the proposed methods MI, IPWCC and DR are confirmed by the simulations that are discussed in §1.6 and §2.7. We summarize in Table 3.5 the results for the case, where both $\pi(W)$ and $\rho(W)$ are correctly specified and MAR holds. It is evident that the IPWCC estimator is less efficient, while the MI estimator is on par with the DR estimator in terms of efficiency.

3.1.3 Robustness

Since in applications it is very common that the missingness probability depends on both time and covariates where both BCC method and GR approach lead to biased estimators, we here focus on assessing the robustness of the approaches LT, MI, IPWCC and DR for the following two cases. Case 1: The model for the probability of missingness $\pi(W)$ is correctly specified, while the model $\rho(W)$ for the probability that the missing cause is the cause of interest is misspecified. Case 2: $\rho(W)$ is truly specified but $\pi(W)$ is misspecified.

For case 1, the imputation methods LT and MIE are not expected to work well, but both IPWCC and DR work very well as confirmed by simulations. For case 2, IPWCC does not perform well while LT, MIE and DR all perform well as confirmed by simulations. These results are summarized in Tables 3.6, 3.7 and 3.8.

3.2 Recommendations

Based on the discussion above, the DR estimator is recommended to be used for general applications, as DR allows using auxiliary covariates, is suitable for arbitrary linear transformation models, is valid under the commonly used general MAR assumptions, is highly efficient, and is doubly robust against mis-specification of either one of the models $\pi(W)$ and $\rho(W)$.

3.3 Further research

We have not yet studied the situations where the missingness is NINR. In addition, model diagnostic techniques for detecting the most appropriate transformation model in the presence of missing cause of failure is a new research area.

Table 3.1: Models for hazard functions

	BCC	GR	LT	IPWCC	DR	MI
PH	Yes	Yes	Yes	Yes	Yes	Yes
LTM[a]	Yes	No	No	Yes	Yes	Yes

[a] Linear Transformation Models.

Table 3.2: Inculsion of Auxiliary Covariates

BCC	GR	LT	IPWCC	DR	MI
No	No	Yes	Yes	Yes	Yes

Table 3.3: Missing Data Mechanism

	BCC	GR	LT	IPWCC	DR	MI
MCAR	Yes	Yes	Yes	Yes	Yes	Yes
RMAR	No	Yes	Yes	Yes	Yes	Yes
MAR	No	No	Yes	Yes	Yes	Yes
NINR	No	No	No	No	No	No

Table 3.4: Efficiency

BCC	GR	LT	IPWCC	DR	MI
Poor	Excellent	Excellent	Good	Excellent	Excellent

Table 3.5: Efficiency–Simulation Results: both $\pi(W, \psi)$ and $\rho(W, \gamma)$ are correctly specified

	Proportional Hazards Model			Proportional Odds Model		
	IPWCC	DR	MI	IPWCC	DR	MI
SSE	0.351	0.319	0.310	0.429	0.389	0.352
SEE	0.375	0.308	0.313	0.449	0.378	0.360

Table 3.6: Robustness

$\pi(W)$	$\rho(W)$	BCC	GR	LT	IPWCC	DR	MI
Correct	Incorrect	No	No	No	Yes	Yes	No
Incorrect	Correct	No	No	Yes	No	Yes	Yes

Table 3.7: Robustness Comparison: The Proportional Hazards Model

	Incorrect $\pi(.)$ & Correct $\rho(.)$				Correct $\pi(.)$ & Incorrect $\rho(.)$			
	BCC	IPWCC	DR	MI	BCC	IPWCC	DR	MI
Bias	-0.347	-0.118	-0.028	-0.035	-0.347	-0.029	-0.027	0.276

Table 3.8: Robustness Comparison: The Proportional Odds Model

	Incorrect $\pi(.)$ & Correct $\rho(.)$				Correct $\pi(.)$ & Incorrect $\rho(.)$			
	BCC	IPWCC	DR	MI	BCC	IPWCC	DR	MI
Bias	-0.214	-0.099	-0.022	-0.020	-0.214	0.041	-0.018	0.231

Bibliography

ANDERSEN, J., GOETGHEBEUR, E. & RYAN, L. (1996). Missing cause of death information in the analysis of survival data. *Statist. Med.* **15**, 2191–201.

BENNETT, S. (1983). Analysis of survival data by the proportional odds model. *Statist. Med.* **2**, 273–7. 7

CAI, T., WEI, L.J. & WILCOX, M. (2000). Semiparametric regression analysis for clustered failure time data. *Biometrika.* **87**, 867–78.

CHEN, K., JIN, Z. & YING, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika.* **89**, 659–68.

CHENG, S.C., WEI, L.J. & YING, Z. (1995). Analysis of transformation models with censored data. *Biometrika.* **82**, 835–45.

CHENG, S.C., WEI, L.J. & YING, Z. (1997). Prediction of survival probabilities with semi-parametric transformation models. *J. Am. Statist. Assoc.* **92**, 227–35.

COX, D.R. (1972). Regression models and life tables (with Discussion). *J.R.Statist.Soc.* **B.34**, 187–220.

COX, D.R. (1975). Partial likelihood. *Biometrika* **62**, 269–76.

COX, D.R. & OAKES, D. (1984). *Analysis of Survival Data.* New York: Chapman and Hall.

CUMMINGS, F.J., GRAY, R., DAVIS, T.E. , TORMEY,D.C, HARRIS, J.E., FRALKSON,G.G., & ARSENEAU, J. (1986) Tamoxifen versus placebo: double blind adjuvant trial in elderly women with stage II breast cancer. *National Cancer Institute Monographs* **1**, 119–23.

Fine, J.P., Ying, Z. & Wei, L.J. (1998). On the linear transformation model with censored data. *Biometrika.* **85**, 980–6.

Fleming, T.R., & Harrington, D.P. (1991). *Counting Process and Survival Analysis* New York: Wiley.

Gao, G. and Tsiatis, A.A. (2005) Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure. *Biometrika*, In Revision.

Goetghebeur, E. & Ryan, L. (1995) Analysis of competing risks survival data when some failure types are missing. *Biometrika.* **82**, 821–34.

Holt, J.D. (1978) Competing risk analysis with special reference to matched pair experiments . *Biometrika* **65**, 159–65.

Kalbfleisch, J.D., & Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Lu, K. & Tsiatis, A.A. (2001) Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics.* **57**, 1191–97.

Lu, K. & Tsiatis, A.A. (2005) Semiparametric efficient estimation in the competing risks model with missing cause of failure. *Lifetime Data Analysis.* In Press.

Lu, W. & Ying, Z. (2004) On semiparametric transformation cure models. *Biometrika.* **91**, 331–43.

MURPHY, S.A., ROSSINI, A.J. & VAN DER VAART, A.W.(1997). Maximum likelihood estimation in the proportional odds model. *J. Am. Statist. Assoc.* **92**, 968–76.

PETTITT, A.N. (1982). Inference for the linear model using a likelihood based on ranks. *J. R. Statist. Soc. B* **44**, 234–43.

PETTITT, A.N. (1984). Proportional odds model for survival data and estimates using ranks. *Appl. Statist.* **33**, 169–75.

ROBINS, J.M., ROTNITZKY, A., & ZHAO, L.P. (1994). Estimation of Regression Coefficients When Some Regressors Are Not Always Observed. *J. Am. Statist. Assoc.* **89**, 846–66.

RUBIN, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581–92.

RUBIN, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

RUBIN, D.B. (1996). Multiple imputation after 18+ years. *J. Am. Statist. Assoc.* **91**, 473–89.

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

TSIATIS, A.A. (2004). Personal Communication.

TSIATIS, A.A. Unpublished lecture notes.

VAN DER VAART, A.W. (2000). *Asymptotic Statistics.* Cambridge: Cambridge University Press.

WANG, N. & ROBINS, J.M. (1998) Large sample inference in parametric multiple imputation. *Biometrika.* **85**, 935–948.

Appendices

Appendix A

Proof of the Proposition in §1.5.2

Usual regularity conditions for martingale central limit theorem as those in Fleming & Harrington (1991) are assumed.

Let $U\{O, \beta, H, \hat{\psi}_n, \hat{\gamma}_n\}$ denote the left hand side of (1.13), the first step is to show that the gradient matrix of $U\{O, \beta, H, \hat{\psi}_n, \hat{\gamma}_n\}$ with respect to β, converges in probability to $-A$ that was defined in the Proposition, i.e.,

$$\frac{1}{n}\frac{\partial}{\partial \beta} U\{O, \beta, \hat{H}_n^{DR}(\cdot, \beta, \hat{\psi}_n, \hat{\gamma}_n), \hat{\psi}_n, \hat{\gamma}_n\}|_{\beta=\beta_0} \xrightarrow{P} -A.$$

Let $a > 0$ and b be fixed finite numbers (e.g. lower limits of the integration to ensure finite integrals). Define the following quantities, which were also used in Chen et al. (2002):

$$\lambda^*\{H_0(t)\} = B(t, a), \qquad B_1(t) = \int_a^t E[Y(s)\dot{\lambda}\{\beta^T Z + H_0(s)\}]dH_0(s),$$

$$B_2(t) = E[Y(t)\lambda\{\beta^T Z + H_0(t)\}], \qquad \Lambda^*(t) = \int_b^t \lambda^*(s)ds.$$

By mimicking Steps A2 and A3 in Chen et al. (2002), we have

$$\Lambda^*\{\hat{H}_0(t)\} - \Lambda^*\{H_0(t)\} = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)} dM_i^*(s) + o_p(n^{-1/2}), \tag{A.1}$$

$$\frac{\partial \hat{H}_n^{DR}(t, \beta, \hat{\psi}_n, \hat{\gamma}_n)}{\partial \beta}|_{\beta=\beta_0} = -\int_0^t \frac{E[Z\dot{\lambda}\{\beta_0^T Z_i + H_0(s)\}Y(s)]}{\lambda^*\{H_0(t)\}} \frac{\lambda^*\{H_0(s)\}}{B_2(s)} dH_0(s) + o_p(1), \tag{A.2}$$

$$
\frac{\partial \hat{H}_n^{DR}(t,\beta,\hat{\psi}_n,\hat{\gamma}_n)}{\partial \psi}|_{\psi=\psi^*} = \int_0^t \frac{\lambda^*\{H_0(u)\}}{\lambda^*\{H_0(t)\}B_2(u)} E\left[\frac{R\pi_\psi^T\{\rho(W,\gamma^*)-I(\Delta=2)\}}{\pi^2(Q,\psi^*)} dN^*(u)\right] + o_p(1), \tag{A.3}
$$

$$
\frac{\partial \hat{H}_n^{DR}(t,\beta,\hat{\psi}_n,\hat{\gamma}_n)}{\partial \gamma}|_{\gamma=\gamma^*} = \int_0^t \frac{-\lambda^*\{H_0(u)\}}{\lambda^*\{H_0(t)\}B_2(u)} E\left[\frac{\{R-\pi(Q,\psi^*)\}\rho^T(W,\gamma^*)}{\pi(Q,\psi^*)} dN^*(u)\right] + o_p(1). \tag{A.4}
$$

Define $\Phi(O,\psi,\gamma) = [RI(\Delta = 2)/\pi(Q,\psi) - \{R-\pi(Q,\psi)\}/\pi(Q,\psi)\rho(W,\gamma)]$, then by mimicking *Step* A3 in Chen et al. (2002), and by the law of large numbers,

$$
\begin{aligned}
& \frac{1}{n}\frac{\partial U\{O,\beta,\hat{H}_n^{DR}(\cdot,\beta,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\}}{\partial \beta}|_{\beta=\beta_0} \\
= & \frac{1}{n}\frac{\partial \sum_{i=1}^n \int_0^\tau Z_i\Phi(O_i,\hat{\psi}_n,\hat{\gamma}_n)dN_i^*(t)}{\partial\beta}|_{\beta=\beta_0} \\
& -\frac{1}{n}\frac{\partial \sum_{i=1}^n Z_iY_i(t)\lambda\{\beta^TZ_i+\hat{H}_n^{DR}(t,\beta,\hat{\psi}_n,\hat{\gamma}_n)\}\hat{H}_n^{DR}(t,\beta,\hat{\psi}_n,\hat{\gamma}_n)}{\partial\beta}|_{\beta=\beta_0} \\
= & -\frac{1}{n}\sum_{i=1}^n \frac{\partial \int_0^\tau Z_iY_i(t)\lambda\{\beta^TZ_i+\hat{H}_n^{DR}(t,\beta,\hat{\psi}_n,\hat{\gamma}_n)\}d\hat{H}_n^{DR}(t,\beta,\hat{\psi}_n,\hat{\gamma}_n)}{\partial\beta}|_{\beta=\beta_0} \\
= & -\frac{1}{n}\sum_{i=1}^n Z_i\lambda\{\beta_0^TZ_i+\hat{H}_n^{DR}(T_i,\beta,\hat{\psi}_n,\hat{\gamma}_n)\}\{Z_i+\frac{\partial\hat{H}_n^{DR}(T_i,\beta,\hat{\psi}_n,\hat{\gamma}_n)}{\partial\beta}|_{\beta=\beta_0}\}Y_i(T_i) \\
= & -\int_0^\tau E\{(Z-\frac{E[Z\lambda\{\beta_0^TZ+H_0(T)\}Y(s)/\lambda^*\{H_0(T)\}]}{B_2(s)/\lambda^*\{H_0(s)\}}) \\
& \times[Z^T\dot{\lambda}\{\beta_0^TZ+H_0(s)\}Y(s)]\}dH_0(s)+o_p(1) \\
= & -A+o_p(1),
\end{aligned}
$$

where A is defined in (1.18), and the fourth equality follows from (A.2).

To complete the proof of the proposition, we here go through a series of expansions. We first expand $U\{O,\beta_0,\hat{H}_n^{DR}(\cdot,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\}$ at $\hat{H}_n^{DR}(\cdot,\beta_0,\hat{\psi}_n,\hat{\gamma}_n)$ about $H_0(\cdot,\beta_0,\hat{\psi}_n,\hat{\gamma}_n)$.

$$
\begin{aligned}
& U\{O,\beta_0,\hat{H}_n^{DR}(t,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\} \\
= & \sum_{i=1}^n \int_0^\tau Z_i dM_i^*\{t,Z_i,\beta_0,\hat{H}_n^{DR}(t,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}\int_0^{\tau} Z_i dM_i^*\{t, Z_i, \beta_0, H_0(t,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\} \\
&\quad -\sum_{i=1}^{n} Z_i\left[\Lambda\{\beta_0^T Z_i + \hat{H}_n^{DR}(T_i,\beta_0,\hat{\psi}_n,\hat{\gamma}_n)\} - \Lambda\{\beta_0^T Z_i + H_0(T_i,\beta_0,\hat{\psi}_n,\hat{\gamma}_n)\}\right] \\
&= \sum_{i=1}^{n}\int_0^{\tau} Z_i dM_i^*\{t, Z_i, \beta_0, H_0(t,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\} \\
&\quad -\sum_{i=1}^{n} Z_i\{\lambda\{\beta_0^T Z_i + \hat{H}_n^{DR}(T_i,\beta_0,\hat{\psi}_n,\hat{\gamma}_n)\}\{\hat{H}_n^{DR}(T_i,\beta_0,\hat{\psi}_n,\hat{\gamma}_n) - H_0(T_i,\beta_0,\hat{\psi}_n,\hat{\gamma}_n)\} \\
&= \sum_{i=1}^{n}\int_0^{\tau} Z_i dM_i^*\{t, Z_i, \beta_0, H_0(t,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\} \\
&\quad -\sum_{i=1}^{n}\left[Z_i \frac{\lambda\{\beta_0^T Z_i + H_0(T_i,\beta_0,\hat{\psi}_n,\hat{\gamma}_n)\}}{\lambda^*\{H_0(T_i,\beta_0,\hat{\psi}_n,\hat{\gamma}_n)\}} n^{-1}\sum_{j=1}^{n}\int_0^{T_i}\frac{\lambda^*\{H_0(u,\beta_0,\hat{\psi}_n,\hat{\gamma}_n)\}}{B_2(u)}\right. \\
&\quad \left.\times dM_i^*\{u, Z_i, \beta_0, H_0(t,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\}\right] + o_p(n^{1/2}) \\
&= \sum_{i=1}^{n}\int_0^{\tau} Z_i dM_i^*\{t, Z_i, \beta_0, H_0(t,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\} \\
&\quad -\sum_{j=1}^{n}\int_0^{\tau} E\left[\frac{Z\lambda\{\beta_0^T Z_i + H_0(T,\beta_0,\hat{\psi}_n,\hat{\gamma}_n)\}Y(t)}{\lambda^*\{H_0(T,\beta_0,\hat{\psi}_n,\hat{\gamma}_n)\}}\right]\frac{\lambda^*\{H_0(t,\beta_0,\hat{\psi}_n,\hat{\gamma}_n)\}}{B_2(t)} \\
&\quad \times dM_i^*\{u, Z_i, \beta_0, H_0(\cdot,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\} + o_p(n^{1/2}) \\
&= \sum_{i=1}^{n}\int_0^{\tau}[Z_i - \mu_z\{t, Z_i, \beta_0, H_0(\cdot,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\}]\times \\
&\quad \times dM_i^*\{t, Z_i, \beta_0, H_0(\cdot,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\} + o_p(n^{1/2}) \\
&= U^*\{O, \beta_0, H_0(\cdot,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\} + o_p(n^{1/2}).
\end{aligned}
$$

By (A.3) and (A.4), we can similarly expand $U^*\{O, \beta_0, H_0(\cdot,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\}$ at $\hat{\psi}_n$ about ψ^* and $\hat{\gamma}_n$ about γ^*. First of all, we derive the partial derivatives of $U^*(\cdot)$ with respect to ψ and γ.

$$
\begin{aligned}
&\frac{\partial U^*\{O, \beta_0, H_0(\cdot,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\}}{\partial\psi}|_{\psi=\psi^*} \\
&= \sum_{i=1}^{n}\int_0^{\tau}\frac{\partial\{Z_i - \mu_z(t,\cdot)\}}{\partial\psi}|_{\psi=\psi^*} dM_i^*\{t, Z_i, \beta_0, H_0(t,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\} \\
&\quad +\sum_{i=1}^{n}\int_0^{\tau}\{Z_i - \mu_z(t,\cdot)\}\frac{\partial dM_i^*\{t, Z_i, \beta_0, H_0(t,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\}}{\partial\psi}|_{\psi=\psi^*}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}\int_0^{\tau}\{Z_i-\mu_z(t,\cdot)\}\frac{\partial dM_i^*\{t,Z_i,\beta_0,H_0(t,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\}}{\partial\psi}|_{\psi=\psi^*}+o_p(1)\\
&= -\sum_{i=1}^{n}\int_0^{\tau}\{Z_i-\mu_z(t,\cdot)\}\frac{\{I(\Delta_i=2)-\rho(W_i,\gamma^*)\}R_i\pi_{\psi}^T(Q_i)}{\pi^2(Q_i,\psi^*)}dN_i^*(t)\\
&\quad -\sum_{i=1}^{n}\int_0^{\tau}\{Z_i-\mu_z(t,\cdot)\}Y_i(t)\dot{\lambda}\{\beta_0^TZ_i+H_0(t)\}/\lambda^*\{H_0(t)\}\\
&\quad \times 1/n\sum_{i=1}^{n}\int_0^{t}\frac{\lambda^*\{H_0(u)\}}{B_2(u)}\frac{R_i\pi_{\psi}^T(Q_i)\{I(\Delta_i=2)-\rho(W_i,\gamma^*)\}}{\pi^2(Q_i,\psi^*)}dN_i^*(u)dH_0(t)\\
&\quad -\sum_{i=1}^{n}\int_0^{\tau}\{Z_i-\mu_z(t,\cdot)\}Y_i(t)\lambda\{\beta_0^TZ_i+H_0(t)\}/\lambda^*\{H_0(t)\}\\
&\quad \times 1/n\sum_{i=1}^{n}\frac{\lambda^*\{H_0(t)\}}{B_2(t)}\frac{R_i\pi_{\psi}^T(Q_i)\{I(\Delta_i=2)-\rho(W_i,\gamma^*)\}}{\pi^2(Q_i,\psi^*)}dN_i^*(t)+o_p(1),
\end{aligned}
$$

where the second equality follows from the fact that $\sum_{i=1}^{n}\int(\partial\{Z_i-\mu_z(t,\cdot)\}/\partial\psi)dM_i^*(t)\xrightarrow{P}0$.

Similarly, we obtain

$$
\begin{aligned}
&\frac{\partial U^*\{O,\beta_0,H_0(\cdot,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\}}{\partial\gamma}|_{\gamma=\gamma^*}\\
&= -\sum_{i=1}^{n}\int_0^{\tau}\{Z_i-\mu_z(t,\cdot)\}\frac{\{R_i-\pi(Q_i,\psi^*)\}\rho^T(W_i,\gamma^*)}{\pi(Q_i,\psi^*)}dN_i^*(t)\\
&\quad +\sum_{i=1}^{n}\int_0^{\tau}\{Z_i-\mu_z(t,\cdot)\}Y_i(t)\dot{\lambda}\{\beta_0^TZ_i+H_0(t)\}/\lambda^*\{H_0(t)\}\\
&\quad \times 1/n\sum_{i=1}^{n}\int_0^{t}\frac{\lambda^*\{H_0(u)\}}{B_2(u)}\frac{\{R_i-\pi(Q_i,\psi^*)\}\rho^T(W_i,\gamma^*)}{\pi(Q_i,\psi^*)}dN^*(u)dH_0(t)\\
&\quad +\sum_{i=1}^{n}\int_0^{\tau}\{Z_i-\mu_z(t,\cdot)\}Y_i(t)\lambda\{\beta_0^TZ_i+H_0(t)\}/\lambda^*\{H_0(t)\}\\
&\quad \times 1/n\sum_{i=1}^{n}\frac{\lambda^*\{H_0(t)\}}{B_2(t)}\frac{\{R_i-\pi(Q_i,\psi^*)\}\rho^T(W_i,\gamma^*)}{\pi(Q_i,\psi^*)}dN^*(t)+o_p(1).
\end{aligned}
$$

Therefore, define

$$
P_{\psi}=-\lim_{n\to\infty}\left[n^{-1}\frac{\partial U^*\{O,\beta_0,H_0(\cdot,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\}}{\partial\psi}|_{\psi=\psi^*}\right]
$$

and

$$
P_{\gamma}=-\lim_{n\to\infty}\left[n^{-1}\frac{\partial U^*\{O,\beta_0,H_0(\cdot,\beta_0,\hat{\psi}_n,\hat{\gamma}_n),\hat{\psi}_n,\hat{\gamma}_n\}}{\partial\gamma}|_{\gamma=\gamma^*}\right].
$$

It is easy to see that $P_\psi = 0$ if $\rho(W_i, \gamma)$ is correct, and $P_\gamma = 0$ if $\pi(W_i, \psi)$ is correct. By a Taylor series expansion we obtain

$$\begin{aligned} U^*\{O, \beta_0, H_0(\cdot, \beta_0, \hat{\psi}_n, \hat{\gamma}_n), \hat{\psi}_n, \hat{\gamma}_n\} &= U^*\{O, \beta_0, H_0(\cdot), \psi^*, \gamma^*\} \\ &\quad - \sum_{i=1}^n (P_\psi I_\psi^{-1} S_{\psi i} + P_\gamma I_\gamma^{-1} S_{\gamma i}) + o_p(1). \end{aligned}$$

Putting the results together we immediately obtain

$$\begin{aligned} & n^{\frac{1}{2}}(\hat{\beta}_n^{DR} - \beta_0) \\ = \; & -A^{-1}[n^{\frac{-1}{2}} U^*\{O, \beta_0, H_0(\cdot), \psi^*, \gamma^*\} \\ & -n^{\frac{-1}{2}} \sum_{i=1}^n \{P_\psi I_\psi^{-1} S_{\psi i} + P_\gamma I_\gamma^{-1} S_\gamma\} + o_p(n^{\frac{-1}{2}})] + o_p(1) \\ = \; & -A^{-1} \left(n^{\frac{-1}{2}} \sum_{i=1}^n \left[\int \{Z_i - \mu_z(t, \cdot)\} dM_i^*(t, \cdot) - P_\psi I_\psi^{-1} S_{\psi i} - P_\gamma I_\gamma^{-1} S_{\gamma i} \right] \right) + o_p(1). \end{aligned}$$

This completes the proof.

Note: P_ψ^{IPWCC} in (1.23) can be defined similarly to P_ψ, namely,

$$\begin{aligned} P_\psi^{IPWCC} &= -\lim_{n\to\infty} \left[n^{-1} \sum_{i=1}^n \frac{\partial \int_0^\tau \frac{R_i}{\pi(Q_i,\psi)} \{Z_i - \mu_z(t)\} dM_i(t)}{\partial \psi} |_{\psi=\psi_0} \right] \\ &= -\lim_{n\to\infty} \left[n^{-1} \sum_{i=1}^n \int_0^\tau \frac{R_i}{\pi(Q_i, \psi)} \{Z_i - \mu_z(t)\} \frac{\partial dM_i(t)}{\partial \psi} |_{\psi=\psi_0} \right] + o_p(1) \\ &= -\lim_{n\to\infty} \left[-\sum_{i=1}^n \int_0^\tau \frac{R_i}{\pi(Q_i)} \{Z_i - \mu_z(t, \cdot)\} Y_i(t) \dot{\lambda}\{\beta_0^T Z_i + H_0(t)\} / \lambda^*\{H_0(t)\} \right. \\ &\quad \times 1/n \sum_{i=1}^n \int_0^t \frac{\lambda^*\{H_0(u)\}}{B_2(u)} \frac{R_i \pi_\psi^T(Q_i)}{\pi^2(Q_i)} dM_i(u) dH_0(t) \\ &\quad - \sum_{i=1}^n \int_0^\tau \frac{R_i}{\pi(Q_i)} \{Z_i - \mu_z(t, \cdot)\} Y_i(t) \lambda\{\beta_0^T Z_i + H_0(t)\} / \lambda^*\{H_0(t)\} \\ &\quad \left. \times 1/n \sum_{i=1}^n \frac{\lambda^*\{H_0(t)\}}{B_2(t)} \frac{R_i \pi_\psi^T(Q_i)}{\pi^2(Q_i)} dM_i(t) \right]. \end{aligned}$$

Appendix B

Proof of Theorems in §2.6

Here we prove the theorems in 5 steps. We first derive the influence function for the single imputation estimator for a typical completed data set (the jth completed data set), in steps B1, B2, and B3. Write the left hand side of (2.10) as function $U\{O, \beta, H, \hat{\gamma}_n\}$. In step B1, The gradient matrix of $U\{O, \beta, \hat{H}_j, \hat{\gamma}_n\}$ with respect to β is proved to converge in probability to $-\Sigma_*$ that was defined in §2.6. To finish the derivation for the influence function, a series of expansions are presented in steps B2 and B3. The influence function then leads to the proof of Theorem 1 in step B4, and later the proof of Theorem 2 in step B5.

Step B1. Here we show that

$$\frac{1}{n}\frac{\partial}{\partial\beta}U\{O, \beta, \hat{H}_j(\cdot, \beta, \hat{\gamma}_n), \hat{\gamma}_n\}|_{\beta=\beta_0} \xrightarrow{P} -\Sigma_*.$$

Let $a > 0$ and b some fixed numbers that ensure finite integrals. We also need to define the following quantities, which were used in Chen et al. (2002):

$$\lambda^*\{H_0(t)\} = B(t, a), \qquad B_1(t) = \int_a^t E[Y(s)\dot{\lambda}\{\beta^T Z + H_0(s)\}]dH_0(s),$$

$$B_2(t) = E[Y(t)\lambda\{\beta^T Z + H_0(t)\}], \qquad \Lambda^*(t) = \int_b^t \lambda^*(s)ds.$$

By mimicking Steps B2 and B3 in Chen et al. (2002), we have

$$\Lambda^*\{\hat{H}_0(t)\} - \Lambda^*\{H_0(t)\} = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)}dM_i(s) + o_p(n^{-1/2}), \qquad \text{(B.1)}$$

$$
\begin{aligned}
& \frac{\partial \hat{H}_j(t,\beta,\hat{\gamma}_n)}{\partial \beta}|_{\beta=\beta_0} \\
= \quad & -\int_0^t \frac{E[Z\dot{\lambda}\{\beta_0^T Z_i + H_0(s)\}Y(s)]}{\lambda^*\{H_0(t)\}} \frac{\lambda^*\{H_0(s)\}}{B_2(s)} dH_0(s) + o_p(1), \qquad \text{(B.2)}
\end{aligned}
$$

$$
\begin{aligned}
& \frac{\partial \hat{H}_j(t,\beta,\hat{\gamma}_n)}{\partial \gamma}|_{\gamma=\gamma_0} \\
= \quad & \int_0^t \frac{\lambda^*\{H_0(s)\}}{\lambda^*\{H_0(t)\}B_2(s)} E\{\pi_\gamma^T(W,\gamma_0)\text{pr}(R=0|W)dI(T \le s)\} \\
& + o_p(1). \qquad \text{(B.3)}
\end{aligned}
$$

By (B.2) and by the law of large numbers,

$$
\begin{aligned}
& \frac{1}{n}\frac{\partial U\{O,\beta,\hat{H}_j(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\}}{\partial \beta}|_{\beta=\beta_0} \\
= \quad & -\frac{1}{n}\sum_{i=1}^n \frac{\partial \int_0^\tau Z_i Y_i(t)\lambda\{\beta^T Z_i + \hat{H}_j(t,\beta,\hat{\gamma}_n)\}d\hat{H}_j(t,\beta,\hat{\gamma}_n)}{\partial \beta}|_{\beta=\beta_0} \\
= \quad & -\frac{1}{n}\sum_{i=1}^n Z_i \lambda\{\beta_0^T Z_i + \hat{H}_j(T_i,\beta,\hat{\gamma}_n)\}\{Z_i + \frac{\partial \hat{H}_j(T_i,\beta,\hat{\gamma}_n)}{\partial \beta}|_{\beta=\beta_0}\}Y_i(T_i) \\
= \quad & -\int_0^\tau E\{(Z - \frac{E[Z\lambda\{\beta_0^T Z + H_0(T)\}Y(s)/\lambda^*\{H_0(T)\}]}{B_2(s)/\lambda^*\{H_0(s)\}}) \\
& \times [Z^T \dot{\lambda}\{\beta_0^T Z + H_0(s)\}Y(s)]\}dH_0(s) + o_p(1) \\
= \quad & -\Sigma_* + o_p(1),
\end{aligned}
$$

Here we expand $U\{O,\beta_0,\hat{H}_j(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\}$ at $\hat{H}_n$ about H_0

$$
\begin{aligned}
& U\{O,\beta_0,\hat{H}_j(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\} \\
= \quad & \sum_{i=1}^n \int_0^\tau Z_i dM_{ij}\{t,Z_i,\beta_0,\hat{H}_j(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\} \\
= \quad & \sum_{i=1}^n \int_0^\tau Z_i dM_{ij}\{t,Z_i,\beta_0,H_0(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\} \\
& -\sum_{i=1}^n Z_i \left[\Lambda\{\beta_0^T Z_i + \hat{H}_j(T_i,\beta_0,\hat{\gamma}_n)\} - \Lambda\{\beta_0^T Z_i + H_0(T_i,\beta_0,\hat{\gamma}_n)\}\right] \\
= \quad & \sum_{i=1}^n \int_0^\tau Z_i dM_{ij}(t,Z_i,\beta_0,H_0(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n)
\end{aligned}
$$

$$
\begin{aligned}
&-\sum_{i=1}^{n} Z_i \frac{\lambda\{\beta_0^T Z_i + H_0(T_i,\beta_0,\hat{\gamma}_n)\}}{\lambda^*\{H_0(T_i,\beta_0,\hat{\gamma}_n)\}} \times \\
&\times[\Lambda^*\{\hat{H}_j(T_i,\beta_0,\hat{\gamma}_n)\} - \Lambda^*\{H_0(T_i,\beta_0,\hat{\gamma}_n)\}] + o_p(n^{1/2}) \\
= \; & \sum_{i=1}^{n}\int_0^{\tau} Z_i dM_{ij}(t, Z_i, \beta_0, H_0(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n) \\
&-\sum_{i=1}^{n} Z_i \frac{\lambda\{\beta_0^T Z_i + H_0(T_i,\beta_0,\hat{\gamma}_n)\}}{\lambda^*(H_0(T_i,\beta_0,\hat{\gamma}_n))}\left[n^{-1}\sum_{j=1}^{n}\int_0^{T_i}\frac{\lambda^*\{H_0(u,\beta_0,\hat{\gamma}_n)\}}{B_2(u)}\times\right. \\
&\times dM_{ij}\{u, Z_i, \beta_0, H_0(.,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\}] + o_p(n^{1/2}) \\
= \; & \sum_{i=1}^{n}\int_0^{\tau} Z_i dM_{ij}\{t, Z_i, \beta_0, H_0(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\} \\
&-\sum_{j=1}^{n}\int_0^{\tau} E\left[\frac{Z\lambda\{\beta_0^T Z_i + H_0(T,\beta_0,\hat{\gamma}_n)\}Y(t)}{\lambda^*\{H_0(T,\beta_0,\hat{\gamma}_n)\}}\right]\frac{\lambda^*\{H_0(t,\beta_0,\hat{\gamma}_n)\}}{B_2(t)} \\
&\times dM_{ij}\{u, Z_i, \beta_0, H_0(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n)\} + o_p(n^{1/2}) \\
= \; & \sum_{i=1}^{n}\int_0^{\tau}[Z_i - \mu_z\{t,\beta_0,H_0(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\}] \times \\
& dM_{ij}\{t, Z_i, \beta_0, H_0(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\} + o_p(n^{1/2}) \\
= \; & U^*\{O,\beta_0,H_0(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\} + o_p(n^{1/2}),
\end{aligned}
$$

where the fifth equality follows from Chen et al.(2002).

Step B3. Here we expand $U^*\{O,\beta_0,H_0(\cdot,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\}$ at $\hat{\gamma}_n$ about γ_0, as $\rho(W,\gamma)$ is correctly specified. Define function

$$
\varphi_{ij} \equiv \int_0^{\tau}[Z_i - \mu_z\{t,\beta_0,H_0(t,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\}]dM_{ij}\{t,Z_i,\beta_0,H_0(t,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\}.
$$

Therefore $\sum_{i=1}^{n}\varphi_{ij} \equiv U^*$. Let $\mu_\varphi = E\{\varphi\}$. Notice $\mu_\varphi(\gamma_0) = 0$. Also define $G = \sum_{i=1}^{n}\{\varphi_{ij} - \mu_\varphi\}$, then it can be shown using the theory of empirical processes (van der varrt, 2000) that

$$
n^{-1/2}\{G(\beta_0,\hat{\gamma}_n) - G(\beta_0,\gamma_0)\} \xrightarrow{P} 0;
$$

hence

$$
n^{-1/2}U^*\{O,\beta_0,H_0(t,\beta_0,\hat{\gamma}_n),\hat{\gamma}_n\}
$$

$$\begin{aligned}
&\equiv n^{-1/2}\sum_{i=1}^{n}\varphi_{ij}(O,\beta_0,H_0(\cdot,\beta_0,\hat\gamma_n),\hat\gamma_n)\\
&= n^{-1/2}\sum_{i=1}^{n}\varphi_{ij}(O,\beta_0,H_0(t),\gamma_0)\\
&+ \left\{\frac{\partial\mu_\varphi(O,H_0(t),\gamma)}{\partial\gamma^T}\right\}|_{\gamma=\gamma_0}n^{1/2}(\hat\gamma_n-\gamma_0)+o_p(1)
\end{aligned}$$

First we need to compute μ_φ

$$\begin{aligned}
&\mu_\varphi(\beta_0,\gamma)\\
=\ & E\{\int\{Z-\mu_z(t;\gamma)\}dM(t;\gamma)\}\\
=\ & E[\int\{Z_1-\mu_z(t;\gamma)\}dN_{1j}(t,\gamma)\\
&-\int\{Z_1-\mu_z(t;\gamma)\}\lambda\{\beta_0^TZ_1+H_0(t,\beta_0,\gamma)\}Y_1(t)dH_0(t,\beta_0,\gamma)]\\
=\ & E[\{Z-\mu_z(T)\}\mathrm{pr}(R=0|W)\{\pi(W,\gamma)-\pi(W,\gamma_0)\}]\\
&-E[\int\{Z-\mu_z(t;\gamma)\}\lambda\{\beta_0^TZ+H_0(t,\beta_0,\gamma)\}Y(t)dH_0(t,\beta_0,\gamma)].
\end{aligned}$$

Therefore by the second equality, the partial derivative of $\mu_\varphi(\beta_0,\gamma)$ with respect to γ, evaluated at $\gamma=\gamma_0$, equals

$$\begin{aligned}
&\frac{\partial}{\partial\gamma}\mu_\varphi(\beta_0,\gamma)|_{\gamma=\gamma_0}\\
=\ & -E\{\int\frac{\partial}{\partial\gamma}\mu_z(t;\gamma)dM(t;\gamma_0)\}|_{\gamma=\gamma_0}\\
&+E[\int\{Z-\mu_z(t;\gamma_0)\}\frac{\partial}{\partial\gamma}dM(t;\gamma)]|_{\gamma=\gamma_0}.
\end{aligned}$$

Since the first piece is the expectation of a stochastic integral, it equals zero at the truth. Hence,

$$\begin{aligned}
&\frac{\partial}{\partial\gamma}\mu_\varphi(\beta_0,\gamma)|_{\gamma=\gamma_0}\\
=\ & E\{\int\{Z-\mu_z(t;\gamma_0)\}\frac{\partial}{\partial\gamma}dM(t;\gamma)\}|_{\gamma=\gamma_0}\\
=\ & E\{\int\{Z-\mu_z(t;\gamma_0)\}\frac{\partial}{\partial\gamma}dN(t;\gamma_0)\}|_{\gamma=\gamma_0}\\
&+E\left(\int\{Z-\mu_z(t;\gamma_0)\}\frac{\partial}{\partial\gamma}[\lambda\{\beta_0^TZ+H_0(t,\beta_0,\gamma)\}Y(t)dH_0(t,\beta_0,\gamma)]\right)|_{\gamma=\gamma_0}.
\end{aligned}$$

It is straightforward to show that the first term and the second term, respectively, equal $E[\{Z-\mu_z(t;\gamma)\}\text{pr}(R=0|W)\{\pi_\gamma^T(W,\gamma_0)\}] \triangleq C_1$, and

$$\begin{aligned}
&\frac{\partial}{\partial\gamma}E\left[\int (Z-\mu_z(t;\gamma))\lambda\{\beta_0^T Z + H_0(t,\beta_0,\gamma)\}Y(t)dH_0(t,\beta_0,\gamma)\right]|_{\gamma=\gamma_0} \\
= \;& E\left[\int \{Z-\mu_z(t;\gamma)\}Y(t)d\frac{\partial\Lambda}{\partial\gamma}\right] \\
= \;& E\left(\int \{Z-\mu_z(t;\gamma)\}Y(t)d[\lambda\{\beta_0^T Z + H_0(t,\beta_0,\gamma)\}\frac{\partial H_0(t)}{\partial\gamma}]\right) \\
= \;& E\left(\int \{Z-\mu_z(t;\gamma)\}Y(t)\left[\dot\lambda\{\beta_0^T Z + H_0(t)\}\frac{\partial H_0(t)}{\partial\gamma}dH_0(t)\right.\right. \\
& \left.\left. +\lambda\{\beta_0^T Z + H_0(t)\}d\frac{\partial H_0(t)}{\partial\gamma}\right]\right)|_{\gamma=\gamma_0} \\
\triangleq \;& C_2
\end{aligned}$$

Here C_2 depends on $\frac{\partial H_0}{\partial\gamma}|_{\gamma=\gamma_0}$ that can be approximated by (B.3). Now putting things together we have

$$\frac{\partial}{\partial\gamma}\mu_\varphi(\beta_0,\gamma)|_{\gamma=\gamma_0} = C_1 - C_2 = F,$$

where F was used in Theorem 1. So

$$\begin{aligned}
n^{-1/2}U^*\{O,\beta_0,\hat H_n(t,\beta_0,\hat\gamma_n),\hat\gamma_n\} &= n^{-1/2}\sum_{i=1}^n \varphi_{ij}\{O_i,\beta_0,H_0(t),\gamma_0\} \\
&+ n^{1/2}F(\hat\gamma_n-\gamma_0) + o_p(1).
\end{aligned}$$

When there is no missingness, $\text{pr}(R=0|W)=0$, hence by definition $C_1=0$, and by (B.3) $C_2=0$, consequently $F=0$. Since γ is estimated by a logistic regression model, we have

$$n^{1/2}(\hat\gamma_n-\gamma_0) = n^{-1/2}\phi(O_i,\gamma_0) + o_p(1),$$

where

$$\begin{aligned}
\phi(O_i,\gamma_0) &= J^{-1}(\gamma_0)\rho_\gamma^T(W_i,\gamma_0)I(R_i=1,\Delta_i>0) \\
&\times \left[\frac{D_i-\rho(W_i,\gamma_0)}{\rho(W_i,\gamma_0)\{1-\rho(W_i,\gamma_0)\}}\right],
\end{aligned}$$

where $J^{-1}(\gamma_0)$ is the inverse of the fisher's information matrix. Hence for $\hat{\beta}_j$, the single imputation estimator for the jth completed data set,

$$n^{1/2}(\hat{\beta}_j - \beta_0) \quad = \quad \Sigma_*^{-1} n^{-1/2} \sum_{i=1}^{n} \{\varphi_{ij}(\beta_0, \gamma_0) + F\phi(O_i, \gamma_0)\} + o_p(1).$$

Since this influence function is a sum of independent and identically distributed random variables whose means are zero, by central limit theorem, $n^{1/2}(\hat{\beta}_j - \beta_0)$ converges to a mean zero normal variable with asymptotic variance given by the variance of a single summand.

Step B.4 We now show that this variance is just V_{SI} defined in Theorem 1. Define $L_i = (R_i, I(\Delta_i > 0), W_i)$. Notice that at the truth

$$E\{D|L\} = E\{D|I(\Delta > 0), W\} = I(\Delta > 0)\rho(W, \gamma_0). \qquad \text{(B.4)}$$

Since $\varphi_{ij}(\beta_0, \gamma_0) = \int_0^\tau [Z_i - \mu_z\{t, \beta_0, H_0(t), \gamma_0\}] dM_{ij}\{t, Z_i, \beta_0, H_0(t)\}$, $E\{\varphi_{ij}(\beta_0, \gamma_0)\} = 0$ and $\mathrm{Var}\{\varphi_{ij}(\beta_0, \gamma_0)\} = \sum^*$ for all i. Also for the parametric model, since $E\{\phi(O_i, \gamma_0)\} = 0$ and $\mathrm{Var}\{\phi(O_i, \gamma_0)\} = J^{-1}(\gamma_0)$, by (B.4), the covariance of φ_{ij} and $\phi(O_i, \gamma_0)$ equals

$$\begin{aligned}
& E\{\varphi_{ij}(\beta_0, \gamma_0)\phi(O_i, \gamma_0)\} \\
= \; & E[E\{\varphi_{ij}(\beta_0, \gamma_0)\phi(O_i, \gamma_0)\}|R_i, I(\Delta_i > 0), W_i] \\
= \; & J^{-1}(\gamma_0) E\left(\frac{\rho_\gamma^T(W, \gamma_0) R I(\Delta > 0)}{\rho(W, \gamma_0)\{1 - \rho(W, \gamma_0)\}} \{Z - \mu_z(T)\} E[\{D - \rho(W, \gamma_0)\} D | L] \right) \\
& - J^{-1}(\gamma_0) E\left(\frac{\rho_\gamma^T(W, \gamma_0) R I(\Delta > 0)}{\rho(W, \gamma_0)\{1 - \rho(W, \gamma_0)\}} \times \right. \\
& \left. \times \int Y(t) d\Lambda\{\beta_0^T Z + H_0(t)\} E[\{D - \rho(W, \gamma_0)\} | L] \right) \\
= \; & J^{-1}(\gamma_0) E\left[\frac{\rho_\gamma^T(W, \gamma_0) R I^2(\Delta > 0)}{\rho(W, \gamma_0)\{1 - \rho(W, \gamma_0)\}} \{Z - \mu_z(T)\} \rho(W, \gamma_0)\{1 - \rho(W, \gamma_0)\} \right] \\
& - J^{-1}(\gamma_0) E\left[\frac{\rho_\gamma^T(W, \gamma_0) R I(\Delta > 0)}{\rho(W, \gamma_0)\{1 - \rho(W, \gamma_0)\}} \times \right. \\
& \left. \times \int Y(t) d\Lambda\{\beta_0^T Z + H_0(t)\} \rho(W, \gamma_0)\{1 - I(\Delta > 0)\} \right] \\
= \; & J^{-1}(\gamma_0) E\left[\rho_\gamma^T(W, \gamma_0) R I^2(\Delta > 0)\{Z - \mu_z(T)\} \right] -
\end{aligned}$$

$$
\begin{aligned}
& -J^{-1}(\gamma_0)E\left[\frac{\rho_\gamma^T(W,\gamma_0)RI(\Delta>0)I(\Delta=0)}{\{1-\rho(W,\gamma_0)\}}\int Y(t)d\Lambda\{\beta_0^T Z+H_0(t)\}\right] \\
= & \; J^{-1}(\gamma_0)E\left[\rho_\gamma^T(W,\gamma_0)RI(\Delta>0)\{Z-\mu_z(T)\}\right]-0 \\
= & \; J^{-1}(\gamma_0)E[\rho_\gamma^T(W_i,\gamma_0)\mathrm{pr}(R_i=1,\Delta_i>0|W_i)\{Z_i-\mu_z(T_i)\}].
\end{aligned}
$$

Putting these results together we have $\mathrm{Var}\{\varphi_{ij}(\beta_0,\gamma_0)+F\phi(O_i,\gamma_0)\}$ equals V_{SI}, which in turn proves Theorem 1.

Step B.5 We now prove Theorem 2. Since only D_{ij} is imputed, $j=1,\cdots,m$,

$$
n^{1/2}(\hat\beta-\beta_0) \;=\; \Sigma_*^{-1}n^{-1/2}\sum_{i=1}^n\left\{m^{-1}\sum_{j=1}^m\varphi_{ij}(\beta_0,\gamma_0)+F\phi(O_i,\gamma_0)\right\}+o_p(1),
$$

which is also a sum of independent and identically distributed mean zero random variables. It then follows from the central limit theorem that $n^{1/2}(\hat\beta-\beta_0)$ is asymptotically normal with mean zero and asymptotic variance given by the variance of a single summand. It is straight forward that for $j\neq j'$,

$$
\begin{aligned}
\mathrm{Var}\{\varphi_{ij},\varphi_{ij'}\} &= E\{\varphi_{ij}\varphi_{ij'}\} \\
&= E[\{Z_i-\mu_z(T_i)\}^{\bigotimes 2}D_{ij}D_{ij'}] \\
&= E[\{Z_i-\mu_z(T_i)\}^{\bigotimes 2}\mathrm{pr}(R_i=0|W_i)\rho(W_i,\gamma_0)\{1-\rho(W_i,\gamma_0)\}].
\end{aligned}
$$

Thus the variance of a single summand is just given by V_{MI}, which gives Theorem 2.

Appendix C

Semiparametric Theory

In this section we will study the theory for finding semiparametric estimators for parameters in semiparametric models with missing data. This theory was developed in Robins et al. (1994) and later described in great detail in the lecture notes written by Dr. Tsiatis for his course called 'Semiparametric Inference and Missing Data'. A general approach for finding the semiparametric estimators is as follows. First, we need to identify an influence function for the full data, in the absence of missingness. Then we can obtain the related observed data influence function incorporating the missingness, which is the residual of the full data influence function after projecting it onto a linear subspace of the observed data Hilbert space. This linear subspace consists of all the functions of the observed data with mean zero conditioned on the covariates Z. Finally the semiparametric estimators are the solution to the estimating equations that are formed by the observed data influence functions. Dr. Kaifeng Lu in his Ph.D thesis found the locally semiparametric efficient estimator for the proportional hazards model with missing cause of failure. We will see that it is in general very difficult to find the class of full data influence functions for the class of linear transformation models. However, a generalization of the partial likelihood scores for the proportional hazards model led Chen et al. (2002) to a set of estimating equations (1.5) and (1.6) that result in consistent and asymptotically normal estimators for linear transformation models with full data. Although their estimators are not semiparametric efficient, according to a technical report written by Dr. Tsiatis (2004), these estimators are very similar to the locally semiparametric efficient estimators in the sense that the efficiency of these two estimators are almost identical, at least in finite samples. In the presence of missing data, for (1.5) and (1.6), we used the techniques developed in Robins et al. (1994) to obtain estimating equations that lead to optimal estimators with minimum variances. These estimators are the DR estimators that were proposed in chapter 1.

C.1 The difficulties in finding full data influence functions

For the full data, the density for a typical observation can be written as

$$\begin{aligned} & P_{T,\Delta,Z,A}(t,\delta,z,a) \\ = \; & p_{A|T,\Delta,Z}(a|t,\delta,z) \\ & \times \exp[-\{\Lambda_2(t|z)+\Lambda_1(t|z)+\Lambda_0(t|z)\}] \\ & \times \lambda_2(t|z)^{I(\delta=2)}\lambda_1(t|z)^{I(\delta=1)}\lambda_0(t|z)^{I(\delta=0)} \\ & \times p_Z(z), \end{aligned}$$

where p_Z is the marginal density of the covariates Z, $\lambda_1(t|z)$ is the conditional cause-specific hazard function for the failures from the competing cause and $\lambda_0(t|z)$ is the conditional cause-specific hazard function for C given $Z=z$, $\{\Lambda_2(t|z), \Lambda_1(t|z), \Lambda_0(t|z)\}$ are the corresponding cumulative hazard functions, i.e. $\Lambda_k(t|z) = \int_0^t \lambda_k(u|z)du$, $k = 0, 1, 2$, and finally $p_{A|T,\Delta,Z}$ is the conditional density of A given $\{T, \Delta, Z\}$.

This representation yields the corresponding log-likelihood

$$\begin{aligned} \text{like}^F &= \log\ P_{T,\Delta,Z,A}(t,\delta,z,a) \\ &= \log\ p_{A|T,\Delta,Z}(a|t,\delta,z) \end{aligned}$$

$$
\begin{aligned}
&-\{\Lambda_2(t|z) + \Lambda_1(t|z) + \Lambda_0(t|z)\} \\
&+I(\delta = 2)\log\ \lambda_2(t|z) + I(\delta = 1)\log\ \lambda_1(t|z) + I(\delta = 0)\log\ \lambda_0(t|z) \\
&+\log\ p_Z(z), \qquad \text{(C.1)}
\end{aligned}
$$

For the linear transformation model (1.2),

$$
\begin{aligned}
\Lambda_2(t|z) &= -\log\ S_{T_2|Z}(t|Z) \\
&= -\log\ P(T_2 > t|Z) \\
&= -\log\ P\{H(T_2) + \beta^T Z \geq H(t) + \beta^T Z|Z\} \\
&= \Lambda\{H(t) + \beta^T Z\},
\end{aligned}
$$

where $S_{T_2|Z}(t|Z)$ is the conditional survival function of T_2 given Z and Λ is the cumulative hazard function for ε. Therefore

$$
\begin{aligned}
\lambda_2(t|Z) &= -\frac{d}{dt}\Lambda\{H(t) + \beta^T Z\} \\
&= \lambda\{H(t) + \beta^T Z\}h(t),
\end{aligned}
$$

where $h(s) = \frac{d}{ds}H(s)$ and $\lambda(s) = \frac{d}{ds}\Lambda(s)$.

Thus, for linear transformation models, (C.1) reduces to

$$
\begin{aligned}
\text{like}^F = &\log\ p_{A|T,\Delta,Z}(a|t, \delta, z) \\
&-[\Lambda\{H(t) + \beta^T z\} + \Lambda_1(t|z) + \Lambda_0(t|z)] \\
&+I(\delta = 2)\log\ \lambda\{H(t) + \beta^T z\}h(t) + I(\delta = 1)\log\ \lambda_1(t|z) + I(\delta = 0)\log\ \lambda_0(t|z) \\
&+\log\ p_Z(z). \qquad \text{(C.2)}
\end{aligned}
$$

Since the nuisance parameters $\{H(t), \lambda_1(t|z), \lambda_0(t|z), p_Z(z), p_{A|T,\Delta,Z}(a|t, \delta, z)\}$ are separate from one another in (C.2), the full data nuisance tangent space is a direct sum of five orthogonal spaces,

$$
\Lambda^F = \Lambda_{1s} \oplus \Lambda_{2s} \oplus \Lambda_{3s} \oplus \Lambda_{4s} \oplus \Lambda_{5s},
$$

where Λ_{1s} is the space associated with $H(t)$, Λ_{2s} corresponds to $\lambda_1(t|z)$, Λ_{3s} corresponds to $\lambda_0(t|z)$, Λ_{4s} is associated with p_Z and Λ_{5s} is associated with $p_{A|T,\Delta,Z}$. To find Λ_{1s}, we first consider parametric sub-models for $H(t)$, say $H(t,\theta) = \int_0^t h(s,\theta)du$, and $H(t) = H(t,\theta_0)$, where θ_0 denotes the truth. The score vector for θ is by taking the derivative of the log-likelihood with respect to θ, evaluated at the truth ($\theta = \theta_0$),

$$\begin{aligned} S_\theta &= \int \frac{\partial}{\partial\theta}\log[\lambda\{H(t,\theta) + \beta^T Z\}h(t,\theta)]dM(t)|_{\theta=\theta_0} \\ &= \int [\varsigma\{H(t) + \beta^T Z\}\frac{\partial}{\partial\theta}H(t,\theta) + \frac{\partial}{\partial\theta}\log\ h(t,\theta)]dM(t)|_{\theta=\theta_0}, \end{aligned} \tag{C.3}$$

where $\varsigma(t) = \frac{\dot\lambda(t)}{\lambda(t)}$. Both $\varsigma(t)$ and $\dot\lambda(t)$ are known functions. By (C.3), we postulate that

$$\Lambda_{1s} = \left(\int \left[\varsigma\{H(t) + \beta^T Z\}\int_0^t w(u)h(u)du + w(t)\right] dM(t) : \forall w(t)\right). \tag{C.4}$$

This conjecture is easily verified by first observing that all score vectors for θ derived in (C.3) are included in this space, and finally by considering the parametric sub-model of the form $h(t,\theta) = h(t)e^{\theta w(t)}$ for arbitrary function $w(t)$.

On the other hand, if we put no restrictions on the form of the cause-specific hazard $\lambda_2(t|z)$, the log-likelihood (C.1) is just a saturated (nonparametric) model, and the nuisance tangent space is the entire full data Hilbert space that can be written as the direct sum of the five orthogonal spaces,

$$\mathcal{H}^F = \Lambda_{1s}^* \oplus \Lambda_{2s} \oplus \Lambda_{3s} \oplus \Lambda_{4s} \oplus \Lambda_{5s},$$

where Λ_{1s}^* corresponds to $\lambda_2(t|z)$. It is well known, also easy to show, that

$$\Lambda_{1s}^* = \left\{\int a(t,Z)dM(t) : \forall a(t,Z)\right\}.$$

Thus the space that is orthogonal to the full data nuisance tangent space consists of elements of Λ_{1s}^* perpendicular to Λ_{1s}, i.e. allowing abuse of notation,

$$\Lambda^{F\perp} = \Lambda_{1s}^* - \Pi(\Lambda_{1s}^*|\Lambda_{1s}), \tag{C.5}$$

where $\Pi(S_1|S_2)$ means the projection of S_1 onto S_2. All full data influence function for linear transformation models belong to $\Lambda^{F\perp}$. Dr. Tsiatis in his lecture notes gave general theory for finding $\Lambda^{F\perp}$ for proportional hazards models. However, due to the complex structure of Λ_{1s}, it is in general very difficult to compute $\Lambda^{F\perp}$, hence it is almost impossible to identify all of the full data influence function for linear transformation models. Interested readers are refereed to the technical report (Tsiatis, 2004). Consequently, the usual approach for finding estimating equations for semiparametric models can not be applied to general linear transformation models. Nonetheless, the partial likelihood scores for the proportional hazards model, a special case of linear transformation models, motivated Chen et al. (2002) to consider the estimating equations (1.5) and (1.6) for (β, H) jointly for arbitrary linear transformation models, as (1.5) and (1.6) reduce to the Cox Partial likelihood score equation when the proportional hazards model is considered.

C.2 Estimating equations in the observed world

In the presence of missing cause of failure, if (1.5) and (1.6) are considered, according to the theory developed in section 2.4 of Robins et al. (1994), in the observed world the estimators of the truth (β_0, H_0) are the solutions to the following class of estimating equations for (β, H),

$$\sum_{i=1}^{n} \int_0^{\infty} \left[\frac{R_i}{\pi(Q_i, \hat{\psi}_n)} Z_i \left[dN_i(t) - Y_i(t) d\Lambda\{\beta^T Z_i + H(t)\} \right] - \frac{R_i - \pi(Q_i, \hat{\psi}_n)}{\pi(Q_i, \hat{\psi}_n)} \{f(t, Q_i)\} \right] = 0, \tag{C.6}$$

$$\sum_{i=1}^{n} \left[\frac{R_i}{\pi(Q_i, \hat{\psi}_n)} \left[dN_i(t) - Y_i(t) d\Lambda\{\beta^T Z_i + H(t)\} \right] - \frac{R_i - \pi(Q_i, \hat{\psi}_n)}{\pi(Q_i, \hat{\psi}_n)} \{g(t, Q_i)\} \right] = 0, \tag{C.7}$$

where f and g are arbitrary functions. Clearly these equations reduce to the inverse probability weighted estimating equations (1.8) and (1.9), if $f(t, Q_i) \equiv g(t, Q_i) \equiv 0$ for all i and all $0 \leq t \leq \tau$. However, Proposition 2.3 of Robins et al. (1994) shows that for the fixed full data estimating equations (1.5) and (1.6), the optimal choice for the observed data estimating equations is by choosing $f(t, Q_i) = \mathrm{E}\{Z_i dM_i(t)|Q_i\}$ and $g(t, Q_i) = \mathrm{E}\{dM_i(t)|Q_i\}$ for all i and all $0 \leq t \leq \tau$, i.e.,

$$\begin{aligned} f(t, Q_i) &= \mathrm{E}\{Z_i dM_i(t)|Q_i\} \\ &= \mathrm{E}(Z_i[dN_i(t) - Y_i(t)d\Lambda\{H(t) + \beta^T Z_i\}]|Q_i) \\ &= Z_i[\mathrm{E}\{dN_i(t)|Q_i\} - Y_i(t)d\Lambda\{H(t) + \beta^T Z_i\}] \\ &= Z_i[\mathrm{E}\{I(\Delta_i = 2)dN_i^*(t)|Q_i\} - Y_i(t)d\Lambda\{H(t) + \beta^T Z_i\}] \\ &= Z_i[\mathrm{E}\{I(\Delta_i = 2)|Q_i\}dN_i^*(t) - Y_i(t)d\Lambda\{H(t) + \beta^T Z_i\}] \\ &= Z_i[\rho(W_i)dN_i^*(t) - Y_i(t)d\Lambda\{H(t) + \beta^T Z_i\}], \end{aligned}$$

and similarly

$$g(t, Q_i) = \rho(W_i)dN_i^*(t) - Y_i(t)d\Lambda\{H(t) + \beta^T Z_i\},$$

which leads to the augmented inverse probability weighted estimating equations (1.13) and (1.14) that lead to the doubly robust estimator for (β, H).