

Comparison of parameter estimators for cause-specific hazards in competing risks data with missing cause of failure

Jimin Lee¹ and Kedai Cheng²

¹Department of Mathematics and Statistics
University of North Carolina Asheville,
Asheville, NC 28804, USA
Email: jlee@unca.edu

²Department of Mathematics and Statistics
University of North Carolina Asheville,
Asheville, NC 28804, USA
Email: kcheng@unca.edu

ABSTRACT

We consider the semiparametric proportional hazards model for the cause-specific hazard function in analysis of competing risks data with missing cause of failure. The inverse probability weighting equation is well used for estimating the regression parameters under the missing at random (MAR), provided the model for the nonmissing cause case is correctly specified. In this study, we consider a parametric logistic regression, discriminant analysis, and neural network for the model of missingness probability. Simulation studies demonstrate that the inverse probability weighted (IPW) estimators by a correctly specified logistic regression model and by discriminant analysis perform well and are appropriate for practical use. The IPW estimator by neural network works considerably well for a sufficient large sample. The application of the proposed methods are illustrated with data from a bone marrow transplant study and Women's Interagency HIV study.

Keywords: Missing cause of failure, Inverse probability weighted estimator, Logistic regression, Neural network, Discriminant analysis.

Mathematics Subject Classification: 62D10, 62F12, 62N01, 62N02

1 Introduction

In medical studies, the responses to a treatment can be classified in terms of failure from disease of interest or from other causes. Hence, in the competing risks framework, each individual is exposed to K distinct types of risks and the eventual failure is classified to precisely one of the risks. Let T^* denote the time to failure, Δ^* the cause of failure, and Z a p -dimensional vector of possibly time-dependent covariates. Then, an estimable quantity in competing risks

data is the cause-specific hazard function of cause $k \in \{1, \dots, K\}$ defined by

$$\lambda_k^*(t|Z) = \lim_{h \rightarrow 0} \frac{1}{h} P(t \leq T^* < t+h, \Delta^* = k | T^* \geq t, Z),$$

which is the instantaneous rate of experiencing the event of type k at time t , having not experienced any of the K competing events until time t . Without loss of generality in our study, we consider only two causes of failure, the cause of interest as cause 1 and the other as cause 2 (i.e. $\Delta^* = 1$ or 2). In many applications involving follow-up studies, however, individuals may be subject to censoring. Let C be a censoring time and $T^* = \min(T_1^*, T_2^*)$, where T_1^* and T_2^* denote the latent failure times from causes 1 and 2, respectively. Then the observed data consist of observations of (T, Δ, Z) , where $T = \min(T^*, C)$ and $\Delta = \Delta^* I(T^* \leq C)$. If the failure time T^* is observed, Δ is the cause of failure and $\Delta = 0$ otherwise. The observable cause-specific hazard function of cause k in the presence of censoring is given by

$$\lambda_k(t|Z) = \lim_{h \rightarrow 0} \frac{1}{h} P(t \leq T < t+h, \Delta = k | T \geq t, Z), \quad k = 1, 2.$$

Throughout the paper, we assume that Z is an external covariate process (Kalbfleish and Prentice, 2002) and that the censoring time C is conditionally independent of (T^*, Δ^*) given Z . Under this assumption, it was shown that $\lambda_k^*(t|Z) = \lambda_k(t|Z)$ if the distribution of C is continuous at t (Lu and Tsiatis, 2001; Gao and Tsiatis, 2005; Lu and Liang, 2008).

A number of statistical models for the relationship between the cause-specific hazard function of interest and regression covariates have been studied, among others, by (Prentice et al., 1978; Benichou and Gail, 1990; Cheng et al., 1998; Shen and Cheng, 1999; Scheike and Zhang, 2003). In this article we study the proportional hazards model for describing the relationship

$$\lambda_1^*(t|Z) = \lambda_0(t) e^{\beta_0^T Z(t)}, \quad (1.1)$$

where $\lambda_0(\cdot)$ is a nonnegative but otherwise unspecified baseline hazard function and β_0 is a p -dimensional vector of regression parameters. The parameter β_0 can be consistently estimated by treating all the failure times with $\Delta \neq 1$ as censored observations and using the partial likelihood score equation proposed by (Cox, 1972; Cox, 1975). The estimator will be called the full-case estimator, denoted by $\hat{\beta}_F$ in the paper. In practice, however, the information needed for the cause of failure may be lost, or it may be difficult to determine the cause of death for some individuals (Andersen et al., 1996). When we have missing causes in data, a naive method for estimating the regression parameter β_0 is to simply ignore the missing data and use the partial likelihood score equation to the complete data only. The so-called complete-case analysis and its estimator, denoted by $\hat{\beta}_C$, is clearly inefficient and can lead to serious bias. Thus, analysis of competing risks data with missing cause of failure has received considerable attention and a number of methods have been proposed. (Lu and Tsiatis, 2001) proposed a parametric model to model the probability that the missing cause is the cause of interest while allowing the inclusion of additional auxiliary covariates and then estimated the regression parameters by using a multiple imputation method. (Gao and Tsiatis, 2005) considered linear transformation models and (Lu and Liang, 2008) considered the additive hazards model for the relationship in analysis of competing risks data with missing cause of failure.

In this study, we derive an estimator of the regression parameters in model (1.1), namely the inverse probability weighted estimator and establish its theoretical properties. The approach, following the idea of (Horvitz and Thompson, 2012), uses the inverse probability weighted complete-case technique to estimate the regression parameter. The fundamental idea of IPW is to weight observed data points where the cause is known by the inverse of the probability that their cause of failure would be observed. This approach uses only the complete cases and relies on the missing at random (MAR) assumption, and the model for the missing cause case must be correctly specified for the IPW estimator to be valid (Scharfstein et al., 1999; Lu and Tsiatis, 2001; Gao and Tsiatis, 2005; Lu and Liang, 2008). Thus, we consider several models for the probability of missing causes; logistic regression model, discriminant analysis, and neural network for the model.

The paper is structured as follows. In Section 2, the inverse probability weighted estimator is developed. The asymptotic properties of the corresponding estimator is established in Section 2. We review logistic regression, discriminant analysis, and neural network to estimate the probability of missing causes in Section 3. In Section 4, we investigate and compare the finite sample properties of the proposed estimators through simulations. A bone marrow transplant data set and Women's Interagency HIV study data set are analyzed in Section 4. Some conclusions and discussions are given in Section 5. Technical notations are detailed in Appendix.

2 Estimating equation and asymptotic properties

Since the cause of failure may not be observed for some individuals, we define the missingness indicator R as follows. If an individual's failure is observed, then $R = 1$ when the cause of failure information Δ is observed and $R = 0$ otherwise. If an individual is censored, we always define $R = 1$. We also introduce auxiliary covariates A which are not of interest for modeling the cause-specific hazard function but may be used to describe the missingness mechanism. The utilization of auxiliary information has been considered by (Lu and Tsiatis, 2001; Gao and Tsiatis, 2005; Lu and Liang, 2008; Gilbert et al., 2008), among others. Then the observed data will consist of

$$O_i = \{T_i, Z_i, A_i, \Delta_i, R_i\} \text{ if } R_i = 1,$$

and

$$O_i = \{T_i, Z_i, A_i, R_i\} \text{ if } R_i = 0$$

for $i = 1, \dots, n$. We assume that $\{O_i, i = 1, \dots, n\}$ are independent identically distributed. We also assume that the cause of failure is missing at random (Rubin, 1976); that is, the probability that the cause of failure is non-missing given $\Delta > 0$ and $W = (T, Z, A)$ depends only on the observed W , but not on the unobserved Δ ,

$$P(R = 1 | \Delta, \Delta > 0, W) = P(R = 1 | \Delta > 0, W). \quad (2.1)$$

The assumption implies that

$$P(R = 1 | \Delta > 0, W) = P(R = 1 | \Delta = 1, \Delta > 0, W) = P(R = 1 | \Delta = 2, \Delta > 0, W).$$

See (Lu and Tsiatis, 2001; Gao and Tsiatis, 2005; Lu and Liang, 2008).

2.1 Inverse probability weighted estimator

Following the inverse selection probability idea of (Horvitz and Thompson, 2012), the method of inversely weighting the probability of complete-case has been commonly used in missing data problems. To do that, we need to estimate the probability of a complete case, $\pi(Q) \equiv P(R = 1|Q)$, where $Q = (W, \Delta)$. By the MAR assumption and $R = 1$ when $\Delta = 0$, we have

$$\begin{aligned}\pi(Q) &= P(R = 1|W, \Delta) \\ &= P(R = 1|W, \Delta, \Delta > 0) \cdot I(\Delta > 0) + P(R = 1|W, \Delta, \Delta = 0) \cdot I(\Delta = 0) \\ &= P(R = 1|W, \Delta, \Delta > 0) \cdot I(\Delta > 0) + 1 \cdot I(\Delta = 0) \\ &= P(R = 1|W, \Delta > 0) \cdot I(\Delta > 0) + I(\Delta = 0) \\ &= r(W) \cdot I(\Delta > 0) + I(\Delta = 0),\end{aligned}\tag{2.2}$$

where $r(W) = P(R = 1|W, \Delta > 0)$. The probability of complete-case $r(W_i)$ may be specified by a parametric model $r(W_i, \psi_0)$, in terms of a few unknown parameters ψ_0 . Accordingly, let $\pi(Q_i, \psi_0) = r(W_i, \psi_0)I(\Delta_i > 0) + I(\Delta_i = 0)$. By (2.1) and (2.2), the likelihood L regarding to $\pi(Q, \psi_0)$ is

$$\begin{aligned}L(\pi) &= \prod_{i=1}^n \pi(Q_i)^{I(R_i=1)} (1 - \pi(Q_i))^{I(R_i=0)} \\ &= \prod_{i=1}^n \pi(Q_i)^{I(R_i=1)I(\Delta_i>0)} \cdot \pi(Q_i)^{I(R_i=1)I(\Delta_i=0)} \cdot (1 - \pi(Q_i))^{I(R_i=0)I(\Delta_i>0)} \\ &= \prod_{i=1}^n r(W_i)^{I(R_i=1)I(\Delta_i>0)} \cdot 1 \cdot (1 - r(W_i))^{I(R_i=0)I(\Delta_i>0)}, \\ &\quad \text{because } \pi(Q_i) = r(W_i) \text{ when } \Delta_i > 0, \text{ and } \pi(Q_i) = 1 \text{ when } \Delta_i = 0 \\ &= \prod_{i=1}^n r(W_i)^{R_i I(\Delta_i>0)} \cdot (1 - r(W_i))^{(1-R_i)I(\Delta_i>0)}.\end{aligned}$$

This implies that the maximum likelihood estimator $\hat{\psi}$ of ψ can be estimated by maximizing the likelihood based on only uncensored data

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

We define the counting process $N_i(t) = I(\Delta_i = 1)I(T_i \leq t)$ and at-risk process $Y_i(t) = I(T_i \geq t)$. Let $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, and $a^{\otimes 2} = aa^T$ for a vector a . Let

$$\begin{aligned}\tilde{S}^{(m)}(t, \beta, \psi) &= \frac{1}{n} \sum_{i=1}^n \frac{R_i}{\pi(Q_i, \psi)} Y_i(t) e^{\beta^T Z_i(t)} Z_i(t)^{\otimes m}, \quad m = 0, 1, 2, \\ \tilde{Z}(t, \beta, \psi) &= \frac{\tilde{S}^{(1)}(t, \beta, \psi)}{\tilde{S}^{(0)}(t, \beta, \psi)}, \\ \tilde{V}(t, \beta, \psi) &= \frac{\tilde{S}^{(2)}(t, \beta, \psi)}{\tilde{S}^{(0)}(t, \beta, \psi)} - \tilde{Z}(t, \beta, \psi)^{\otimes 2}.\end{aligned}$$

Then we consider the following inverse probability weighted estimating equation for β_0 ,

$$U_I(\beta, \hat{\psi}) = \sum_{i=1}^n \int_0^\tau \frac{R_i}{\pi(Q_i, \hat{\psi})} \left(Z_i(t) - \tilde{Z}(t, \beta, \hat{\psi}) \right) dN_i(t),\tag{2.3}$$

where $\tau > 0$ is the end of follow-up time. The inverse probability weighted estimator of β solves the above equation and is denoted by $\hat{\beta}_I$. When there is no missing cause, the equation (2.3) consequently becomes the partial likelihood score equation proposed by (Cox, 1972). The cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ can be estimated by

$$\hat{\Lambda}_{I0}(t) = \sum_{i=1}^n \int_0^t \frac{R_i}{\pi(Q_i, \hat{\psi})} \frac{dN_i(u)}{n\tilde{S}^{(0)}(u, \hat{\beta}_I, \hat{\psi})}.$$

2.2 Asymptotic results

Conditions C.

(C.1) $\lambda_0(t)$ is continuous on $[0, \tau]$. The distribution of C is continuous on $[0, \tau]$ and $P(C > \tau) > 0$. The covariate processes $Z_i(t)$ have paths that are left continuous and of bounded variation, and satisfy the moment condition $E[\|Z_i(t)\|^4 \exp(2M\|Z_i(t)\|)] < \infty$, where M is a constant such that $\beta \in [-M, M]^p$ and $\|A\| = \max_{k,l} |a_{kl}|$ for a matrix $A = (a_{kl})$.

(C.2) Each component of $s^{(j)}(t, \beta)$, $j = 0, 1, 2$ is continuous on $[0, \tau] \times [-M, M]^p$ for some $M > 0$, $s^{(0)}(t, \beta) > 0$ on $[0, \tau] \times [-M, M]^p$, and

$$\sup_{t \in [0, \tau], \beta \in [-M, M]^p} \|\tilde{S}^{(j)}(t, \beta) - s^{(j)}(t, \beta)\| = O_p(n^{-1/2}).$$

(C.3) The matrix $\Sigma = \int_0^\tau v(t, \beta_0) \lambda_0(t) s^{(0)}(t, \beta_0) dt$ is positive definite.

(C.4) There is a $\sigma > 0$ such that $r(W_i) \geq \sigma$ for all i with $\Delta_i > 0$. $r(W_i, \psi)$ is twice continuously differentiable with respect to ψ . There exists ψ^* satisfying the equation $ES_{\psi i}^* = 0$, where $S_{\psi i}^*$ is the corresponding score functions for $r(W_i, \psi)$ given in (5.2) in the Appendix. The information matrix I_ψ^* also given in (5.2) in the Appendix is positive definite.

We let ψ_0 be the true value of ψ such that $r(W_i) = r(W_i, \psi_0)$. In general, under Condition (C.4), there exists ψ^* such that $\hat{\psi}$ converges in probability to ψ^* and $\hat{\psi}$ consistently estimates ψ_0 (Haberman, 1974; Haberman, 1977; Gourieroux and Monfort, 1981; White, 1982). Thus, if the model for $r(W_i)$ is correctly specified, we have $\psi^* = \psi_0$ and $\hat{\psi}$ converges in probability to ψ_0 .

Theorem 2.1. Assume Conditions C are satisfied. If $r(W_i, \psi)$ is correctly specified for $r(W_i)$, then $\hat{\beta}_I$ converges in probability to β_0 and $\sqrt{n}(\hat{\beta}_I - \beta_0)$ converges in distribution to a zero-mean Gaussian random vector with covariance matrix $\Sigma^{-1} E[\omega_1 \omega_1^T] \Sigma^{-1}$, where

$$\Sigma = \int_0^\tau v(t, \beta_0) \lambda_0(t) s^{(0)}(t, \beta_0) dt,$$

$$\omega_i = \int_0^\tau \{Z_i(t) - \bar{z}(t, \beta_0)\} \frac{R_i}{\pi(Q_i, \psi_0)} dM_i(t) - V_\psi I_\psi^{-1} S_{\psi i},$$

$M_i(t) = N_i(t) - \int_0^t Y_i(u) e^{\beta_0^T Z_i(u)} d\Lambda_0(u)$, V_ψ is given in (5.1), I_ψ and $S_{\psi i}$ are given in (5.3) in the Appendix.

The asymptotic covariance matrix $\Sigma^{-1} E [\omega_1 \omega_1^T] \Sigma^{-1}$ can be consistently estimated by

$$\widehat{\Sigma}^{-1} \left(n^{-1} \sum_{i=1}^n \widehat{\omega}_i \widehat{\omega}_i^T \right) \widehat{\Sigma}^{-1},$$

where

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{R_i}{\pi(Q_i, \widehat{\psi})} \widetilde{V}(t, \widehat{\beta}_I, \widehat{\psi}) dN_i(t),$$

$$\widehat{\omega}_i = \int_0^\tau \frac{R_i}{\pi(Q_i, \widehat{\psi})} \{Z_i(t) - \widetilde{Z}(t, \widehat{\beta}_I, \widehat{\psi})\} d\widehat{M}_i(t) - \widehat{V}_\psi \widehat{I}_\psi^{-1} \widehat{S}_{\psi i},$$

and $\widehat{M}_i(t) = N_i(t) - \int_0^t Y_i(u) e^{\widehat{\beta}_I^T Z_i(u)} d\widehat{\Lambda}_{I0}(u)$. Here $\widehat{V}_\psi$, $\widehat{I}_\psi$ and $\widehat{S}_{\psi i}$ are obtained by replacing with their respective sample estimators and substituting $(\widehat{\beta}_I, \widehat{\psi})$ for (β_0, ψ_0) in V_ψ , I_ψ , and $S_{\psi i}$. See (Hyun et al., 2012) for the proof of Theorem 2.1.

3 Several models for $r(W)$

Logistic regression

Logistic regression is often used to model the probability that an experimental unit falls into one of the two groups based on predictor variables measured on the experimental unit. Since R is binary, one can posit a parametric logistic model, $\text{logit}\{r(W_i, \psi)\} = W_i^T \psi$ which was often used in this study area (Lu and Tsiatis, 2001; Gao and Tsiatis, 2005; Lu and Liang, 2008; Hyun et al., 2012), among others. However, model misspecification occurs, leading to biased parameter estimates and inaccurate inferences.

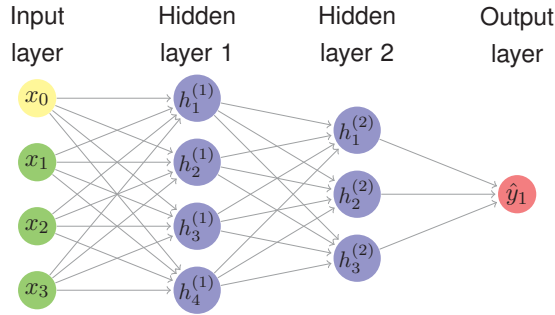
Discriminant analysis

Discriminant analysis is a method used to classify observations into nonoverlapping groups based on predictor variables. For example, a doctor might use discriminant analysis to identify patients at low or high risk for a disease based on factors like cholesterol levels, sex, blood pressure, and other variables. There are different types of discriminant analysis, including linear discriminant analysis which is a common approach that assumes a linear relationship between the predictor variables and the groups, and quadratic discriminant analysis that allows non-linear relationship between the predictor variables and the groups. Discriminant analysis relies on certain assumptions, including independence of observations, normality of predictor variables, and equal variance-covariance matrices across groups for linear discriminant analysis. The only difference from quadratic discriminant analysis is that we do not assume that the variance-covariance matrix is identical for different groups. We consider IPW estimator under linear and quadratic discriminant analysis for the probability of non-missingness, $r(W) = P(R = 1 | W, \Delta > 0)$.

Neural network

A (artificial) neural network is a computational model inspired by the structure of biological neural networks. A neural network typically consists of an input layer, one or more hidden layers, and an output layer. Each node in the hidden layer receives input, processes it, and send an output to other connected nodes. See Figure 1 for neural network diagram. Neural networks are being used in the area of prediction and classification of outcomes in medicine areas where regression models have traditionally been used. However, logistic regression assumes a linear relationship between input variables and the probability of the outcome, but neural networks can capture complex, non-linear relationship between inputs and the output. We consider IPW estimator under neural network model for the probability of non-missingness, $r(W) = P(R = 1|W, \Delta > 0)$.

Figure 1: A neural network is an interconnected group of nodes. Each circular node represents an artificial neuron and an arrow represents a connection from the output of one neuron to the input of another.



4 Numerical results

In this study, we estimate the parameter β_0 under these methods and evaluate the estimates based on Mean Squared Error (MSE) and Mean Absolute Error (MAE), where MSE calculates the average of the squared differences between IPW estimates and β_0 , and MAE is the average of the absolute differences between IPW estimates and β_0 .

4.1 Simulation studies

We present simulation studies conducted to evaluate the performance of the proposed methods. We consider a univariate covariate Z , where Z follows a normal distribution $N(0.5, 1/12)$. Given Z , the latent failure time T_1^* of interest is generated from the proportional hazards model $\lambda_1^*(t|Z) = \lambda e^{\beta_0 Z}$, where $\lambda = 1$ and $\beta_0 = -0.5$. The other latent failure time T_2^* is generated from a Gompertz distribution with a hazard function $\lambda_2^*(t|Z) = e^{\theta + \nu t}$, where $\theta = -0.5$ and $\nu = 0.2$. The censoring time C is generated from an exponential distribution which yields about 25% censoring level. We also consider a single auxiliary covariate A which follows the standard normal distribution $N(0, 1)$ or a Bernoulli distribution with success probability of 0.5. For missing

cause of failure, we consider a logistic regression model $\text{logit}\{r(W, \psi)\} = \psi_1 + \psi_2 T_b + \psi_3 Z + \psi_4 A$, where $T_b = \frac{T^\lambda - 1}{\lambda}$, the Box–Cox transformation with $\lambda = 0.25$ and A is the standard normal distribution (Model 1). We also consider $\text{logit}\{r(W, \psi)\} = \psi_1 + \psi_2 T + \psi_3 Z + \psi_4 A$, where A is the Bernoulli distribution (Model 2). With $\psi = (0.5, 0.5, 0.5, 0.5)$, we have about 32% missingness (Model 1) and about 17% missingness (Model 2). To study the performance of the estimators when $r(W)$ is misspecified, we posit two different parametric models of $r(W, \psi)$, where one is a correctly specified logistic model and the other is a misspecified model $\text{logit}\{r(W, \psi)\} = \psi_1 + \psi_4 A$. The simulation studies consist of 1000 runs with the sample size $n = 400$ and $n = 800$. Neural network consists of 3 nodes in one hidden layer. We consider IPW estimators from linear and quadratic discriminant analysis.

Model 1 has all normally distributed predictors which are required for discriminant analysis. The results from Tables 1 and 2 show that as expected, the full-case estimator $\hat{\beta}_F$ shows the smallest biases, SE, MSE, and MAE in all the settings, but the complete-case estimator $\hat{\beta}_C$ shows larger biases, SE, MSE, and MAE in almost all the settings. When the parametric logistic model for $r(W)$ is correctly specified, the IPW estimator $\hat{\beta}_{Ic}$ shows small biases, but the IPW estimator $\hat{\beta}_{Im}$ tends to have larger bias, MSE, and MAE when $\text{logit}\{r(W, \psi)\}$ is misspecified and $\hat{\beta}_{Im}$ is clearly sensitive to the misspecification. Correctly specifying a statistical model in real-world data analysis is incredibly difficult in most cases.

The IPW estimator under linear discriminant analysis, $\hat{\beta}_{IL}$ performs quite well and shows similar result as $\hat{\beta}_{Ic}$. Since neural networks generally require a large number of observations to achieve good performance, $\hat{\beta}_{IN}$ shows poor performance when $n = 400$ and as sample size increase its performance improves.

Model 2 has that only the auxiliary predictor A is normally distributed and the other two predictors are not normally distributed. Thus, it does not meet the normally condition for discriminant analysis. However, the results from Table 2 show that $\hat{\beta}_{IL}$ and $\hat{\beta}_{IQ}$ perform well and all the other estimators show similar results and patterns as in Table 1.

In conclusion, $\hat{\beta}_{IL}$ and $\hat{\beta}_{IQ}$ under discriminant analysis and the correctly specified IPW estimator $\hat{\beta}_{Ic}$ show similar performance and have better performance than the other estimators, except the full-case estimator.

4.2 Bone marrow transplant

(Sierra et al., 2002) described the characteristics and outcomes of 452 patients with primary myelodysplastic syndrome (MDS) who received transplants from HLA-identical siblings and were registered with the International Bone Marrow Transplant Registry (IBMTR). The study has two competing risks; treatment related death defined as death in complete remission and relapse defined as recurrence of myelodysplasia. In this example we consider 408 patients with complete covariate information obtained from the *timereg* R-package. Among these 408 patients, 161 patients died in complete remission, 87 patients relapsed, and 160 patients were censored. The covariates considered in our study are *age* of patient standardized at mean of 35 years old and *platelet* before transplantation (1 for more than 100×10^9 per L, or 0 for less). We apply the semiparametric proportional hazards model for the cause-specific hazard function to the data set with the two covariates: platelet and age. We also consider *tcell* (T-cell

depleted BMT; 1 for yes, 0 for no) for auxiliary variable. In the data set, the causes of failure are all known. We can obtain $\hat{\beta}_F$. For illustration purposes, we delete some failure causes by the three following missing mechanisms; missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR).

For the MCAR, the causes of failure are randomly selected to yield about 23% missing causes. For the MAR, the logistic model is chosen as $\text{logit}\{r(W)\} = 0.1 + 0.5T - 1.0age$ which yields about 23% missing causes, where T is the failure time. For the NMAR, the logistic model is chosen as $\text{logit}\{r(W)\} = 0.1 + 1.0T - 1.0age - 0.5I(\Delta = 1)$ which yields about 22% missing causes, where $\Delta = 1$ corresponds to the death in complete remission, $\Delta = 2$ to relapse, and $\Delta = 0$ does to being censored. Because correctly specifying a statistical model is often difficult, if not impossible, in real-world data analysis, we posit two logistic models for $r(W)$ with $\text{logit}\{r(W, \psi)\} = \psi_1 + \psi_2T_b + \psi_3platelet + \psi_4age + \psi_5tcell$ to obtain the estimators for β , where where $T_b = \frac{T^\lambda - 1}{\lambda}$ with $\lambda = 0.06$ (Model 1) and $\text{logit}\{r(W, \psi)\} = \psi_1 + \psi_2T + \psi_3platelet + \psi_4age + \psi_5tcell$ (Model 2). The results of the estimation of β are summarized in Tables 3 and 4. For comparison they also includes the full-case estimator, $\hat{\beta}_F$, based on the original data without artificial missing. The results are close under all the missing mechanisms and almost all estimators are closer to the full-case estimator than the IPW estimator from misspecified logistic regression. The analyses are consistent with the findings from the earlier study that patients with high platelet counts have a lower risk of treatment related mortality than those with low platelet counts and a higher risk rate is seen among the older patients.

4.3 Women's Interagency HIV Study

(Lau et al., 2009) analyzed the Women's Interagency HIV Study (WIHS) data. The WIHS was established in August 1993 to investigate the impact of HIV infection on US women. In 1994–1995, 2,054 HIV-positive and 569 HIV-negative women were enrolled. The study sample consisted of 1,164 women enrolled in WIHS, who were infected with HIV, alive, and free of clinical AIDS on December 6, 1995. Women were followed until the first of the following: time to initiation of therapy prior to AIDS or death (679 women, cause=1), time to the occurrence of AIDS or death prior to initiation therapy (359 women, cause=2), or administrative censoring on September 28, 2006 (126 women, cause=0). The data set also includes history of injection drug use at WIHS enrollment (drug, 1=yes, 0=no), whether an individual was African American (black, 1=yes, 0=no), age on December 6, 1995, standardized at mean 36.4 years, and CD4 nadir prior to December 6, 1995 (CD4). We apply the proportional hazards model for the cause-specific hazard function to the data set with the three covariates: the history of injection drug use, age, African American, and CD4 as auxiliary variable.

Because the cause of failure are all known, we delete some causes of failure by missing at random (MAR) mechanism by $\text{logit}\{r(W)\} = 0.1 + 0.5T + 0.5drug - 1.5age + 0.5black$ which yields about 20% missing causes. However, we posit two logistic models for $r(W)$ with $\text{logit}\{r(W, \psi)\} = \psi_1 + \psi_2T_b + \psi_3drug + \psi_4age + \psi_5CD4$ to obtain the estimators for β , where where $T_b = \frac{T^\lambda - 1}{\lambda}$ with $\lambda = 0.06$ (Model 1) and $\text{logit}\{r(W, \psi)\} = \psi_1 + \psi_2T + \psi_3drug + \psi_4age + \psi_5CD4$ (Model 2).

The study found that women with an injection drug use history were less likely to initiate therapy

prior to progression to AIDS or death than those without a history of injection drug use and that older women were more likely to initiate therapy prior to AIDS or death than young women.

5 Conclusion

This study investigates the use of the inverse probability weighted estimator to analyze competing risks data with missing cause of failure, where the Cox proportional hazard model is utilized to examine the covariate effects on the cause-specific hazard function for the failure type of interest. The simulation study shows that the complete-case only estimator $\hat{\beta}_C$ leads to significant bias in all the settings, highlighting the necessity of a method to account for the missing data. We explore three different approaches for modeling $r(W)$. The core challenge of this method lies in accurately modeling the probability of non-missingness $r(W)$. The study finds that IPW estimator using a parametric logistic regression model for $r(W)$ performs very well and is appropriate for practical use only when the parametric model for $r(W)$ is correctly specified. This suggests that when the underlying relationship is accurately captured by a logistic model, this estimator is unbiased and effective choice. Similar to a logistic regression model, both linear and quadratic IPW estimators based on discriminant analysis for $r(W)$ perform well, indicating that discriminant analysis can be a strong alternative to logistic regression for modeling $r(W)$. On the other hand, neural network method offers greater flexibility as it don't require a specific parametric model assumption. Our analysis of the WIHS data shows that $\hat{\beta}_N$ performs exceptionally well and is close to the $\hat{\beta}_F$. However, this approach requires a sufficiently large sample to provide stable and reliable results. In summary, the choice of model for $r(W)$ is critical. While logistic regression and discriminant analysis are effective and practical when their assumptions hold, neural networks offer a more flexible and robust solution for large datasets.

References

- Andersen, J., Goetghebeur, E. and Ryan, L. 1996. Missing cause of death information in the analysis of survival data, *Statistics in Medicine* **15**: 2191–2201.
- Benichou, J. and Gail, M. 1990. Estimates of absolute cause-specific risk in cohort studies, *Biometrics* **46**: 813–826.
- Cheng, S., Fine, J. and Wei, L. 1998. Prediction of cumulative incidence function under the proportional hazards model, *Biometrics* **54**: 219–228.
- Cox, D. 1972. Regression models and life tables (with discussion), *Journal of the Royal Statistical Society Series B* **34**: 187–220.
- Cox, D. 1975. Partial likelihood, *Biometrika* **62**: 269–276.
- Gao, G. and Tsiatis, A. 2005. Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure, *Biometrika* **92**: 875–891.

- Gilbert, P., McKeague, I. and Sun, Y. 2008. The two-sample problem for failure rates depending on a continuous mark: An application to vaccine efficacy, *Biostatistics* **9**: 263–276.
- Gourieroux, C. and Monfort, A. 1981. Asymptotic properties of the maximum likelihood estimator in dichotomous logit models, *Journal of Econometrics* **17**: 83–97.
- Haberman, S. 1974. *The Analysis of Frequency Data*, University of Chicago Press, Chicago.
- Haberman, S. 1977. Maximum likelihood estimates in exponential response models, *Annals of Statistics* **5**: 815–841.
- Horvitz, D. and Thompson, D. 2012. A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association* **142**: 1767–1779.
- Hyun, S., Lee, J. and Sun, Y. 2012. Proportional hazards model for competing risks data with missing cause of failure, *Journal of Statistical Planning and Inference* **142**: 1767–1779.
- Kalbfleish, J. and Prentice, R. 2002. *The Statistical Analysis of Failure Time Data. 2nd edition*, Wiley, New York.
- Lau, B., Cole, S. and Gange, S. J. 2009. Competing risk regression models for epidemiologic data, *American Journal of Epidemiology* **170**: 244–256.
- Lu, K. and Tsiatis, A. 2001. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure, *Biometrics* **57**: 1191–1197.
- Lu, W. and Liang, Y. 2008. Analysis of competing risks data with missing cause of failure under additive hazards model, *Statistica Sinica* **18**: 219–234.
- Prentice, R., Kalbfleisch, J., Peterson, A., Flournoy, N., Farewell, V. and Breslow, N. 1978. The analysis of failure times in the presence of competing risks, *Biometrics* **34**: 541–554.
- Rubin, D. 1976. Inference and missing data, *Biometrika* **63**: 581–592.
- Scharfstein, D., Rotnitzky, A. and Robins, J. 1999. Adjusting for nonignorable drop-out using semiparametric nonresponse models: rejoinder, *Journal of the American Statistical Association* **94**: 1135–1146.
- Scheike, T. and Zhang, M. 2003. Extensions and applications of the cox-aalen survival model, *Biometrics* **59**: 1036–1045.
- Shen, Y. and Cheng, S. 1999. Confidence bands for cumulative incidence curves under the additive risk model, *Biometrics* **55**: 1093–1100.
- Sierra, J., Perez, W., Rozman, W., Carreras, C., Klein, J., Rizzo, J., Davies, J., Lazarus, S., Bredeson, C., Marks, D., Canals, C., Boogaerts, M., Goldman, J., Champlin, R., Keating, A., Weisdorf, D., deWitte, T. and Horowitz, M. 2002. Bone marrow transplantation from hla-identical siblings as treatment for myelodysplasia, *Blood* **100**: 1997–2004.
- White, H. 1982. Maximum likelihood estimation of misspecified models, *Econometrica* **50**: 1–25.

Appendix

Let $\dot{\pi}(\cdot, \psi) = \partial\pi(\cdot, \psi)/\partial\psi$, $\dot{r}(\cdot, \psi) = \partial r(\cdot, \psi)/\partial\psi$, $\ddot{r}(\cdot, \psi) = \partial\dot{r}(\cdot, \psi)/\partial\psi^T$, and let

$$\begin{aligned} s^{(m)}(t, \beta) &= E[Y_1(t)e^{\beta^T Z_1(t)} Z_1(t)^{\otimes m}] \\ \bar{z}(t, \beta) &= \frac{s^{(1)}(t, \beta)}{s^{(0)}(t, \beta)} \\ v(t, \beta) &= \frac{s^{(2)}(t, \beta)}{s^{(0)}(t, \beta)} - \bar{z}(t, \beta)^{\otimes 2} \\ h^{(1)}(t, \beta, \psi) &= E\left[Y_1(t)e^{\beta^T Z_1(t)} \frac{R_1 \dot{\pi}^T(Q_1, \psi)}{\pi^2(Q_1, \psi)}\right] \\ h^{(2)}(t, \beta, \psi) &= E\left[Y_1(t)e^{\beta^T Z_1(t)} Z_1(t) \frac{R_1 \dot{\pi}^T(Q_1, \psi)}{\pi^2(Q_1, \psi)}\right]. \end{aligned}$$

Let

$$\begin{aligned} V_\psi &= E\left[\int_0^\tau \left(Z_1(t) - \bar{z}(t, \beta_0)\right) \frac{R_1 \dot{\pi}^T(Q_1, \psi_0)}{\pi^2(Q_1, \psi_0)} dN_1(t) \right. \\ &\quad \left. - \left(\frac{h^{(2)}(t, \beta_0, \psi_0)}{\bar{S}^{(0)}(t, \beta_0, \psi_0)} - \frac{\tilde{S}^{(1)}(t, \beta_0, \psi_0) h^{(1)}(t, \beta_0, \psi_0)}{(\tilde{S}^{(0)}(t, \beta_0, \psi_0))^2}\right) \frac{R_1}{\pi(Q_1, \psi_0)} dN_1(t)\right]. \end{aligned} \quad (5.1)$$

Let $S_{\psi i}^*$ and I_ψ^* be the score vector and the information matrix under the parametric model $r(W_i, \psi)$, respectively. Specifically,

$$\begin{aligned} S_{\psi i}^* &= \frac{I(\Delta_i > 0)\{R_i - r(W_i, \psi^*)\}\dot{r}(W_i, \psi^*)}{r(W_i, \psi^*)\{1 - r(W_i, \psi^*)\}}, \\ I_\psi^* &= E\left[S_{\psi 1} S_{\psi 1}^T - \frac{I(\Delta_1 > 0)\{R_1 - r(W_1, \psi^*)\}\dot{r}(W_1, \psi^*)}{r(W_1, \psi^*)\{1 - r(W_1, \psi^*)\}}\right]. \end{aligned} \quad (5.2)$$

Similarly, let $S_{\psi i}$ and I_ψ be the score vector and the information matrix under the parametric model $r(W_i, \psi_0)$, respectively, i.e.,

$$\begin{aligned} S_{\psi i} &= \frac{I(\Delta_i > 0)\{R_i - r(W_i, \psi_0)\}\dot{r}(W_i, \psi_0)}{r(W_i, \psi_0)\{1 - r(W_i, \psi_0)\}}, \\ I_\psi &= E\left[S_{\psi 1} S_{\psi 1}^T - \frac{I(\Delta_1 > 0)\{R_1 - r(W_1, \psi_0)\}\dot{r}(W_1, \psi_0)}{r(W_1, \psi_0)\{1 - r(W_1, \psi_0)\}}\right]. \end{aligned} \quad (5.3)$$

Table 1: Summary statistics of simulation results from Model 1.

Est.	$n = 400$				$n = 800$			
	Bias	SSE	MSE	MAE	Bias	SSE	MSE	MAE
$\hat{\beta}_F$	-0.0017	0.2758	0.0760	0.2205	-0.0028	0.1882	0.0354	0.1492
$\hat{\beta}_C$	0.0622	0.3775	0.1463	0.3028	0.0554	0.2599	0.0706	0.2122
$\hat{\beta}_{Ic}$	0.0050	0.3636	0.1321	0.2844	-0.0041	0.2546	0.0648	0.2029
$\hat{\beta}_{Im}$	0.0819	0.3703	0.1437	0.2992	0.0665	0.2560	0.0699	0.2120
$\hat{\beta}_{IL}$	0.0057	0.3632	0.1319	0.2842	-0.0036	0.2545	0.0647	0.2028
$\hat{\beta}_{IQ}$	0.0214	0.3519	0.1242	0.2771	0.0145	0.2428	0.0591	0.1933
$\hat{\beta}_{IN}$	-0.0377	1.5891	2.5241	0.4007	-0.0046	0.2886	0.0832	0.2154

Bias, the mean of the estimates of $\beta - \beta_0$; SSE, the sample standard deviation of the estimates of β ; MSE, the mean squared error; MAE, the mean absolute error; $\hat{\beta}_F$, the full-case estimator; $\hat{\beta}_C$, the complete-case estimator; $\hat{\beta}_{Ic}$ and $\hat{\beta}_{Im}$, the IPW estimators under logistic regression; $\hat{\beta}_{IL}$ and $\hat{\beta}_{IQ}$, the IPW estimators under linear and quadratic discriminant analysis, respectively; $\hat{\beta}_{IN}$, the IPW estimator under neural network. Here, c denotes the correctly specified logistic model and m the misspecified logistic model for $\tau(\cdot)$, respectively.

Table 2: Summary statistics of simulation results from Model 2.

Est.	$n = 400$				$n = 800$			
	Bias	SSE	MSE	MAE	Bias	SSE	MSE	MAE
$\hat{\beta}_F$	0.0062	0.2905	0.0844	0.2315	0.0010	0.1975	0.0390	0.1576
$\hat{\beta}_C$	0.0371	0.3294	0.1097	0.2669	0.0268	0.2277	0.0525	0.1835
$\hat{\beta}_{Ic}$	0.0078	0.3168	0.1003	0.2540	0.0038	0.2200	0.0484	0.1744
$\hat{\beta}_{Im}$	0.0436	0.3234	0.1064	0.2616	0.0328	0.2237	0.0511	0.1809
$\hat{\beta}_{IL}$	0.0094	0.3172	0.1006	0.2543	0.0024	0.2201	0.0484	0.1745
$\hat{\beta}_{IQ}$	0.0066	0.3155	0.0995	0.2518	-0.0074	0.2133	0.0455	0.1681
$\hat{\beta}_{IN}$	0.0255	0.3629	0.1322	0.2696	-0.0006	0.2521	0.0635	0.1799

Bias, the mean of the estimates of $\beta - \beta_0$; SSE, the sample standard deviation of the estimates of β ; MSE, the mean squared error; MAE, the mean absolute error; $\hat{\beta}_F$, the full-case estimator; $\hat{\beta}_C$, the complete-case estimator; $\hat{\beta}_{Ic}$ and $\hat{\beta}_{Im}$, the IPW estimators under logistic regression; $\hat{\beta}_{IL}$ and $\hat{\beta}_{IQ}$, the IPW estimators under linear and quadratic discriminant analysis, respectively; $\hat{\beta}_{IN}$, the IPW estimator under neural network. Here, c denotes the correctly specified logistic model and m the misspecified logistic model for $\tau(\cdot)$, respectively.

Table 3: Estimation of the effects of platelet and age in the bone marrow transplant data, Model 1.

Missingness	Estimator	Platelet		Age	
		Est.	SE	Est.	SE
None	$\hat{\beta}_F$	-0.543	0.186	0.0035	0.377
				0.0035	0.087
MAR	$\hat{\beta}_I$	-0.401	0.225	0.0756	0.416
	$\hat{\beta}_{IL}$	-0.408	0.222	0.0660	0.395
	$\hat{\beta}_{IQ}$	-0.559	0.263	0.0333	0.378
	$\hat{\beta}_{IN}$	-0.338	0.291	0.2440	0.454
MCAR	$\hat{\beta}_I$	-0.793	0.213	0.0002	0.359
	$\hat{\beta}_{IL}$	-0.793	0.213	0.0002	0.359
	$\hat{\beta}_{IQ}$	-0.709	0.227	0.0018	0.355
	$\hat{\beta}_{IN}$	-0.806	0.222	0.0003	0.372
NMAR	$\hat{\beta}_I$	-0.635	0.234	0.0066	0.450
	$\hat{\beta}_{IL}$	-0.623	0.230	0.0067	0.411
	$\hat{\beta}_{IQ}$	-0.684	0.235	0.0036	0.371
	$\hat{\beta}_{IN}$	-0.758	0.233	0.0111	0.459

Est., the estimate; SE, the standard error estimate; p -value pertaining to testing no covariate effect; $\hat{\beta}_F$, the full-case estimator with no missing causes; $\hat{\beta}_I$, the IPW estimator under logistic regression; $\hat{\beta}_{IL}$ and $\hat{\beta}_{IQ}$, the IPW estimators under linear and quadratic discriminant analysis, respectively; $\hat{\beta}_{IN}$, the IPW estimator under neural network.

Table 4: Estimation of the effects of platelet and age in the bone marrow transplant data, Model 2.

Missingness	Estimator	Platelet			Age		
		Est.	SE	p-value	Est.	SE	p-value
None	$\hat{\beta}_F$	-0.543	0.186	0.0035	0.377	0.087	< 0.0001
	$\hat{\beta}_I$	-0.422	0.217	0.0522	0.393	0.107	0.0002
	$\hat{\beta}_{IL}$	-0.428	0.214	0.0405	0.372	0.103	0.0003
	$\hat{\beta}_{IQ}$	-0.558	0.287	0.0522	0.392	0.109	0.0003
MAR	$\hat{\beta}_{IN}$	-0.636	0.219	0.0038	0.375	0.097	0.0001
	$\hat{\beta}_I$	-0.775	0.209	0.0002	0.362	0.087	< 0.001
	$\hat{\beta}_{IL}$	-0.774	0.209	0.0002	0.362	0.088	< 0.001
	$\hat{\beta}_{IQ}$	-0.412	0.229	0.0725	0.443	0.144	0.0012
MCAR	$\hat{\beta}_{IN}$	-0.830	0.221	0.0002	0.366	0.086	< 0.001
	$\hat{\beta}_I$	-0.568	0.228	0.0128	0.392	0.128	0.0021
	$\hat{\beta}_{IL}$	-0.563	0.222	0.0113	0.330	0.109	0.0025
	$\hat{\beta}_{IQ}$	-0.735	0.230	0.0014	0.378	0.105	0.0003
NMAR	$\hat{\beta}_{IN}$	-0.683	0.235	0.0037	0.453	0.141	0.0013

Est., the estimate; SE, the standard error estimate; p-value pertaining to testing no covariate effect; $\hat{\beta}_F$, the full-case estimator with no missing causes; $\hat{\beta}_I$, the IPW estimator under logistic regression; $\hat{\beta}_{IL}$ and $\hat{\beta}_{IQ}$, the IPW estimators under linear and quadratic discriminant analysis, respectively; $\hat{\beta}_{IN}$, the IPW estimator under neural network.

Table 5: Estimation of the effects of drug use, age, and African American in the WHIV study.

Missingness	Estimator	Drug Use			Age			African American		
		Est.	SE	p-value	Est.	SE	p-value	Est.	SE	p-value
None	$\hat{\beta}_F$	-0.383	0.187	< 0.0001	0.132 0.038 (Model 1)	0.038	0.0005	-0.394	0.077	< 0.0001
	$\hat{\beta}_I$	-0.494	0.117	< 0.0001				-0.321	0.090	0.0004
MAR	$\hat{\beta}_{IL}$	-0.513	0.120	< 0.0001				-0.317	0.091	0.0005
	$\hat{\beta}_{IQ}$	-0.943	0.244	0.0001				-0.181	0.123	0.1418
	$\hat{\beta}_{IN}$	-0.472	0.108	< 0.0001				-0.320	0.089	0.0003
MAR	$\hat{\beta}_I$	-0.511	0.117	< 0.0001	0.165 0.069 (Model 2)	0.069	0.0173	-0.308	0.090	0.0006
	$\hat{\beta}_{IL}$	-0.511	0.121	< 0.0001				-0.309	0.090	0.0006
	$\hat{\beta}_{IQ}$	-0.960	0.201	< 0.0001				-0.102	0.115	0.3745
	$\hat{\beta}_{IN}$	-0.490	0.107	< 0.0001				-0.320	0.088	0.0003

Est., the estimate; SE, the standard error estimate; p-value pertaining to testing no covariate effect; $\hat{\beta}_F$, the full-case estimator with no missing causes; $\hat{\beta}_I$, the IPW estimator under logistic regression; $\hat{\beta}_{IL}$ and $\hat{\beta}_{IQ}$, the IPW estimators under linear and quadratic discriminant analysis, respectively; $\hat{\beta}_{IN}$, the IPW estimator under neural network.