

ABSTRACT

LU, XIAOMIN. Improving the Efficiency of Tests and Estimators of Treatment Effect with Auxiliary Covariates in the Presence of Censoring . (Under the direction of Professor Anastasios A. Tsiatis).

In most randomized clinical trials, the primary response variable, for example, the survival time, is not observed directly after the patients enroll in the study but rather observed after some period of time (lag time). It is often the case that such a response variable is missing for some patients because of censoring such as administrative censoring that occurs when the study ends before all the patients had the opportunity to observe their response but also censoring may result from patient dropout. It is often assumed that censoring occurs at random which is referred to as noninformative censoring; however, in many cases such an assumption may not be reasonable. If the missing data are not analyzed properly, the estimate or test for the treatment effect may be biased. In this paper, we consider two situations. In the first situation, we only consider the special case where the censoring time is noninformative and the survival time itself is the time-lagged response. We use semiparametric theory to derive a class of consistent and asymptotically normal estimators for the unconditional log-hazard ratio parameter. The prognostic auxiliary covariates are used to derive estimators that are more efficient than the traditional maximum partial likelihood estimator and the corresponding Wald tests are more powerful than the logrank test. In the second situation, we extended the results under the first situation to a general case where the censoring time can be informative and the time-lagged response can be any type. We also use the semiparametric theory to derive a class of consistent and asymptotic normal estimator for the treatment effect estimator. The prognostic baseline auxiliary covariates and post-treatment auxiliary covariates, which may be time-dependent, are also used to derive estimators that both account for informative censoring and are more efficient than the estimators which do not consider the auxiliary covariates.

KEY WORDS: Informative censoring; Influence function; Logrank test; Nuisance tangent space; Proportional hazard model; Regular and asymptotically linear estimators.

**Improving the Efficiency of Tests and Estimators of Treatment Effect with
Auxiliary Covariates in the Presence of Censoring**

by

Xiaomin Lu

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, NC

May, 2007

Approved By:

Dr. Wenbin Lu

Dr. Hao Zhang

Dr. Anastasios A. Tsiatis
Chair of Advisory Committee

Dr. Marie Davidian

Dedication

To
My Mother, Lamei
My Husband, Heshan
and
My Brother, Jing

Biography

Xiaomin Lu was born in Wuhan, Hubei, China to parents Lamei Lu and Hongqing Chen on November 8, 1976. She received her B.S. in Applied Mathematics from South China University of Technology in June, 1999. In August, 2001, she entered the Ph.D program in Mathematics at Central Michigan University and received a M.A. in Mathematics in August, 2003. Thereafter, she joined the Ph.D program in Statistics at North Carolina State University. Upon completion of her doctoral degree, she will start working at the Department of Epidemiology and Biostatistics @ University of Florida as an assistant professor.

Acknowledgements

I would like to thank the members of my advisory committee for their advice and support during the preparation of my dissertation.

I am extremely grateful for having Dr. Anastasios A. Tsiatis as my academic advisor. His enthusiasm, knowledgeable guidance, and constant support, including financial support, have meant a tremendous amount to me. Besides being one of leading researchers in the field of biostatistics, he is also one of the most respectable people one could ever associate with. Without his great guidance and patience, it would have been much more difficult for me to complete the doctorate degree.

I would also like to thank Dr. Marie Davidian for her useful tips on LATEX and helpful advice on job opportunities, Dr. Bibhuti Bhattacharyya, Dr. Dennis Boos, Dr. Daowen Zhang and Dr. Cavell Brownie for imparting valuable knowledge and tremendous help in their lecturing courses. I am also grateful to Dr. Sastry G. Pantula, Dr. Bill Swallow, Dr. Pam Arroway, and Dr. Leonard A. Stefanski for their continuous support of my graduate study, Terry Byron for his aid on computing, Adrian Blue for his assistance toward graduation. My thanks also go to other faculty, staff and students in our department. The congenial and conducive learning environment they provide has made my study such a pleasant and fruitful experience.

Finally and foremost, I would like to thank my mother, my husband and my brother for their love and support all the way in my life.

Contents

List of Figures	vii
List of Tables	viii
1 Improving the efficiency of the logrank test using auxiliary covariates	1
1.1 Introduction	1
1.2 Model framework and notation	4
1.3 Major results and basic ideas to improve the efficiency	6
1.4 The class of all semiparametric estimators for β	7
1.5 Improving the efficiency of the logrank test	11
1.5.1 Deriving more efficient estimators for β than the maximum partial likelihood estimator	11
1.5.2 Variance estimator for $\hat{\beta}_{\text{AUG}}$	14
1.6 Simulation	15
1.7 A brief example	16
1.8 Concluding Remarks	19
2 Estimation of treatment effect with time-lagged response in the presence of informative censoring	21
2.1 Introduction	21
2.2 Model framework and notation	24
2.2.1 Notation and assumptions	24
2.2.2 Introduction to semiparametrics theory	25
2.3 The class of all semiparametric estimators for β	26
2.4 Application to the proportional hazard model	29
2.4.1 Class of all semiparametric estimators for β	29
2.4.2 Variance estimator for $\hat{\beta}$	32
2.5 Simulation	33
2.6 A brief example	37
3 Conclusion	40
Bibliography	42

Appendices	44
A.1 Sketch proof of Theorem 1.1	45
A.2 Sketch proof of Theorem 1.2	47
A.3 Sketch proof of Theorem 1.3	50
A.4 Proof of theorem 2.1	53
A.5 Proof of corollary 2.2	54
A.6 Proof of theorem 2.3	56

List of Figures

1.1	Log negative log survival function of time to death for each treatment	18
-----	--	----

List of Tables

1.1	Simulation Results for $\beta_0 = 0$ (2000 Monte-Carlo samples, entries in parentheses for SE and Monte-Carlo SE are relative efficiencies with respect to $\hat{\beta}_{PH}$.)	17
1.2	Simulation Results for $\beta_0 = 0.25$ (2000 Monte-Carlo samples, entries in parentheses for SE and Monte-Carlo SE are relative efficiencies with respect to $\hat{\beta}_{PH}$.) .	17
1.3	Estimates of $\hat{\beta}_{PH}$ and $\hat{\beta}_{AUG}$ on the ACTG 175 data (RE is the relative efficiencies with respect to $\hat{\beta}_{PH}$.)	19
2.1	Simulation Results for the 1st Scenario of Censoring (1000 Monte-Carlo samples, entries in parentheses for SE and Monte-Carlo SE are relative efficiencies with respect to $\hat{\beta}_1$.)	35
2.2	Simulation Results for the 2nd Scenario of Censoring (1000 Monte-Carlo samples, entries in parentheses for SE and Monte-Carlo SE are relative efficiencies with respect to $\hat{\beta}_1$.)	36
2.3	Simulation Results for the 3rd Scenario of Censoring (1000 Monte-Carlo samples, entries in parentheses for SE and Monte-Carlo SE are relative efficiencies with respect to $\hat{\beta}_1$.)	36
2.4	Estimates of $\hat{\beta}_{PH}$ and $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ on the ACTG 175 data (RE is the relative efficiencies with respect to $\hat{\beta}_{PH}$.)	38

Chapter 1

Improving the efficiency of the logrank test using auxiliary covariates

1.1 Introduction

A major objective in a randomized clinical trial of chronic disease, such as AIDS, Cancer, etc. is to compare the distribution of a time to event (for example, time to death, time to relapse of disease etc.) between two competing treatments, denoted here by treatments 1 and 0. The time to event, so-called survival time, may be right censored. The most commonly used method to test for treatment differences with censored survival data is the logrank test ([1], [2]). It is well known that under the assumption of proportional hazards, that is, the survival time T conditional on treatment assignment Z follows the proportional hazards model

$$\lambda_{T|Z}(t|z) = \lambda(t) \exp(\beta z), \quad (1.1)$$

where $\lambda_{T|Z}(t|z)$ denotes the conditional hazard rate of failing at time t given treatment $Z = z$ for $z = 0, 1$, respectively, the logrank test is optimal for testing for treatment difference, i.e., testing the null hypothesis of $H_0 : \beta = 0$.

The logrank test is also equivalent to the partial likelihood score test of Cox ([3]) for the proportional hazards regression model, and is asymptotically equivalent to the Wald test induced by $\hat{\beta}_{\text{PH}}$, the maximum partial likelihood estimator for the proportional hazard regression parameter. From the theory of semiparametrics, see for example [4] and [5], we know that the maximum partial likelihood estimator is the efficient estimator for β under the assumption of proportional hazards. However, these results are based on the premise that only information on time to event, which may be censored, and treatment assignment are used for the analysis.

In most clinical trials, in addition to the data on survival and censoring times and treatment assignment, auxiliary information (for example, age, gender, health conditions, etc.) are also collected and some of these variables may be important prognostic factors that are correlated with the time to event.

In randomized studies that test for, or estimate, treatment difference with censored survival data, there are two main sources of efficiency loss. Clearly, there is a loss of efficiency due to censoring of the time to event for some individuals in the study. However, there is also a loss of efficiency due to the impossibility of observing both the response to treatment 1 and the response to treatment 0 on the same individual. That is, when an individual is randomized to treatment 1 (or 0), then we do not observe his/her survival time on treatment 0 (or 1). Consequently, the idea proposed here is to use auxiliary covariates that correlate with survival time to recover as much of this lost information as is possible without making any additional assumptions other than those already necessary for the logrank test and the maximum partial likelihood estimator to hold.

This will be accomplished by using semiparametric theory to derive estimators and tests for the proportional hazards regression parameter when auxiliary covariates are introduced that are more efficient than the traditional maximum partial likelihood estimator and logrank test that do not consider such auxiliary information.

Remark 1.1. We wish to emphasize that the primary goal of this paper is to estimate the treatment-specific log hazard ratio β given by (1.1) and to test the null hypothesis that $\beta = 0$. Therefore, we will not be considering inference that is conditional on the auxiliary covariates.

For example, a popular model is the proportional hazards regression model which assumes

$$\lambda(t|Z, X) = \lambda(t) \exp(\beta_* Z + \beta_1^T X). \quad (1.2)$$

For such a model, the parameter β_* represents the treatment effect conditional on the covariates X . We first caution that post-baseline covariates should not be included as they will distort the causal interpretation of treatment effect. However, baseline covariates X may be included in such a model and the parameter β_* in (1.2) would denote the conditional treatment effect; i.e., the treatment log hazard ratio conditional on the covariates X under the assumption that this treatment log hazard ratio was constant for all values of X . Although such a conditional treatment effect may be of interest, the primary analysis for most randomized clinical trials is the unconditional treatment effect. Conditional treatment effects are generally estimated in a secondary analysis.

For nonlinear models such as Cox's proportional hazards model it is well known that the estimate for β_* obtained from standard inference will in general not be a consistent estimator for the parameter β in model (1.1). As we will also see later in §1.6, if the null hypothesis is true (i.e., $\beta = 0$ in (1.1)), then the corresponding test using the conditional proportional hazards regression model (1.2) will be biased (not control the type I error) and the resulting estimator, although unbiased, will in some cases be inefficient as compared to the unadjusted analysis.

This article is organized as follows. §1.2 describes the notation and model assumptions which will be used throughout this article. §1.3 gives the major results and basic ideas to improve the efficiency of the logrank test. §1.4 briefly reviews the theory of semiparametrics, and then uses the methodology of semiparametrics to derive the class of regular and asymptotic linear estimators for β . §1.5 characterizes a subclass of regular and asymptotic linear estimators for β which include the maximum partial likelihood estimator and derives the most efficient estimator for β within this subclass. Some easy-to-compute estimators which are more efficient than the maximum partial likelihood estimator are also provided in this section. To implement the theory, we perform a series of simulation studies in §1.6. A brief example is given in §1.7.

1.2 Model framework and notation

The data from a randomized clinical trial which compares the survival distribution between two treatments can be summarized as n realizations of independent and identically distributed random vectors $D_i = (U_i, \Delta_i, Z_i, X_i)$, $i = 1, \dots, n$. For the i -th individual, $U_i = \min(T_i, C_i)$, where T_i denotes the underlying survival time, and C_i denotes the potential censoring time, $\Delta_i = I(T_i \leq C_i)$ denotes the failure indicator, Z_i denotes the treatment indicator with value either 0 or 1 corresponding to treatment 0 or 1, and X_i denotes a vector of auxiliary covariates. Furthermore, we let $X_i = (X_{1i}^T, X_{2i}^T)^T$, where X_{1i} denotes the vector of baseline auxiliary covariates measured prior to randomization and X_{2i} denotes the vector of auxiliary covariates measured after randomization. The notation X^T denotes the transpose of a vector.

In randomized clinical trials, a common assumption that is often made, and in many cases is reasonable, is that of proportional hazards; that is, the survival time T conditional on Z follows the proportional hazards model (1.1). This model is particularly attractive because of its flexibility and robustness, i.e., the underlying hazard function $\lambda(t)$ is left unspecified and the parameter $\exp(\beta)$ measures the strength of the treatment effect as it corresponds to the hazard ratio (relative risk) between treatment 1 and treatment 0. Moreover, the logrank test, which is the most commonly used method to test for treatment differences with censored data, has greatest power to detect the nonparametric null hypothesis $H_0 : \lambda_{T|Z}(t|z = 1) = \lambda_{T|Z}(t|z = 0)$, $t \geq 0$, i.e., $\beta = 0$, against the proportional hazards alternative $H_A : \beta \neq 0$. With censored data, the parameter β is estimated by maximizing the partial likelihood of Cox ([3], [6]), which has been shown by Begun et al ([4]) to be the semiparametric efficient estimator for β . The maximum partial likelihood estimator for β is denoted by $\hat{\beta}_{\text{PH}}$. A key assumption, which is necessary for the maximum partial likelihood estimator of β to be a consistent, asymptotically normal estimator for β and for the logrank test to be asymptotically normal under the null hypothesis, is that $C \perp\!\!\!\perp T|Z$, that is, the potential censoring time is independent of the underlying survival time given treatment. This is also referred to as noninformative censoring conditional on treatment. For randomized clinical trials that were designed properly, this is a reasonable assumption, because, for such trials, the primary reason for censoring is administrative censoring where patients enter the study in a staggered fashion

and some are still alive at the end of the study when the data are analyzed.

Together with the survival and treatment data (U_i, Δ_i, Z_i) , additional auxiliary covariates X_i are also collected on the i -th individual, some of which may be correlated with the survival time, i.e., so-called prognostic factors. Because of randomization, it is reasonable to assume that the treatment indicator Z is independent of the auxiliary baseline covariates X_1 and that the randomization probability to treatment 1 is equal to π with $0 < \pi < 1$ which is known to us, that is,

$$Z \perp\!\!\!\perp X_1 \text{ and } Pr(Z = 1) = \pi. \quad (1.3)$$

We also assume that the censoring time C is independent of (T, X) conditional on Z , denoted by

$$C \perp\!\!\!\perp (T, X) | Z, \quad (1.4)$$

and denote the conditional survival distribution of censoring time given Z as $K_{C|Z}(u|Z)$, which is left unspecified. This assumption may be questionable in some situations especially if the belief is that patients that are prognostically worse or better, as measured through X , are more likely to drop out of the study. However, this is the assumption that is implicitly made in order for the properties of the logrank test and the maximum partial likelihood estimator for β to hold. As mentioned earlier, we need that $T \perp\!\!\!\perp C | Z$ in order for standard methods of analysis to hold. If the failure time T is related to X as well as Z (which is a property that is expected to hold and a property we will exploit to gain efficiency) and the censoring time C were also related to X as well as Z , then, conditional on Z only, T and C would no longer be conditionally independent. Informative censoring, conditional on Z , would be induced thus invalidating the standard methods of analysis. Therefore, in all that follows, assumption (1.4) will be made.

Without making any additional assumptions, other than those given by (1.1), (1.3) and (1.4), we now consider how to take advantage of the correlation of the auxiliary covariates with the survival time to obtain consistent, asymptotically normal estimators for β which are more efficient than the maximum partial likelihood estimator $\hat{\beta}_{\text{PH}}$. In so doing, we will be deriving a test, either a score test or Wald test, that will be more powerful than the logrank test to detect proportional hazards alternatives.

1.3 Major results and basic ideas to improve the efficiency

Without auxiliary covariates, it is well known (see [7], chapter 5) that all semiparametric estimators for β can be written as the solution to the estimating equation

$$\sum_{i=1}^n \int \{a(u, Z_i) - \bar{a}(u; \beta)\} dN_i(u) = 0, \quad (1.5)$$

where $\bar{a}(u; \beta) = \sum_{j=1}^n \{a(u, Z_j) \exp(\beta Z_j) Y_j(u)\} / \sum_{j=1}^n \{\exp(\beta Z_j) Y_j(u)\}$ for arbitrary functions $a(u, Z)$, $N(u) = I(U \leq u, \Delta = 1)$ is the counting process which counts the number of observed deaths up to and including time u and $Y(u) = I(U \geq u)$ is the “at risk” indicator at time u . In addition, by letting $a(u, Z) = Z$, the estimating equation (1.5) leads us to the partial likelihood score equation, whose solution is the maximum partial likelihood estimator for β , that is, the most efficient estimator within this class.

With auxiliary covariates and assumptions (1.1), (1.3) and (1.4), we will show in §1.4 that all semiparametric estimators for β can be represented as the solution to the following estimating equation

$$\sum_{i=1}^n \left[\int \{a(u, Z_i) - \bar{a}(u; \beta)\} dN_i(u) + (Z_i - \pi) f(X_{1i}) + \int \{g(u, Z_i, X_i) - \bar{g}(u, Z_i)\} dN_{C_i}(u) \right] = 0, \quad (1.6)$$

where $\bar{g}(u, Z_i) = \sum_{j=1}^n \{g(u, Z_j, X_j) Y_j(u) I(Z_j = Z_i)\} / \sum_{j=1}^n \{Y_j(u) I(Z_j = Z_i)\}$ for arbitrary functions $a(u, Z)$, $f(X_1)$ and $g(u, Z, X)$, and $N_C(u) = I(U \leq u, \Delta = 0)$ is the counting process which counts the number of observed censored observations up to and including time u . Both the second and the third summands of (1.6) have mean zero and hence the estimator obtained by solving the estimating equation (1.6) is an unbiased estimator for β . In addition, we immediately observe that choosing $a(u, Z) = Z$ and $f(X_1) = g(u, Z, X) = 0$ leads to the maximum partial likelihood estimator for β without auxiliary covariates. Consequently, the maximum partial likelihood estimator for β is included as a member of the class of estimators given by (1.6). This suggests that fixing $a(u, Z) = Z$ and judicious choices of $f(X_1)$ and $g(u, Z, X)$ will lead us to more efficient estimators.

Therefore, we restrict our attention to the subclass of estimating equations for fixed

$a(u, Z) = Z$ but arbitrary functions $f(X_1)$ and $g(u, Z, X)$. We show in §1.5 that the optimal functions $f(X_1)$ and $g(u, Z, X)$ which lead to the most efficient estimator for β within this subclass are $f_0(X_1) = E\{(Z - \pi)e_{a(u, Z)}(D; \beta_0)|X_1\}/\{\pi(1 - \pi)\}$ and $g_0(u, Z, X) = E\{e_{a(u, Z)}(D; \beta_0)|T \geq u, Z, X\}/K_C(u, Z)$, where

$$e_{a(u, Z)}(D; \beta) = \int \{a(u, Z) - a^*(u; \beta)\} dM(u, Z; \beta), \quad (1.7)$$

$a^*(u; \beta) = E\{a(u, Z) \exp(\beta Z)Y(u)\}/E\{\exp(\beta Z)Y(u)\}$, and $dM(u, Z; \beta)$ is the martingale increment $dN(u) - \lambda(u) \exp(\beta Z)Y(u)du$.

Although this result is of theoretical interest, the optimal functions $f_0(X_1)$ and $g_0(u, Z, X)$ involve the conditional expectations which are difficult, if not impossible, to estimate consistently. In §1.5, we also propose a simple but feasible method to estimate β . To be specific, we consider a specific class of estimating equations for fixed $a(u, Z) = Z$, but $f(X_1; \mathbf{a}) = \mathbf{a}^T q(X_1)$ and $g(u, Z, X; \mathbf{b}) = \mathbf{b}^T w(u, Z, X)$, where $\mathbf{a} \in \mathbb{R}^{r_a}$ and $\mathbf{b} \in \mathbb{R}^{r_b}$ are unknown parameters, $q(\cdot)$ is an r_a -dimensional vector of known functions of X_1 and $w(\cdot)$ is an r_b -dimensional vector of known functions of (u, Z, X) . Again, the maximum partial likelihood estimator is in this class (let $\mathbf{a} = \mathbf{b} = 0$). We will argue in §1.5 that the optimal values of \mathbf{a} and \mathbf{b} which result in the most efficient estimator within this specific class can be estimated by using standard regression methods. Moreover, if the conditional expectations $f_0(X_1)$ and $g_0(u, Z, X)$ are contained within this restricted class of parametric models, the proposed estimator will be efficient.

1.4 The class of all semiparametric estimators for β

In this section, we will use the theory of semiparametrics to derive the class of all regular and asymptotically linear estimators for β subject only to the restrictions given by assumptions (1.1), (1.3), and (1.4). We briefly introduce some necessary terminology and results used in semiparametric theory.

An estimator $\hat{\beta}_n$ for β is asymptotically linear if there exists a random variable $\varphi(D)$, which, under the truth, $\beta = \beta_0$, has mean zero and finite variance, such that $n^{1/2}(\hat{\beta}_n - \beta_0) =$

$n^{-1/2} \sum_{i=1}^n \varphi(D_i) + o_p(1)$. The function $\varphi(D_i)$ is referred to as the i -th *influence function* of the estimator $\hat{\beta}_n$. For a description of a “regular” estimator, please refer to [8]. The influence function of a regular and asymptotic linear estimator for β is uniquely defined and the asymptotic properties of such an estimator is determined by its influence function. It is clear from the definition of the influence function given above and a simple application of the central limit theorem, that, the asymptotic variance of an regular and asymptotic linear estimator $\hat{\beta}_n$ is equal to the variance of its influence function.

In semiparametric theory, influence functions of regular and asymptotic linear estimators can be viewed as geometric objects, i.e., vectors in a Hilbert space \mathcal{H} consisting of all mean zero finite variance random functions, $h(D)$, of a single observation D , equipped with the covariance inner product $\langle h_1, h_2 \rangle = E\{h_1(D)h_2(D)\}$ for any two elements $h_1, h_2 \in \mathcal{H}$. For such a Hilbert space, the square of the norm of an element $h \in \mathcal{H}$ is given by its variance, i.e., $\langle h, h \rangle = \text{Var}\{h(D)\}$. The key result that allows us to construct regular and asymptotic linear estimators for β is that influence functions of regular and asymptotic linear estimators for β must be orthogonal to the so-called *nuisance tangent space*; for more details see [7], chapter 4.

Consequently, to derive the class of regular and asymptotic linear estimators for β , we need to derive the nuisance tangent space Λ and its orthogonal complement space Λ^\perp .

The nuisance parameter η is defined as $(\eta_1, \eta_2, \eta_3, \eta_4)$, where (η_1, η_3, η_4) is used to define the conditional joint density of (T, X) given Z . Specifically, η_1 is used to denote the baseline hazard function $\lambda(t)$ for the proportional hazards model (1.1). The parameter $\eta_3 = (\eta_{31}, \eta_{32})$ is used to denote the conditional density of X given Z satisfying assumption (1.3), that is,

$$p_{X|Z}(x|z; \eta_3) = p_{X_2|(Z, X_1)}(x_2|z, x_1; \eta_{31})p_{X_1}(x_1; \eta_{32}), \quad (1.8)$$

where η_{31} is used to denote an arbitrary conditional density of X_2 given (Z, X_1) and η_{32} is used to denote an arbitrary marginal density of X_1 , which, by assumption (i.e., because of randomization), is independent of Z . The parameter η_4 is used to define the class of conditional densities of (T, X) given Z that satisfy the constraints imposed by (1.1) and (1.3), namely, that

$$\int p_{(T, X)|Z}(u, x|z; \beta_0, \eta_1, \eta_3, \eta_4) dx = \lambda(u; \eta_1) \exp(\beta_0 z) \exp\{-L(u; \eta_1) \exp(\beta_0 z)\}, \quad (1.9)$$

where $p_{(T,X)|Z}(\cdot)$ denotes the conditional density of (T, X) given Z , $L(u) = \int_0^u \lambda(v)dv$ is the cumulative baseline hazard function, and that

$$\int p_{(T,X)|Z}(u, x|z; \beta_0, \eta_1, \eta_3, \eta_4) du = p_{X|Z}(x|z; \eta_3). \quad (1.10)$$

The parameter η_2 is used to denote an arbitrary conditional density of C given Z , without any additional restrictions.

Theorem 1.1. *The nuisance tangent space Λ can be written as a sum of tangent spaces associated with each of the components η_1, \dots, η_4 making up η , that is,*

$$\Lambda = (\Lambda_1 + \Lambda_3 + \Lambda_4) \oplus \Lambda_2, \quad (1.11)$$

where the nuisance tangent space Λ_1 that is associated with the parameter η_1 satisfies the following properties:

$$\Lambda_1 \subset \left\{ a_1(U, \Delta, Z, X) : a_1 \in H_1 \right\}, \quad (1.12)$$

and

$$E\{\Lambda_1 | U, \Delta, Z\} = \Lambda_1^*, \quad (1.13)$$

where

$$H_1 = \left\{ \int a(u, Z, X) dM(u, Z, X) \text{ for all functions } a(u, Z, X) \right\},$$

$dM(u, Z, X)$ is the martingale increment $dN(u) - \lambda_{T|(Z,X)}(u|z, x)Y(u)du$, $\lambda_{T|(Z,X)}(\cdot)$ is the conditional hazard rate of T given (Z, X) , $\Lambda_1^* = \left\{ \int a(u) dM(u, Z) \text{ for all functions } a(u) \right\}$, $dM(u, Z)$ is the martingale increment $dN(u) - \lambda(u) \exp(\beta_0 Z) Y(u) du$ and $E\{\Lambda_1 | U, \Delta, Z\}$ is used as shorthand notation to denote the linear space consisting of elements $E\{h(U, \Delta, Z, X) | U, \Delta, Z\}$ for all $h(\cdot) \in \Lambda_1$.

The nuisance tangent space Λ_3 that is associated with the parameter η_3 satisfies

$$\Lambda_3 \subset \left\{ a_3(U, \Delta, Z, X) : a_3 \in H_1, E(a_3 | U, \Delta, Z) = 0 \right\}, \quad (1.14)$$

and

$$E\{\Lambda_3|Z, X\} = \Lambda_3^*, \quad (1.15)$$

in which $\Lambda_3^* = \{a_{31}(Z, X_1, X_2) + a_{32}(X_1) : E(a_{31}|Z, X_1) = 0 \text{ and } E(a_{32}) = 0\}$.

The nuisance tangent space Λ_4 that is associated with the parameter η_4 is given by

$$\Lambda_4 = \left\{ a_4(U, \Delta, Z, X) : a_4 \in H_1, \text{ and } E(a_4|U, \Delta, Z) = 0 \right\}, \quad (1.16)$$

and the nuisance tangent space Λ_2 that is associated with the parameter η_2 is given by

$$\Lambda_2 = \left\{ \int a_2(u, Z) dM_C(u, Z) \text{ for all functions } a_2(u, Z) \right\}, \quad (1.17)$$

in which $dM_C(u, Z)$ is the martingale increment $dN_C(u) - \lambda_C(u|Z)Y(u)du$.

The proof of Theorem 1.1 is given in Appendix A.1.

Remark 1.2. We wish to note that (1.12) and (1.13) do not uniquely characterize the space Λ_1 . Nonetheless, as we will illustrate, these conditions will suffice to define the orthogonal complement of the nuisance tangent space Λ^\perp . The situation is similar for Λ_3 .

Theorem 1.2. *The orthogonal complement of the nuisance tangent space Λ^\perp is given by*

$$\Lambda^\perp = \mathcal{E} + \{\mathcal{R} \oplus \mathcal{C}\} \quad (1.18)$$

where

$$\mathcal{E} = \left(\int \left[a(u, Z) - \frac{E\{a(u, Z) \exp(\beta_0 Z) Y(u)\}}{E\{\exp(\beta_0 Z) Y(u)\}} \right] dM(u, Z) \text{ for all functions } a(u, Z) \right), \quad (1.19)$$

$$\mathcal{R} = \left\{ (Z - \pi)f(X_1) \text{ for all functions } f(X_1) \right\}, \text{ and} \quad (1.20)$$

$$\mathcal{C} = \left(\int dM_C(u, Z) \left[g(u, Z, X) - E\{g(u, Z, X) | T \geq u, Z\} \right] \text{ for all functions } g(u, Z, X) \right). \quad (1.21)$$

The proof of Theorem 1.2 is given in Appendix A.2. We refer to the space \mathcal{E} as the *estimation space* because this is the space that contains all the regular and asymptotic linear

estimators for β if we didn't include any auxiliary covariates. The space \mathcal{R} and \mathcal{C} are two mutually orthogonal spaces that involve the auxiliary covariates and can be used to derive estimators that are more efficient. The space \mathcal{R} only uses the baseline (pre-randomization) covariates X_1 and the inclusion of this space allows us to gain back some of the efficiency lost due to the impossibility of observing both the response to treatment 1 and the response to treatment 0 on the same individual and is therefore referred to as the *randomization space*. The space \mathcal{C} uses all the auxiliary covariates and the inclusion of this space allows us to gain back some of the efficiency lost due to censoring and is therefore referred to as the *censoring space*.

Because a regular and asymptotic linear estimator must have an influence function that is orthogonal to the nuisance tangent space, this suggests that all regular and asymptotic linear estimators for β can be represented as the solution to the estimating equation (1.6).

Remark 1.3. The theoretical argument thus far implicitly assumes that the post-treatment covariates X_2 are available for all patients. For time-dependent covariates, say $X_2(t)$, where $X_2(t)$ may be potentially measured at times t_0, \dots, t_m , we clearly cannot observe these values beyond the time that a patient is at risk. Nonetheless, without loss of generality, by defining the function $g(u, Z, X)$ as $g(u, X_1, X_2(t_0)I(u \geq t_0), \dots, X_2(t_m)I(u \geq t_m), Z)$, a patient who is not at risk at time t_j will not have his/her covariates $X_2(t_j)$ be considered in the estimating equation (1.6) allowing the use of such time-dependent covariates.

1.5 Improving the efficiency of the logrank test

1.5.1 Deriving more efficient estimators for β than the maximum partial likelihood estimator

We begin by restricting our attention to the subclass of regular and asymptotic linear estimators for β which are the solution to the estimating equation (1.6) for a fixed function $a(u, Z)$ but arbitrary functions $f(X_1)$ and $g(u, Z, X)$. Denote such a subclass by $\mathcal{B}_{a(u, Z)}$. Clearly, each estimator in this subclass $\mathcal{B}_{a(u, Z)}$ has an influence function proportional to an

element in

$$\Lambda_{a(u,Z)}^\perp = \left(\int \{a(u, Z) - a^*(u; \beta_0)\} dM(u, Z) + (Z - \pi)f(X_1) + \int dM_C(u, Z) \left[g(u, Z, X) - E\{g(u, Z, X) | T \geq u, Z\} \right] : \right. \\ \left. \text{for all functions } f(X_1) \text{ and } g(u, Z, X) \right). \quad (1.22)$$

The proportionality constant C_a , according to the theory of semiparametrics ([7], chapter 4), is equal to the inverse of the expectation of the partial derivative of $e_{a(u,Z)}(D; \beta)$ with respect to β evaluated at the true value β_0 , that is, $C_a = [E\{\partial e_{a(u,Z)}(D; \beta_0)/\partial \beta\}]^{-1}$, where $e_{a(u,Z)}(D; \beta)$ is defined in (1.7). After some algebra, we derive the proportionality constant to be

$$\left(E[\Delta\{a(U, 1) - a^*(U; \beta_0)\}Z^*(U)] \right)^{-1}, \quad (1.23)$$

where $Z^*(u) = E\{Z \exp(\beta_0 Z)Y(u)\} / E\{\exp(\beta_0 Z)Y(u)\}$.

Theorem 1.3. *Consider the subclass of influence functions proportional to the elements in $\Lambda_{a(u,Z)}^\perp$. The efficient influence function for β within the subclass is proportional to*

$$\int \{a(u, Z) - a^*(u)\} dM(u, Z) - (Z - \pi)f_0(X_1) + \int \frac{dM_C(u, Z)}{K_C(u, Z)} \left[g_0(u, Z, X) - E\{g_0(u, Z, X) | T \geq u, Z\} \right], \quad (1.24)$$

where

$$f_0(X_1) = \frac{1}{\pi(1 - \pi)} E\left\{ (Z - \pi)e_{a(u,Z)}(D; \beta_0) | X_1 \right\}, \quad (1.25)$$

and

$$g_0(u, Z, X) = E\left\{ e_{a(u,Z)}(D; \beta_0) | T \geq u, Z, X \right\} \quad (1.26)$$

with $e_{a(u,Z)}(D; \beta_0)$ defined by (1.7).

The proof of Theorem 1.3 is given in Appendix A.3. This theorem shows that the optimal functions $f_0(X_1)$ and $g_0(u, Z, X)$ only depend on $a(u, Z)$ through the function $e_{a(u,Z)}(D; \beta_0)$ in the estimation space \mathcal{E} . In addition, we can directly see from Theorem 1.3 that the influ-

ence function proportional to $e_{a(u,Z)}(D; \beta_0) - r_{a(u,Z)}(D) + c_{a(u,Z)}(D)$ has smaller variance than the influence function proportional to $e_{a(u,Z)}(D; \beta_0)$, where $r_{a(u,Z)}(D) = (Z - \pi)f_0(X_1)$, and $c_{a(u,Z)}(D) = \int [g_0(u, Z, X) - E\{g_0(u, Z, X)|T \geq u, Z\}]dM_C(u, Z)/K_C(u, Z)$.

If we knew the true functions $f_0(\cdot)$ and $g_0(\cdot)$ (which is impossible in practice), then we could obtain the most efficient regular and asymptotic linear estimator for β within the subclass $\mathcal{B}_{a(u,Z)}$ by solving the estimating equation

$$\sum_{i=1}^n \left\{ \hat{e}_{a(u,Z)}(D_i; \beta) - \hat{r}_{a(u,Z)}(D_i) + \hat{c}_{a(u,Z)}(D_i) \right\} = 0, \quad (1.27)$$

where $\hat{e}_{a(u,Z)}(D_i; \beta) = \int \{a(u, Z_i) - \bar{a}(u; \beta)\}dN_i(u)$, $\hat{r}_{a(u,Z)}(D_i) = (Z_i - \pi)f_0(X_{1i})$, $\hat{c}_{a(u,Z)}(D_i) = \int \{g_0(u, Z_i, X_i) - \bar{g}_0(u, Z_i)\}dN_{C_i}(u)/\hat{K}_C(u, Z_i)$, and $\hat{K}_C(u, Z)$ is the Kaplan-Meier estimator for the censoring time C given Z .

In practice, however, the functions $f_0(\cdot)$ and $g_0(\cdot)$ are unknown to us and have to be estimated using the available data. We suggest the following strategy which is relatively easy to implement. Assume that the conditional expectation $f_0(X_1)$ given by (1.25) can be modeled by a parametric model $f(X_1; \mathbf{a}) = \mathbf{a}^T q(X_1)$ that is linear in \mathbf{a} and the conditional expectation of $g_0(u, Z, X)$ given by (1.26) can be modeled by $g(u, Z, X; \mathbf{b}) = \mathbf{b}^T w(u, Z, X)$ that is linear in \mathbf{b} , where \mathbf{a} and \mathbf{b} are r_a -dimensional and r_b -dimensional vectors of unknown parameters, respectively, $q(\cdot)$ is an r_a -dimensional vector of functions of X_1 and $w(\cdot)$ is an r_b -dimensional vector of functions of (u, Z, X) , and consider the subclass of regular and asymptotic linear estimators which solve the estimating equations $\sum_{i=1}^n \{\hat{e}_{a(u,Z)}(D_i; \beta) - (Z_i - \pi)f(X_{1i}; \mathbf{a}) - \int \{g(u, Z_i, X_i; \mathbf{b}) - \bar{g}(u, Z_i; \mathbf{b})\}dN_{C_i}(u)/\hat{K}_C(u, Z_i)\} = 0$, for all $\mathbf{a} \in \mathbb{R}^{r_a}$ and $\mathbf{b} \in \mathbb{R}^{r_b}$.

We define \mathbf{a}_0 and \mathbf{b}_0 to be the values leading to the smallest asymptotic variance of the estimator $\hat{\beta}$ within this subclass. Using standard regression methods, we obtain that, \mathbf{a}_0 satisfies $E[\{e_{a(u,Z)}(D; \beta_0) - (Z - \pi)\mathbf{a}_0^T q(X_1)\}(Z - \pi)q(X_1)^T \mathbf{a}] = 0$ for all $\mathbf{a} \in \mathbb{R}^{r_a}$. After some algebra, we derive that $\mathbf{a}_0 = [\pi(1 - \pi)E\{q(X_1)q(X_1)^T\}]^{-1}E\{q(X_1)(Z - \pi)e_{a(u,Z)}(D; \beta_0)\}$ and, consequently, \mathbf{a}_0 is estimated by

$$\hat{\mathbf{a}} = \left\{ \pi(1 - \pi) \sum_{i=1}^n q(X_{1i})q(X_{1i})^T \right\}^{-1} \sum_{i=1}^n \left\{ q(X_{1i})(Z_i - \pi)\hat{m}_{a(u,Z)}(D_i; \hat{\beta}_{\text{PH}}) \right\}, \quad (1.28)$$

where $\hat{m}_{a(u,Z)}(D_i, \beta) = \int \{a(u, Z_i) - \bar{a}(u; \beta)\} \{dN_i(u) - \hat{\lambda}(u; \beta) du \exp(\beta Z_i) Y_i(u)\}$ and $\hat{\lambda}(u; \beta) du$ is estimated using the increment of the Breslow estimate for the underlying cumulative hazard function, i.e., $\hat{\lambda}(u; \beta) du = \sum_i dN_i(u) / \sum_i \{\exp(\beta Z_i) Y_i(u)\}$. Similarly, we derive that $\mathbf{b}_0 = [E\{H_w(u, Z, X)H_w(u, Z, X)^T\}]^{-1} E\{H_w(u, Z, X)e_{a(u,Z)}(D; \beta_0)\}$, where $H_w(u, Z, X) = \int dM_C(u, Z)/K_C(u, Z)[w(u, Z, X) - E\{w(u, Z, X)|T \geq u, Z\}]$ and, \mathbf{b}_0 is estimated by

$$\hat{\mathbf{b}} = \left\{ \sum_{i=1}^n \hat{H}_w(u, Z_i, X_i) \hat{H}_w(u, Z_i, X_i)^T \right\}^{-1} \sum_{i=1}^n \left\{ \hat{H}_w(u, Z_i, X_i) \hat{m}_{a(u,Z)}(D_i; \hat{\beta}_{PH}) \right\}, \quad (1.29)$$

where $\hat{H}_w(u, Z_i, X_i) = \int \{dN_{C_i}(u) - \hat{\lambda}_{C|Z}(u|Z_i) du Y_i(u)\} \{w(u, Z_i, X_i) - \bar{w}(u, Z_i)\} / \hat{K}_C(u, Z_i)$, $\hat{\lambda}_{C|Z}(u|Z) du$, for $Z = 0, 1$ are estimated using the increment of the treatment-specific Nelson-Aalen estimator for the cumulative hazard function of the censoring distribution, that is, $\hat{\lambda}_{C|Z}(u|z) = \sum_i dN_{C_i}(u, z) / \sum_i Y_i(u, z)$, and $\bar{w}(u, Z_i)$ is calculated the same way as $\bar{g}(u, Z_i)$ in the estimating equation (1.6). We denote the estimated function $f(X_1; \hat{\mathbf{a}})$ by $\hat{f}_0(X_1)$ and the estimated function $g(u, Z, X; \hat{\mathbf{b}})$ by $\hat{g}_0(u, Z, X)$.

Obviously, from the definitions of $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$, using the estimated functions $\hat{f}_0(\cdot)$ and $\hat{g}_0(\cdot)$ in the estimating equation (1.27) will result in a more efficient estimator for β than the estimating equation

$$\sum_{i=1}^n \hat{e}_{a(u,Z)}(D_i; \beta) = 0, \quad (1.30)$$

even if the true conditional expectations (1.25) and (1.26) were not correctly specified. If, however, the true conditional expectations f_0 and g_0 are contained in the class of parametric models, the resulting estimator for β will be the most efficient within the subclass $\mathcal{B}_{a(u,Z)}$.

We consider the special case $a(u, Z) = Z$, where the estimating equation (1.30) becomes the partial likelihood score equation leading to the maximum partial likelihood estimator $\hat{\beta}_{PH}$ for β without auxiliary covariates. Let $\hat{\beta}_{AUG}$ be the estimator for β obtained by solving the equation (1.27) when the fitted functions $\hat{f}_0(\cdot)$ and $\hat{g}_0(\cdot)$ are used.

1.5.2 Variance estimator for $\hat{\beta}_{AUG}$

Using standard methods for censored survival analysis, we can show that the estimator $\hat{\beta}_{AUG}$ has influence function proportional to $m_2(D; \beta_0)$ with proportionality constant equal to

$(E[\Delta\{1 - Z^*(U; \beta_0)\}Z^*(U; \beta_0)])^{-1}$, where

$$\begin{aligned} m_2(D; \beta) &= e_{a(u, Z)}(D; \beta)|_{a(u, Z)=Z} - (Z - \pi)f(X_1; \mathbf{a}_0) \\ &+ \int \frac{dM_C(u, Z)}{K_C(u, Z)} \left[g(u, Z, X; \mathbf{b}_0) - E\{g(u, Z, X; \mathbf{b}_0) | T \geq u, Z\} \right]. \end{aligned}$$

Therefore, according to the results in chapter 4 of [7], we can estimate the variance of $\hat{\beta}_{\text{AUG}}$ by the sandwich variance

$$\hat{\sigma}_{\text{AUG}}^2 = \frac{\sum_{i=1}^n \hat{m}_2^2(D_i; \hat{\beta}_{\text{AUG}})}{\left[\sum_{i=1}^n \Delta_i \{ \bar{Z}(U_i; \hat{\beta}_{\text{AUG}}) - \bar{Z}^2(U_i; \hat{\beta}_{\text{AUG}}) \} \right]^2}, \quad (1.31)$$

where

$$\begin{aligned} \hat{m}_2(D_i; \beta) &= \hat{m}_{a(u, Z)}(D_i; \beta)|_{a(u, Z)=Z} - (Z_i - \pi)\hat{f}_0(X_{1i}) \\ &+ \int \frac{\{dN_{Ci}(u) - Y_i(u)\hat{\lambda}_{C|Z}(u|Z_i)du\}}{\hat{K}_C(u, Z_i)} \left[\hat{g}_0(u, Z_i, X_i) - \bar{g}_0(u, Z_i) \right]. \end{aligned}$$

Here, $\bar{Z}(u; \beta)$ and $\bar{g}_0(u, Z)$ are calculated as in the estimating equation (1.6).

1.6 Simulation

We performed a Monte-Carlo simulation study to evaluate the performance of the proposed estimator $\hat{\beta}_{\text{AUG}}$ and to compare it to the standard maximum partial likelihood estimator $\hat{\beta}_{\text{PH}}$. For this study, we considered only one baseline covariate X . In order to ensure that the covariate X was independent of treatment assignment Z , was correlated to the survival time T , and that the conditional distribution of T given Z followed a proportional hazards relationship, we generated the data in the following manner. First we generated bivariate data (Y, X) from a bivariate normal density with mean zero, variance 1, and correlation ρ . We then independently generated the treatment indicator Z as a Bernoulli(π). Using inverse transformation, the survival time T was taken to be $T = -\exp(-\beta Z) \log\{1 - \Phi(Y)\}$, where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal. This guarantees that the distribution of T given Z will follow a proportional hazards relationship $\lambda(t|z) = \lambda(t) \exp(\beta z)$, with $\lambda(t) = 1$, that is,

$T \sim \text{Exp}\{\exp(\beta Z)\}$. Censoring for each treatment $Z = 0, 1$ was generated independently as an exponential distribution $C|Z \sim \text{Exp}(c|z)$. As for the functions $f_0(X)$ and $g_0(u, Z, X)$, we posited the models $a_0 + a_1X + a_2X^2$ for $f_0(X)$, and $a_0 + a_1X + a_2X^2 + a_3XZ$ for $g_0(u, Z, X)$.

For this demonstration, treatment was assigned with probability $\pi = .5$, the correlation between the bivariate normal random variable was taken to be $\rho = .7$ which resulted in a sample correlation of approximately 0.6 between the survival time T and baseline covariate X . Two values for the proportional hazards regression coefficient were considered, $\beta = 0$ (null hypothesis) and $\beta = .25$, and for each treatment $Z = 0, 1$, the value c for the exponential distribution of the censoring variable was taken to be two values that would result in roughly 25% and 50% of the data being censored. Sample sizes of 250 and 600 were considered and each scenario used 2000 Monte-Carlo simulations. In tables 1.1 and 1.2, we compare the bias, standard error estimate, Monte-Carlo standard error, relative efficiency (ratio of variance estimate and ratio of Monte-Carlo variance), type I error and the power of the maximum partial likelihood estimator $\hat{\beta}_{\text{PH}}$ and our proposed estimator $\hat{\beta}_{\text{AUG}}$ for the various simulation scenarios. To assess the performance of covariate-adjusted estimators for treatment effect, we also considered two additional estimators. These two estimators are $\hat{\beta}_*$, the estimate of β_* in model (1.2), and $\hat{\beta}_{**}$, the estimate of β_* in (1.2) after adding the quadratic term X^2 in the model.

Surprisingly, the results show that, under the null hypothesis ($\beta = 0$), the traditional estimators, $\hat{\beta}_*$ and $\hat{\beta}_{**}$, based on the conditional proportional hazards model on X and X^2 are less efficient than the maximum partial likelihood estimator $\hat{\beta}_{\text{PH}}$ with 0-27% drops in efficiency. Moreover, the Wald tests induced by $\hat{\beta}_*$ and $\hat{\beta}_{**}$ are biased with type I errors around 0.08 at the nominal level 0.05. Under the case $\beta = 0.25$, the estimates of $\hat{\beta}_*$ and $\hat{\beta}_{**}$ are biased as would be expected.

In contrast, our proposed estimator $\hat{\beta}_{\text{AUG}}$ is unbiased but more efficient than the maximum partial likelihood estimator $\hat{\beta}_{\text{PH}}$ with 40-80% gains in efficiency.

1.7 A brief example

We also applied this methodology to the data from 2139 patients from the AIDS Clinical Trials Group (ACTG) protocol 175 ([9]), a study that randomized patients to four

Table 1.1: Simulation Results for $\beta_0 = 0$ (2000 Monte-Carlo samples, entries in parentheses for SE and Monte-Carlo SE are relative efficiencies with respect to $\hat{\beta}_{\text{PH}}$.)

	Cen	n	$\hat{\beta}_{\text{PH}}$	$\hat{\beta}_{\text{AUG}}$	$\hat{\beta}_*$	$\hat{\beta}_{**}$
Bias	25%	250	0.002	−0.004	−0.004	−0.003
		600	0.001	−0.002	−0.001	−0.001
	50%	250	0.001	−0.004	−0.006	−0.007
		600	−0.002	−0.004	−0.003	−0.003
SE	25%	250	0.148	0.112(1.74)	0.149(0.987)	0.150(0.974)
		600	0.095	0.072(1.72)	0.095(1.00)	0.095(1.00)
	50%	250	0.182	0.141(1.66)	0.183(0.989)	0.184(0.978)
		600	0.116	0.091(1.63)	0.117(0.983)	0.117(0.983)
MCSE	25%	250	0.146	0.118(1.53)	0.170(0.738)	0.170(0.738)
		600	0.095	0.073(1.67)	0.106(0.803)	0.107(0.788)
	50%	250	0.185	0.156(1.40)	0.210(0.776)	0.212(0.762)
		600	0.117	0.095(1.52)	0.131(0.798)	0.132(0.786)
Size	25%	250	0.0495	0.064	0.082	0.079
		600	0.0455	0.0525	0.081	0.083
	50%	250	0.0515	0.073	0.080	0.084
		600	0.049	0.060	0.077	0.080

Table 1.2: Simulation Results for $\beta_0 = 0.25$ (2000 Monte-Carlo samples, entries in parentheses for SE and Monte-Carlo SE are relative efficiencies with respect to $\hat{\beta}_{\text{PH}}$.)

	Cen	n	$\hat{\beta}_{\text{PH}}$	$\hat{\beta}_{\text{AUG}}$	$\hat{\beta}_*$	$\hat{\beta}_{**}$
Bias	25%	250	0.004	−0.002	0.092	0.092
		600	−0.008	−0.007	0.091	0.091
	50%	250	0.004	0.0002	0.083	0.084
		600	−0.003	−0.004	0.083	0.083
SE	25%	250	0.149	0.113(1.73)	0.151(0.974)	0.152(0.961)
		600	0.095	0.073(1.71)	0.096(0.979)	0.097(0.959)
	50%	250	0.183	0.142(1.65)	0.186(0.968)	0.187(0.958)
		600	0.117	0.092(1.62)	0.118(0.983)	0.118(0.983)
MCSE	25%	250	0.147	0.119(1.52)	0.171(0.739)	0.171(0.739)
		600	0.096	0.075(1.65)	0.106(0.820)	0.107(0.805)
	50%	250	0.187	0.157(1.41)	0.212(0.778)	0.214(0.764)
		600	0.119	0.097(1.52)	0.132(0.813)	0.133(0.801)
power	25%	250	0.4055	0.593	0.612	0.609
		600	0.7095	0.915	0.926	0.925
	50%	250	0.282	0.428	0.438	0.438
		600	0.5565	0.7525	0.782	0.779

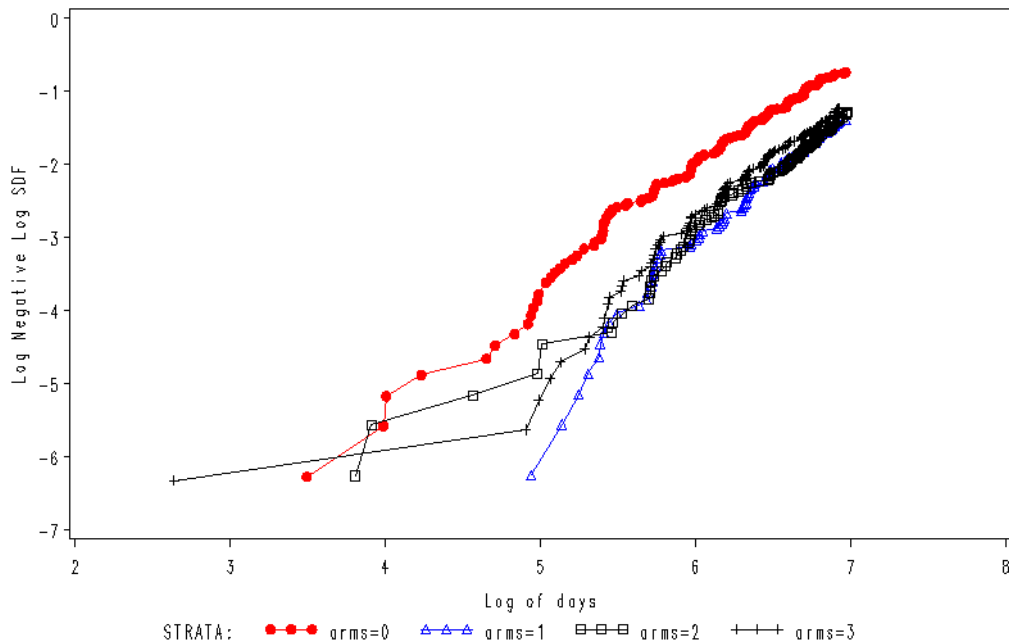


Figure 1.1: Log negative log survival function of time to death for each treatment

antiretroviral regimes in equal proportions. The main focus of this study was two-sample comparisons of treatment 0 (Zidovudine) versus treatment 1 (Zidovudine and didanosine), treatment 2 (Zidovudine and zalcitabine) and treatment 3 (Didanosine), respectively. In this study, 532 patients were randomized to treatment 0, 522 were randomized to treatment 1, 524 were randomized to treatment 2 and 561 were randomized to treatment 3. The primary endpoint was a combined endpoint corresponding to the first time that a patient had a ≥ 50 percent decline in their CD4 cell count, an event indicating progression to the acquired immunodeficiency syndrome (AIDS), or death. Roughly 76% of the data were censored.

Fig.1.1 is a plot of the logarithm of the negative logarithm of the survival distribution for the time to death for each treatment. The four lines, except for a few points early in time, are approximately parallel suggesting that a proportional hazards relationship between treatments is reasonable. The results of applying the standard analysis using Cox's maximum partial likelihood estimator can be found in Table 2.4. For example, the estimate of the log hazard ratio between treatment 0 and treatment 1 is -0.703 and its standard error is .124, which is highly statistically significant.

Table 1.3: Estimates of $\hat{\beta}_{\text{PH}}$ and $\hat{\beta}_{\text{AUG}}$ on the ACTG 175 data (*RE is the relative efficiencies with respect to $\hat{\beta}_{\text{PH}}$.*)

		Estimates	Standard Errors	RE
Treatment 0 and 1	$\hat{\beta}_{\text{PH}}$	-0.703	0.124	1.00
	$\hat{\beta}_{\text{AUG}}$	-0.723	0.110	1.25
Treatment 0 and 2	$\hat{\beta}_{\text{PH}}$	-0.640	0.121	1.00
	$\hat{\beta}_{\text{AUG}}$	-0.555	0.104	1.36
Treatment 0 and 3	$\hat{\beta}_{\text{PH}}$	-0.528	0.116	1.00
	$\hat{\beta}_{\text{AUG}}$	-0.627	0.105	1.21

In our analysis, we also considered several prognostic baseline auxiliary covariates and post-treatment covariates. The baseline covariates include CD4, CD8, age (years), weight (kg), history of IV drug use (0 = no), Karnofsky score (on a scale of 0-100), Zidovudine in the 30 days prior to 175 (0 = no), number of days pre-175 antiretroviral therapy and symptomatic status indicator (0 = asymptomatic). The post-treatment covariates include, CD4 at 20 ± 5 weeks, CD8 at 20 ± 5 weeks, CD4 at 96 ± 5 weeks ($= -1$ if missing), indicator of off-trt before 96 ± 5 weeks (0 = no, 1 = yes) and Missing CD4 at 96 ± 5 weeks (0 = missing, 1 = observed). The results for our proposed estimators applied to the ACTG 175 data are also shown in Table 2.4.

The estimator $\hat{\beta}_{\text{AUG}}$ has estimates of standard errors (0.110, 0.104 and 0.105, respectively), all are more efficient (roughly, 25%, 36% and 21% increase in efficiency, respectively) than the corresponding maximum partial likelihood estimator for β . These results are consistent with the results from such moderately prognostic auxiliary covariates.

1.8 Concluding Remarks

In this paper, we use the theory of semiparametrics to derive an augmented class of regular and asymptotic linear estimators for β , the proportional hazards regression parameter, when auxiliary covariates are introduced. Within this class of estimators we show how estimators and tests can be derived that are more efficient than the traditional maximum partial likelihood estimator and the logrank test. The proposed method results in easy-to-compute estimators which are guaranteed to be more efficient than the maximum partial likelihood esti-

mator for β . As a consequence, the Wald tests induced by these estimators are more powerful than the logrank test.

In deriving these results, one of the crucial assumption we have made is $C \perp\!\!\!\perp (T, X)|Z$ in assumption (1.4). We do note, however, that if the primary reason for the censoring is administrative, we may be willing to make the stronger assumption that censoring is also independent of treatment assignment Z , that is, $C \perp\!\!\!\perp (T, Z, X)$. Under this stronger assumption, the class of regular and asymptotically linear estimators can be shown to be slightly larger by solving the estimating equations,

$$\sum_{i=1}^n \left[\int \{a(u, Z_i) - \bar{a}(u; \beta)\} dN_i(u) + (Z_i - \pi) f(X_{1i}) + \int \{g(u, Z_i, X_i) - \bar{g}(u)\} dN_{C_i}(u) \right] = 0, \quad (1.32)$$

where (1.32) is similar to the estimating equations given in (1.6) with the only difference being that $\bar{g}(u, Z_i)$ is substituted by $\bar{g}(u) = \sum_{j=1}^n \{g(u, Z_j, X_j) Y_j(u)\} / \sum_{j=1}^n \{Y_j(u)\}$. A slightly more efficient estimator can then be obtained by using the same functions $a(u, Z)$, $\hat{f}_0(X_1)$ and $\hat{g}_0(u, Z, X)$ that were derived in §1.5.

Chapter 2

Estimation of treatment effect with time-lagged response in the presence of informative censoring

2.1 Introduction

In many randomized clinical trials, the primary endpoint of interest, that is, the primary response variable, denoted here by Y , is not observed directly after patients enroll into a study, but rather is observed after some period of time which may vary among patients. The time from a patient's entry into the study until the response is observed is referred to as the lag time or time to ascertainment, and the corresponding response variable is referred to as the time-lagged response ([10]) or marked point process ([11]). The simplest example of a time-lagged response is survival time where the primary endpoint itself is the time lag. In this case, the lag time (i.e., survival time) varies by individual. Another example of a time-lagged response may be a laboratory measurement taken after some fixed period of time. For instance, in a study of HIV disease we may be interested in the CD4 count one year after a patient is randomized to some treatment. In this case, the lag time (one year) is the same for all individuals. Yet another example is the total medical costs incurred during the treatment of some disease. Here the response variable of interest Y is the total medical cost but the time to

ascertainment is the length of disease which will vary by individual.

Our primary goal here is to estimate or test for the treatment effect between two competing treatment groups, e.g., a new treatment versus placebo, which is defined through the statistical model

$$p_{Y|Z}(y|z; \beta, \eta) \tag{2.1}$$

where Z denotes the treatment assignment with $Z = 1$ for new treatment and $Z = 0$ for placebo, $p_{Y|Z}(y|z)$ is used to denote the conditional density of Y given Z , β is the treatment effect parameter of interest and η are nuisance parameters used to describe the class of conditional distributions of $Y|Z$. The lag time for individual i ; i.e., the time it takes for the response variable Y_i to be ascertained will be denoted by T_i . Of course, if all the patients in a study are followed until their response is ascertained, then the lag time itself does not add any useful additional information regarding the estimation of the treatment effect β .

It is very common, however, in such clinical trials that the time-lagged response data Y are missing because of right censoring of some patients. More precisely, it is the lag time T that is right censored and if this is the case then the response variable Y will be missing. Depending on the study, censoring occurs for a variety of reasons. Administrative censoring occurs because patients enter the study in a staggered fashion and not all have been observed at the end of the study when the data are analyzed. Often it is assumed that the censoring time is independent of the primary response Y of interest or the slightly weaker assumption that the censoring time is independent of Y given treatment assignment Z . The assumption of independence between censoring time and response Y given Z is also necessary for some commonly used standard methods, for example, the maximum partial likelihood estimator of Cox ([3]) and the logrank test ([1], [2]), to ensure their properties, such as consistency or asymptotic normality, to hold. This assumption is often referred to as noninformative censoring. However, censoring may also occur due to a patient's drop out of the study before their response data are observed. For example, patients may drop out of the study because of side effects, or prognostically worse or better patients may drop out for reasons that can be attributed to other time-dependent outcomes. Under such situations, the censoring time is likely to be dependent on the response Y given Z and such censoring is usually referred to as the informative censoring. Informative

censoring, if not properly accounted for, may bias the results from standard inferential methods and give overly optimistic or pessimistic estimates of treatment effect.

In addition to the data on the time-lagged response, the lag and censoring times and treatment assignment, some auxiliary information (for example, age, gender, health conditions, etc.) are also collected in most clinical trials. Some of these auxiliary variables are collected prior to randomization, while others may be collected after treatment assignment. Because of randomization, covariates collected prior to randomization, referred to as baseline auxiliary covariates, are independent of treatment assignment and are not affected by treatment, whereas, covariates measured after randomization, referred to as post-treatment auxiliary covariates, may be time-dependent and affected by treatment assignment. Nonetheless, some of these covariates may be important prognostic factors that are correlated with the primary response variable.

Lu and Tsiatis ([12]) considered the special case where the survival time itself is the time-lagged response and censoring is independent of Y conditional on Z , and used semiparametric theory to derive an augmented class of consistent and asymptotically normal estimators for the treatment-specific log hazard ratio regression parameter. The auxiliary covariates were used to derive estimators that are more efficient than the maximum partial likelihood estimator and the logrank test.

In this paper, we use semiparametric theory and the major results in [13] to derive a class of consistent and asymptotically normal estimators for the treatment effect parameter β but weaken the assumption of independence between censoring time and the response variable given treatment assignment to allow for censoring that is informative in a manner that can be explained through the observed auxiliary covariates. The auxiliary covariates here play an important role in deriving a class of augmented semiparametric consistent and asymptotically normal estimators for β when the censoring is informative. The correlations between auxiliary covariates and the primary response variable are also utilized to derive estimators that are more efficient than the estimators without using the auxiliary covariates even if the censoring was noninformative.

This article is organized as follows. §2.2 describes the notation and model assumptions which will be used throughout this article. §2.3 characterizes the class of regular and asymptotic linear estimators for β using a general time-lagged responses. §2.4 considers a specific example

where the underlying relationship between the survival time distribution and treatment assignment follows a proportional hazards model and applies the major results in §2.3 to characterize a subclass of regular and asymptotic linear estimators for β , the treatment-specific log hazard ratio, when informative censoring exists. We also proposed some easy-to-compute estimators and perform a series of simulation studies to compare them with the commonly used logrank test in §2.5. A brief example is given in §2.6.

2.2 Model framework and notation

2.2.1 Notation and assumptions

Consider a randomized clinical trial where n subjects are sampled from a population of interest. Let $D_i = \{U_i, \Delta_i, \Delta_i Y_i, Z_i, X_i(U_i)\}$ denote the observed data that are independent and identically distributed random vectors for $i = 1, \dots, n$. For the i -th subject, $U_i = \min(T_i, C_i)$, where T_i denotes the underlying lag time, and C_i denotes the potential censoring time, $\Delta_i = I(T_i \leq C_i)$ is an indicator of whether the response data were ascertained ($\Delta_i = 1$) or missing ($\Delta_i = 0$), Y_i denotes the response on which the primary analysis will be based, where Y_i may be continuous or discrete and is only observed if $\Delta_i = 1$, Z_i denotes the treatment indicator with value 1 and 0 denoting experimental treatment and placebo, respectively. Furthermore, we let $X_i(U_i) = \{X_{1i}, X_{2i}^H(U_i)\}$, where X_{1i} denotes a vector of baseline auxiliary covariates which are measured prior to randomization, and $X_{2i}^H(U_i)$ defined by $\{X_{2i}(u), 0 \leq u \leq U_i\}$ denotes post-treatment auxiliary covariates which may be time-dependent in which case we would observe the history of these values up to time U_i . In addition, we use $V_i = \{T_i, Y_i, Z_i, X_i(T_i)\}$ to denote the full data had there been no censoring or missing data.

We first notice from the data that, because of randomization, it is reasonable to assume that the treatment indicator Z is independent of the auxiliary baseline covariates X_1 and that the randomization probability of receiving treatment 1 is equal to π with $0 < \pi < 1$ which is known to us; that is,

$$Z \perp\!\!\!\perp X_1 \text{ and } \Pr(Z = 1) = \pi. \quad (2.2)$$

As in any missing data problem it is important to consider assumptions regarding the

process in which the data are missing (censored); that is, we need to consider the conditional distribution of the censoring variable C given the full data V which we define through the conditional hazard function $\lambda_C(u|V)$ at time u given V . Often, in randomized clinical trials, one assumes that the censoring variable C is completely independent of the full-data V . This assumption is similar in spirit to the "missing completely at random" assumption (MCAR) as defined by Rubin ([14]) and may be a reasonable assumption if the data were administratively censored. A slightly weaker assumption that is implicitly made when one uses the logrank test to test for differences in the survival distributions for two treatments with right-censored data is that C is conditionally independent of (T, Y) given Z . When a patient is censored due to drop out or lost to follow-up, then one can imagine some scenarios where poorer prognostic patients may be more likely to be censored and other scenarios where the opposite may happen. We therefore consider the weaker assumption that

$$\lambda_C(u|T \geq u, V) = \lambda_C\{u|T \geq u, Z, X(u)\}. \quad (2.3)$$

In words, this assumption means that the probability of being censored at time u , given that one has not been censored or failed by time u , only depends on observed variables measured prior to time u and not additionally on the future data. This last assumption is similar in spirit to what Rubin refers to as missing at random (MAR) and we will refer to this assumption as censoring at random (CAR). This CAR assumption allows greater flexibility than the usual assumptions with censored data. It is also clear that if the covariates $X(u)$ that make this assumption tenable were not considered then informative censoring of C and Y given Z would be induced.

Without making additional assumptions, other than those given by (2.2) and (2.3), we now consider how to derive a class of semiparametric estimators that are consistent and asymptotically normal for β using the results by Robins and Rotnitzky ([13]).

2.2.2 Introduction to semiparametrics theory

Robins and Rotnitzky ([13]) restricted attention to estimators that are *regular and asymptotically linear* (RAL). An estimator $\hat{\beta}_n$ for β is asymptotically linear if there exists a

random variable $\varphi(D)$, which, under the truth, $\beta = \beta_0$, has mean zero and finite variance, such that $n^{1/2}(\hat{\beta}_n - \beta_0) = n^{-1/2} \sum_{i=1}^n \varphi(D_i) + o_p(1)$. The function $\varphi(D_i)$ is referred to as the i -th *influence function* of the estimator $\hat{\beta}_n$. Regularity is a technical condition that rules out "pathological" estimators with undesirable local properties ([8]), such as the "superefficient" estimator of Hodges (e.g., [7], page 24). The influence function of a regular and asymptotic linear estimator for β is uniquely defined and the asymptotic properties of such an estimator is determined by its influence function. It is clear from the definition of the influence function given above and a simple application of the central limit theorem, that, the asymptotic variance of an regular and asymptotic linear estimator $\hat{\beta}_n$ is equal to the variance of its influence function, i.e., $E\{\varphi(D)^2\}$.

For a general semiparametric model, Robins and Rotnitzky ([13]) provided a series of steps for deducing a class of regular and asymptotically linear estimators when data are censored at random: (1) characterize the class of full data influence functions, (2) characterize the class of observed data influence functions by applying the theory of augmented inverse probability complete case estimators derived by Robins and Rotnitzky ([13]) and (3) identify the observed data estimators with estimating functions in this class.

2.3 The class of all semiparametric estimators for β

The class of full data influence functions for β , whose proof was given by Zhang, Tsiatis and Davidian ([15]), is characterized by

$$\Psi^F = \Psi_{YZ} + \mathcal{R}, \quad (2.4)$$

where $\Psi_{YZ} = \{\psi_{YZ}(Y, Z; \beta_0)\}$ is the class of influence functions for β that only use the information of time-lagged response variable Y and treatment assignment Z that are derived through model (2.1), and

$$\mathcal{R} = \left\{ (Z - \pi)f(X_1) \text{ for all functions } f(X_1) \right\}. \quad (2.5)$$

For the time being, assume the hazard function $\lambda_C\{u|T \geq u, Z, X(u)\}$ to be known and

let $K_C\{t, Z, X(u)\} = \exp \left[- \int_0^t \lambda_C\{u|T \geq u, Z, X(u)\} du \right]$ be the conditional survival function for censoring. Under the CAR assumption, we obtain that $P(\Delta = 1|V) = K\{T, Z, X(T)\}$. The theory of Robins and Rotnitzky ([13]), see also Chapter 9.2 of [7], tells us that the an observed data influence function for β can be written as an augmented inverse probability weighted complete case influence function; namely,

$$\frac{\Delta \psi^F(V; \beta_0)}{K_C\{U, Z, X(U)\}} + \int \frac{dM_C\{u, Z, X(u)\}}{K_C\{u, Z, X(u)\}} L\{u, Z, X(u)\}, \quad (2.6)$$

where the full-data influence function $\psi^F(V; \beta_0) \in \Psi^F$ is given in (2.4), $dM_C\{u, Z, X(u)\}$ is the increment of the martingale process $dN_C(u) - \lambda_C\{u|T \geq u, Z, X(u)\}Y(u)du$, where $dN_C(u)$ is the increment of the counting process for censoring; i.e., $N_C(u) = I(U \leq u, \Delta = 0)$, $Y(u) = I(U \geq u)$, is the “at-risk” process, and $L\{u, Z, X(u)\}$ is an arbitrary function of all the data observed prior to time u ; that is,

Theorem 2.1. *If the hazard function $\lambda_C\{u|T \geq u, Z, X(u)\}$ for censoring is known, the class of all observed data influence functions is given by*

$$\Psi^{obs} = \frac{\Delta \Psi_{YZ}}{K_C\{U, Z, X(U)\}} + \{\mathcal{R} \oplus \mathcal{C}\} \quad (2.7)$$

where Ψ_{YZ} and \mathcal{R} are defined in (2.4) and (2.5), respectively, and

$$\mathcal{C} = \left[\int dM_C\{u, , Z, X(u)\} L\{u, Z, X(u)\} : \text{for all functions } L(\cdot) \right]. \quad (2.8)$$

Theorem 2.1, whose proof is given in Appendix A.4, tells that if $\lambda_C\{u|T \geq u, Z, X(u)\}$ is known, then the class of regular and asymptotically linear estimators for β can be represented as the solution to the following estimating equations

$$\sum_{i=1}^n \left[\frac{\Delta_i \psi_{YZ}(Y_i, Z_i; \beta)}{K_C\{U_i, Z_i, X_i(U_i)\}} + (Z_i - \pi) f(X_{1i}) + \int dM_{C_i}\{u, Z_i, X_i(u)\} L\{u, Z_i, X_i(u)\} \right] = 0, \quad (2.9)$$

for arbitrary functions $\psi_{YZ}(Y, Z; \beta_0) \in \Psi_{YZ}$, $f(X_1)$ and $L\{u, Z, X(u)\}$.

In practice, however, the hazard function $\lambda_C\{u|T \geq u, Z, X(u)\}$ (also used to derive $K_C\{u, Z, X(u)\}$) is not known, and must be estimated from the data. Because of its availability

and versatility, a proportional hazards regression models with time-dependent covariates will be used to model the hazard relationship for the censoring distribution in (2.3) together with the Breslow estimator ([16]) to estimate the underlying censoring survival distribution. For maximum flexibility, we will consider separate models for each treatment group. To be specific, we posit a stratified proportional hazards regression model for the censoring time C , that is, $\lambda_C\{u|T \geq u, Z, X(u)\} = \lambda_{0C}(u, Z)H\{Z, X(u); \alpha\}$, where $\lambda_{0C}(u, Z)$ is the baseline hazard function for each treatment group $Z = 0, 1$, and $H\{Z, X(u); \alpha\} = Z \exp[\alpha_{11}X_1 + \alpha_{21}g_1\{X_2^H(u)\}] + (1 - Z) \exp[\alpha_{12}X_1 + \alpha_{22}g_2\{X_2^H(u)\}]$. The vector of parameters $\alpha = (\alpha_{11}, \dots, \alpha_{22})$ will be estimated by using the standard partial likelihood estimator ([6]), denoted here by $\hat{\alpha}$, and $\lambda_{0C}(u, Z)$ will be estimated by

$$\sum_{j=1}^n \{dN_{C_j}(u)I(Z_j = Z)\} / \sum_{j=1}^n [H\{Z_j, X_j(u); \hat{\alpha}\}Y_j(u)I(Z_j = Z)].$$

Denote the estimated hazard function and corresponding survival function by $\hat{\lambda}_C(\cdot)$ and $\hat{K}_C(\cdot)$, respectively. Using these estimated functions in (2.9), then after some algebra, we have an equivalent expression of the estimating equation

$$\sum_{i=1}^n \left(\frac{\Delta_i \psi_{YZ}(Y_i, Z_i; \beta)}{\hat{K}_C\{U_i, Z_i, X_i(U_i)\}} + (Z_i - \pi)f(X_{1i}) + \int [L\{u, Z_i, X_i(u)\} - \bar{L}(u, Z_i)]dN_{C_i}(u) \right) = 0, \quad (2.10)$$

where

$$\bar{L}(u, Z_i) = \frac{\sum_{j=1}^n \{L(u, Z_j, X_j)H\{Z_i, X_j(u); \hat{\alpha}\}Y_j(u)I(Z_j = Z_i)\}}{\sum_{j=1}^n \{H\{Z_i, X_j(u); \hat{\alpha}\}Y_j(u)I(Z_j = Z_i)\}} \quad (2.11)$$

for arbitrary functions $\psi_{YZ}(Y, Z; \beta_0) \in \Psi_{YZ}$, $f(X_1)$ and $L\{u, Z, X(u)\}$.

In addition, if we consider the subclass of estimators of (2.10) for any fixed function $\psi_{YZ}(Y, Z; \beta_0) \in \Psi_{YZ}$, following the analogous proof as that of Theorem 3 of [12], we can derive the optimal functions

$$f_0(X_1) = \frac{1}{\pi(1 - \pi)} E\{(Z - \pi)\psi_{YZ}(Y, Z; \beta_0)|X_1\}, \quad \text{and} \quad (2.12)$$

$$L_0\{u, Z, X(u)\} = \frac{E\{\psi_{YZ}(Y, Z; \beta_0)|T \geq u, Z, X(u)\}}{K_C\{u, Z, X(u)\}} \quad (2.13)$$

for $f(\cdot)$ and $L(\cdot)$, respectively, that lead to the most efficient estimator with the smallest variance in this subclass. However, the conditional expectations in equations (2.12) and (2.13) are unknown in practice and they will be obtained by positing regression models in a manner similar to that described in Section 5 of [12]. Substituting these estimated functions $\hat{f}(\cdot)$ and $\hat{L}(\cdot)$ for $f(\cdot)$ and $L(\cdot)$ in (2.10) will lead to a more efficient estimator for β than using any other functions $f(\cdot)$ and $L(\cdot)$.

2.4 Application to the proportional hazard model

2.4.1 Class of all semiparametric estimators for β

To illustrate how these estimators are derived, we will focus on the special case where the time-lagged response variable Y is the survival time T and $T|Z$ follows a proportional hazard model

$$\lambda_{T|Z}(t|z) = \lambda(t) \exp(\beta z), \quad (2.14)$$

where $\lambda_{T|Z}(t|z)$ denotes the conditional hazard rate of failing at time t given treatment $Z = z$ for $z = 0, 1$, respectively. This problem is common in chronic disease clinical trials and is of importance in its own right.

Using standard results for the proportional hazards model; see, for example, ([7], Chapter 5.2), the class of full data influence functions Ψ_{YZ} for estimators of β can be described by

$$\Psi_{YZ} = \left\{ C_a^{-1} \int a(u) \{Z - Z_T^*(u; \beta_0)\} dM_T(u, Z; \beta_0) : \text{for any function } a(u) \right\}, \quad (2.15)$$

where $dM_T(u, Z; \beta) = dN_T(u) - \lambda(u) \exp(\beta Z) I(T \geq u) du$, $N_T(u) = I(T \leq u)$, $Z_T^*(u, \beta) = E\{Z \exp(\beta Z) I(T \geq u)\} / E\{\exp(\beta Z) I(T \geq u)\}$, and the proportionality constant $C_a = E[a(T) \{1 - Z_T^*(T; \beta_0)\} Z_T^*(T; \beta_0)]$.

Applying this class of full data influence functions Ψ_{YZ} to Theorem 2.1, we have the following corollary, whose proof is given in Appendix A.5.

Corollary 2.2. *If the hazard function $\lambda_C\{u|T \geq u, Z, X(u)\}$ for censoring is known, and if $T|Z$*

follows a proportional hazard model (2.14), the class of all observed data influence functions is given by

$$\Psi_{PH}^{obs} = \mathcal{E} + \{\mathcal{R} \oplus \mathcal{C}\} \quad (2.16)$$

where

$$\mathcal{E} = \left[C_W^{-1} \int \{Z - Z^*(u; \beta_0)\} \frac{W(u, Z) dM(u, Z; \beta_0)}{K_C\{u, Z, X(u)\}} : \text{for all functions } W(\cdot) \right], \quad (2.17)$$

$$Z^*(u; \beta) = \frac{E[Z \exp(\beta Z) Y(u) W(u, Z) / K_C\{u, Z, X(u)\}]}{E[\exp(\beta Z) Y(u) W(u, Z) / K_C\{u, Z, X(u)\}]}, \quad (2.18)$$

$$C_W = E[\Delta W(U, Z) / K_C\{U, Z, X(U)\} \{1 - Z^*(U; \beta_0)\} Z^*(U; \beta_0)], \quad (2.19)$$

and \mathcal{R} and \mathcal{C} are defined in (2.5) and (2.8), respectively.

Analogous to (2.10), if we use the estimated functions $\hat{\lambda}_C(\cdot)$ and $\hat{K}_C(\cdot)$, assuming the stratified proportional hazards regression model, then the class of all regular and asymptotically linear estimators for β can be represented as a solution to the following estimating equation

$$\begin{aligned} 0 &= \sum_{i=1}^n \left(\int \{Z_i - \bar{Z}(u; \beta)\} \frac{W(u, Z_i) dN_i(u)}{\hat{K}_C\{u, Z, X(u)\}} + (Z_i - \pi) f(X_{1i}) \right. \\ &\quad \left. + \int [L\{u, Z_i, X_i(u)\} - \bar{L}(u, Z_i)] dN_{C_i}(u) \right), \end{aligned} \quad (2.20)$$

where

$$\bar{Z}(u; \beta) = \frac{\sum_j^n [Z_j \exp(\beta Z_j) Y_j(u) W(u, Z_j) / \hat{K}_C\{u, Z_j, X_j(u)\}]}{\sum_j^n [\exp(\beta Z_j) Y_j(u) W(u, Z_j) / \hat{K}_C\{u, Z_j, X_j(u)\}]}, \quad (2.21)$$

for arbitrary functions $W(u, Z)$, $f(X_1)$ and $L\{u, Z, X(u)\}$. Again, we can also derive a more efficient estimator for β by substituting $\hat{f}(\cdot)$ and $\hat{L}(\cdot)$ for $f(\cdot)$ and $L(\cdot)$ in (2.20), where $\hat{f}(\cdot)$ and $\hat{L}(\cdot)$ are obtained using a similar algorithm as that in Section 5 of [12]. To be specific, we posit parametric models $f(X_1; \mathbf{a}) = \mathbf{a}^T q(X_1)$ that is linear in \mathbf{a} and $L\{u, Z, X(u); \mathbf{b}\} = \mathbf{b}^T s\{u, Z, X(u)\}$ that is linear in \mathbf{b} , where \mathbf{a} and \mathbf{b} are r_a -dimensional and r_b -dimensional vectors of unknown parameters, respectively, $q(\cdot)$ is an r_a -dimensional vector of functions of X_1 and $w(\cdot)$ is an r_b -dimensional vector of functions of (u, Z, X) , and consider the subclass of regular and asymptotic

linear estimators which solve the estimating equations

$\sum_{i=1}^n \left\{ \int \{Z_i - \bar{Z}(u; \beta)\} dN_i(u) W(u, Z_i) / \hat{K}_C\{u, Z, X(u)\} - (Z_i - \pi) f(X_{1i}; \mathbf{a}) - \int [L\{u, Z_i, X_i(u); \mathbf{b}\} - \bar{L}(u, Z_i; \mathbf{b})] dN_{Ci}(u) \right\} = 0$, for all $\mathbf{a} \in \mathbb{R}^{r_a}$ and $\mathbf{b} \in \mathbb{R}^{r_b}$. We define \mathbf{a}_0 and \mathbf{b}_0 to be the values leading to the smallest asymptotic variance of the estimator $\hat{\beta}$ within this subclass. Using standard regression methods, we obtain the estimators

$$\hat{\mathbf{a}} = \left\{ \pi(1 - \pi) \sum_{i=1}^n q(X_{1i}) q(X_{1i})^T \right\}^{-1} \sum_{i=1}^n \left\{ q(X_{1i}) (Z_i - \pi) \hat{m}_{W(u, Z)}(D_i; \hat{\beta}_{\text{PH}}) \right\}, \text{ and} \quad (2.22)$$

$$\hat{\mathbf{b}} = \left\{ \sum_{i=1}^n \hat{G}_s\{u, Z_i, X_i(u)\} \hat{G}_s\{u, Z_i, X_i(u)\}^T \right\}^{-1} \sum_{i=1}^n \left\{ \hat{G}_s\{u, Z_i, X_i(u)\} \hat{m}_{W(u, Z)}(D_i; \hat{\beta}_{\text{PH}}) \right\}, \quad (2.23)$$

for \mathbf{a}_0 and \mathbf{b}_0 , respectively, where

$$\hat{m}_{W(u, Z)}(D_i; \beta) = \int \{Z_i - \bar{Z}(u; \beta)\} \frac{W(u, Z_i)}{\hat{K}_C\{u, Z, X(u)\}} \{dN_i(u) - \hat{\lambda}(u; \beta) du \exp(\beta Z_i) Y_i(u)\},$$

$\hat{\lambda}(u; \beta) du$ is estimated using the increment of the Breslow estimate for the underlying cumulative hazard function, i.e., $\hat{\lambda}(u; \beta) du = \sum_i dN_i(u) / \sum_i \{\exp(\beta Z_i) Y_i(u)\}$, $\hat{G}_s\{u, Z_i, X_i(u)\} = \int [dN_{Ci}(u) - Y_i(u) \hat{\lambda}_{0C}(u, Z_i) H\{u, Z_i, X_i(u); \hat{\alpha}\} du] [s\{u, Z_i, X_i(u)\} - \bar{s}(u, Z_i)]$, and $\hat{\lambda}_{0C}(u, Z) du$, for $Z = 0, 1$ are estimated using the increment of the treatment-specific Nelson-Aalen estimator for the cumulative hazard function of the censoring distribution, that is, $\hat{\lambda}_{0C}(u, Z) du = \sum_i dN_{Ci}(u, Z) / \sum_i Y_i(u, Z)$. $\bar{Z}(u; \beta)$ and $\bar{s}(u, Z)$ are calculated as in (2.21) and the estimating equation (2.11). The estimated functions $\hat{f}(\cdot)$ and $\hat{L}(\cdot)$ are given by $f(X_1; \hat{\mathbf{a}})$ and $L\{u, Z, X(u); \hat{\mathbf{b}}\}$, respectively.

Remark 2.1. If the censoring time is independent of the survival time T given treatment assignment Z , then after some algebra, the class of estimating equations described in (2.20) will be identical to the class of estimating equations characterized in (6) of [12].

Remark 2.2. Finding the optimal function of $W(u, Z)$ that gives the efficient estimator in the class of all regular and asymptotic linear estimators by solving estimating equations (2.20) would be very difficult. Therefore our strategy is to choose a function that gives an estimator at least as efficient as the maximum partial likelihood estimator when the censoring time is

noninformative. Here, we choose $W(u, Z) = \hat{K}_C(u, Z)$, where $\hat{K}_C(u, Z)$ is the treatment specific Kaplan-Meier estimator for the censoring time. Clearly, when censoring is noninformative, the first summand of the estimating equation (2.20) will reduce to the standard partial likelihood score function that leads to the maximum partial likelihood estimator for β .

2.4.2 Variance estimator for $\hat{\beta}$

If the potential censoring time truly follows a stratified proportional hazards regression model, $\lambda_C\{u|T \geq u, Z, X(u)\} = \lambda(u|Z)H\{Z, X(u); \alpha\}$, as described in section 1.4, then according to chapter 9.1 of [7], using the estimated function $\hat{\lambda}_C(\cdot)$ and $\hat{K}_C(\cdot)$ in (2.20) will lead to an estimator with its influence function in $\Pi(\Psi_{\text{PH}}^{\text{obs}}|\Lambda_\varphi^\perp)$, where $\Pi(S_1|S_2)$ denotes the projection of space S_1 onto space S_2 , $\Psi_{\text{PH}}^{\text{obs}}$ is defined in (2.16) and $\Lambda_\varphi = \Lambda_{\lambda_{0C}(\cdot, Z)} \oplus \Lambda_\alpha$, in which $\Lambda_{\lambda(\cdot, Z)}$ is the nuisance tangent space associated with the nuisance baseline hazard function $\lambda_{0C}(\cdot, Z)$ which is given in equation (A.20) of Appendix A.3 and Λ_α is the nuisance tangent space associated with the nuisance parameters $\alpha = (\alpha_{11}, \dots, \alpha_{22})$ given in equation (A.21) of Appendix A.3; that is,

Theorem 2.3. *If the hazard function $\lambda_C\{u|T \geq u, Z, X(u)\} = \lambda_{0C}(u, Z)H\{Z, X(u); \alpha\}$, then the class of all observed data influence functions is given by*

$$\Psi = \left[\Pi\{\psi(D)|\Lambda_{\lambda_{0C}(\cdot, Z)}^\perp\} - \Pi(\psi(D)|\Lambda_\alpha) : \psi(D) \in \Psi_{\text{PH}}^{\text{obs}} \right] \quad (2.24)$$

where $\Psi_{\text{PH}}^{\text{obs}}$ is defined in (2.16), $\Pi\{\psi(D)|\Lambda_{\lambda_{0C}(\cdot, Z)}^\perp\}$ is given by

$$\begin{aligned} & C_W^{-1} \int \{Z - Z^*(u; \beta_0)\} \frac{W(u, Z)dM(u, Z; \beta_0)}{K_C\{u, Z, X(u)\}} + (Z - \pi)f(X_1) \\ & + \int dM_C\{u, Z, X(u)\} [L\{u, Z, X(u)\} - L^*(u, Z)] \\ & + \int \frac{E[\psi^F(T, Z)/K_C\{u, Z, X(u)\}H\{Z, X(u)\}Y(u)|Z]}{E[H\{Z, X(u)\}Y(u)|Z]} dM_C\{u, Z, X(u)\} \\ & - \int \frac{E[E\{\psi^F(T, Z)|T \geq u, Z\}/K_C\{u, Z, X(u)\}H\{Z, X(u)\}Y(u)|Z]}{E[H\{Z, X(u)\}Y(u)|Z]} dM_C\{u, Z, X(u)\}, \end{aligned} \quad (2.25)$$

in which C_W and $Z^*(u; \beta)$ are defined in (2.19) and (2.18), respectively,

$$\psi^F(T, Z) = C_W^{-1} \int \{Z - Z^*(u; \beta_0)\} W(u, Z) dM_T(u, Z; \beta_0), \quad (2.26)$$

$$L^*(u, Z) = \frac{E[L\{u, Z, X(u)\}H\{Z, X(u)\}Y(u)|Z]}{E[H\{Z, X(u)\}Y(u)|Z]}, \quad (2.27)$$

and $\Pi(\psi(D)|\Lambda_\alpha)$ is given by (A.24).

Remark 2.3. The proof of Theorem 2.3 is given in Appendix A.6. The expressions for the fourth and fifth summands in (2.25) are very complicated and difficult to evaluate numerically. Therefore we chose to ignore these terms in our variance estimator of the estimator that solves equation (2.20). If the censoring time is noninformative, then these two summands will add to zero giving us a consistent variance estimator, otherwise, our estimator will be slightly conservative. In addition, for the posited model $L\{u, Z, X(u); \mathbf{b}\} = \mathbf{b}^T s\{u, Z, X(u)\}$, if $s\{u, Z, X(u)\}$ contains $\mathbf{Q}(\mathbf{u})$ which is defined by (A.23), then the projection $\Pi(\psi(D)|\Lambda_\alpha)$ is identical to zero. Thus, the a variance estimator for $\hat{\beta}$ is computed by the following sandwich estimator

$$\hat{V}(\hat{\beta}) = \frac{\sum_{i=1}^n \hat{m}_2^2(D_i; \hat{\beta})}{\left[\sum_{i=1}^n \Delta_i W(U_i, Z_i) / \hat{K}_C\{U_i, Z_i, X_i(U)\} \{1 - \bar{Z}(U_i; \beta_0)\} \bar{Z}(U_i; \beta_0) \right]^2}, \quad (2.28)$$

where

$$\begin{aligned} \hat{m}_2(D_i; \beta) &= \int \{Z_i - \bar{Z}(u; \beta)\} \frac{W(u, Z_i)}{\hat{K}_C\{u, Z, X(u)\}} \{dN_i(u) - \hat{\lambda}(u; \beta) du \exp(\beta Z_i) Y_i(u)\} \\ &\quad - (Z_i - \pi) f(X_{1i}) \\ &\quad + \int [dN_{C_i}(u) - Y_i(u) \hat{\lambda}_{0C}(u, Z_i) H\{u, Z_i, X_i(u); \hat{\alpha}\} du] [L(u, Z_i, X_i(u)) - \bar{L}(u, Z_i)]. \end{aligned}$$

2.5 Simulation

We performed a Monte-Carlo simulation study to compare the performance of the maximum partial likelihood estimator $\hat{\beta}_{\text{PH}}$ and our proposed estimators $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ that are obtained by solving the estimating equation (2.20) with $W(u, Z) = \hat{K}_C(u, Z)$ and $\{f(\cdot) = 0, L(\cdot) = 0\}$, $\{f(\cdot) = \hat{f}(\cdot), L(\cdot) = 0\}$ and $\{f(\cdot) = \hat{f}(\cdot), L(\cdot) = \hat{L}(\cdot)\}$, respectively. For this

study, we considered one baseline covariate X_1 and one posttreatment covariate X_2 that can be obtained immediately for all patients. In order to ensure that the covariate X_1 was independent of treatment assignment Z , was correlated to the survival time T , and that the conditional distribution of T given Z followed a proportional hazards relationship, we generated the data in the following manner. First we generated bivariate data (Y, X) from a bivariate normal density with mean zero, variance 1, and correlation ρ . We then independently generated the treatment indicator Z as a $\text{Bernoulli}(\pi)$. Using inverse transformation, the survival time T was taken to be $T = -\exp(-\beta Z) \log\{1 - \Phi(Y)\}$, where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal. $X_1 = \Phi^{-1}(X)$ follows a uniform $(0, 1)$ distribution. This guarantees that the distribution of T given Z will follow a proportional hazards relationship $\lambda(t|z) = \lambda(t) \exp(\beta z)$, with $\lambda(t) = 1$, that is, $T \sim \text{Exp}\{\exp(\beta Z)\}$. The posttreatment covariate was generated using the formula $X_2 = rT + \sqrt{1 - r^2}/\sqrt{2}\chi_1^2$ which results in the correlation of X_2 and T to be r . Censoring time C was generated using the following three scenarios: 1) Exponential distribution with hazard rate $\lambda_C(u|Z) = c \exp(\beta Z)$ reflecting noninformative censoring given Z ; 2) Exponential distribution with hazard rate $\lambda_C(u|X_1, X_2, Z) = c\{Z \exp(\alpha_{11}X_1 + \alpha_{21}X_2) + (1 - Z) \exp(\alpha_{12}X_1 + \alpha_{22}X_2)\}$, reflecting informative censoring with a stratified proportional hazard regression model; 3) Scaled Lognormal distribution $c \cdot \text{LN}(\mu, \sigma)$ with $\mu = Z(\ell_{11}X_1 + \ell_{21}X_2) + (1 - Z)(\ell_{12}X_1 + \ell_{22}X_2)$ and $\sigma = 1$, in which case the censoring time is dependent on the survival time given Z but does not follow a stratified proportional hazard regression model. To calculate the estimated functions $\hat{f}(\cdot)$ and $\hat{L}(\cdot)$, we posited the models $a_0 + a_1X_1 + a_2X_1^2$ for $f(\cdot)$, and $b_0 + b_1X_1 + b_2X_2 + b_3X_1Z + b_4X_2Z$ for $L(\cdot)$.

For this demonstration, treatment was assigned with probability $\pi = .5$, the correlation between the bivariate normal random variable was taken to be $\rho = .5$ which resulted in a sample correlation of approximately 0.4 between the survival time T and baseline covariate X_1 . The correlation of X_2 and T was taken to be $r = 0.7$. Two values for the proportional hazards regression coefficient were considered, $\beta = 0$ (null hypothesis) and $\beta = .3$. For the censoring time, $\alpha_{11} = 1$, $\alpha_{21} = 0.1$, $\alpha_{12} = 2$, $\alpha_{22} = 0.3$, $\ell_{11} = 0.3$, $\ell_{21} = 1$, $\ell_{12} = 1$, $\ell_{22} = 0.3$, and the value c was chosen in different scenarios that would result in roughly 36% the data being censored. Sample sizes of 250 and 600 were considered and each scenario used 1000 Monte-Carlo simulations. In Tables 2.1, 2.2 and 2.3, we compare the bias, standard error estimate, Monte-

Table 2.1: Simulation Results for the 1st Scenario of Censoring (*1000 Monte-Carlo samples, entries in parentheses for SE and Monte-Carlo SE are relative efficiencies with respect to $\hat{\beta}_1$.*)

True β_0	Statistics	n	$\hat{\beta}_{PH}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0	Bias	250	0.004	0.004	0.005	0.005
		600	0.001	0.000	0.001	0.001
	AveSE	250	0.160 (1.00)	0.160 (1.00)	0.145 (1.21)	0.134 (1.42)
		600	0.089 (1.00)	0.089 (1.00)	0.080 (1.24)	0.077 (1.34)
	MCSE	250	0.165 (0.88)	0.155 (1.00)	0.141 (1.21)	0.141 (1.21)
		600	0.091 (0.94)	0.088 (1.00)	0.078 (1.27)	0.078 (1.27)
	Type I Error	250	0.055	0.048	0.041	0.060
		600	0.054	0.043	0.047	0.059
	Bias	250	0.007	0.007	0.008	0.008
		600	0.002	0.003	0.003	0.003
0.3	AveSE	250	0.162 (1.00)	0.161 (1.00)	0.147 (1.21)	0.137 (1.38)
		600	0.103 (1.00)	0.014 (1.00)	0.094 (1.21)	0.088 (1.37)
	MCSE	250	0.165 (0.90)	0.157 (1.00)	0.143 (1.20)	0.143 (1.20)
		600	0.107 (0.91)	0.102 (1.00)	0.091 (1.26)	0.091 (1.26)
	Power	250	0.481	0.474	0.555	0.606
		600	0.826	0.831	0.899	0.923

Carlo standard error, relative efficiency (ratio of variance estimate and ratio of Monte-Carlo variance), type I error and the power of the maximum partial likelihood estimator $\hat{\beta}_{PH}$ and our proposed estimators $\hat{\beta}_k$, $k = 1, 2, 3$, for the various simulation scenarios.

Table 2.1 shows the simulation results under scenario 1 where censoring time is non-informative. As we expect, all the estimators are unbiased and control the type I error. Our proposed estimators are more efficient than the traditional maximum partial likelihood estimator and more powerful than the logrank test.

Table 2.2 shows the simulation results under scenario 2 where censoring time follows a stratified proportional hazards regression model for each treatment group. As we can see the traditional maximum partial likelihood estimator is severely biased whereas all our proposed estimators are unbiased and control the type I error. This is consistent with the theoretical results in this paper.

In addition, from Table 2.3, we see that when the censoring time is informative, but does not follow a stratified proportional hazard regression model, our proposed estimators are still less biased than the traditional maximum partial likelihood estimator.

Table 2.2: Simulation Results for the 2nd Scenario of Censoring (*1000 Monte-Carlo samples, entries in parentheses for SE and Monte-Carlo SE are relative efficiencies with respect to $\hat{\beta}_1$.*)

True β_0	Statistics	n	$\hat{\beta}_{PH}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0	Bias	250	0.159	0.014	0.015	0.015
		600	0.159	0.008	0.008	0.008
	AveSE	250	0.183 (1.07)	0.189 (1.00)	0.174 (1.18)	0.161 (1.38)
		600	0.117 (1.10)	0.123 (1.00)	0.114 (1.17)	0.105 (1.38)
	MCSE	250	0.185 (1.02)	0.187 (1.00)	0.171 (1.20)	0.171 (1.20)
		600	0.121 (1.02)	0.122 (1.00)	0.110 (1.22)	0.110 (1.22)
	Type I Error	250	0.141	0.053	0.049	0.069
		600	0.289	0.043	0.039	0.064
0.3	Bias	250	0.152	0.019	0.020	0.020
		600	0.149	0.010	0.010	0.010
	AveSE	250	0.185 (1.05)	0.190 (1.00)	0.174 (1.18)	0.163 (1.35)
		600	0.119 (1.08)	0.123 (1.00)	0.113 (1.17)	0.106 (1.34)
	MCSE	250	0.188 (1.01)	0.189 (1.00)	0.173 (1.20)	0.173 (1.20)
		600	0.121 (1.04)	0.123 (1.00)	1.112 (1.21)	1.112 (1.21)
	Power	250	0.684	0.397	0.450	0.493
		600	0.962	0.707	0.769	0.803

Table 2.3: Simulation Results for the 3rd Scenario of Censoring (*1000 Monte-Carlo samples, entries in parentheses for SE and Monte-Carlo SE are relative efficiencies with respect to $\hat{\beta}_1$.*)

True β_0	Statistics	n	$\hat{\beta}_{PH}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0	Bias	250	-0.245	0.023	0.019	0.019
		600	-0.254	0.051	0.052	0.052
	AveSE	250	0.193 (1.37)	0.225 (1.00)	0.210 (1.15)	0.183 (1.51)
		600	0.123 (1.86)	0.168 (1.00)	0.160 (1.11)	0.133 (1.60)
	MCSE	250	0.191 (1.43)	0.229 (1.00)	0.215 (1.13)	0.215 (1.13)
		600	0.123 (1.89)	0.169 (1.00)	0.164 (1.06)	0.164 (1.06)
	Type I Error	250	0.247	0.045	0.041	0.071
		600	0.527	0.040	0.037	0.088
0.3	Bias	250	-0.230	0.023	0.019	0.019
		600	-0.240	0.046	0.048	0.048
	AveSE	250	0.185 (1.38)	0.218 (1.00)	0.203 (1.15)	0.177 (1.51)
		600	0.118 (1.87)	0.162 (1.00)	0.154 (1.11)	0.128 (1.60)
	MCSE	250	0.182 (1.42)	0.216 (1.00)	0.202 (1.15)	0.202 (1.15)
		600	0.117 (1.83)	0.158 (1.00)	0.154 (1.06)	0.154 (1.06)
	Power	250	0.059	0.297	0.319	0.450
		600	0.067	0.617	0.666	0.781

2.6 A brief example

We also applied this methodology to a subset of the data from 2139 patients from the AIDS Clinical Trials Group (ACTG) protocol 175 ([9]), a study that randomizes patients to four antiretroviral regimes in equal proportions. The main focus of this study was two-sample comparisons of treatment 0 (Zidovudine) versus treatment 1 (Zidovudine and didanosine), treatment 2 (Zidovudine and zalcitabine) and treatment 3 (Didanosine), respectively. 532 patients were randomized to treatment 0, 522 were randomized to treatment 1, 524 were randomized to treatment 2 and 561 were randomized to treatment 3. The primary endpoint was a ≥ 50 percent decline in the CD4 cell count, an event indicating progression to the acquired immunodeficiency syndrome (AIDS), or death. Roughly 76% of the data were censored.

Fig.1.1 is a plot of the logarithm of the negative logarithm of the survival distribution for the time to death for each treatment. The four lines, except for a few points early in time, are approximately parallel suggesting that a proportional hazards relationship between treatments is reasonable. If the censoring time is noninformative, we know that the maximum partial likelihood estimator is the most efficient estimator and the logrank test is the most powerful nonparametric test without using any additional covariates. The results of applying the standard analysis using Cox's maximum partial likelihood estimator can be found in Table 2.4. For example, the estimate of the log hazard ratio between treatment 0 and treatment 1 is -0.703 and its standard error is $.124$, which is highly statistically significant.

However, in addition to the censored survival times and treatment arms, the data also contains several prognostic baseline auxiliary covariates and post-treatment covariates. The baseline covariates include CD4, CD8, age (years), weight (kg), gender (0 = female), hemophilia indicator (0 = no), homosexual activity (0 = no), race (0 = white, 1 = non-white), history of IV drug use (0 = no), Karnofsky score (on a scale of 0-100), Zidovudine in the 30 days prior to 175 (0 = no), antiretroviral history (1 = 'Antiretroviral Naive', 0 = other), number of days pre-175 antiretroviral therapy and symptomatic status indicator (0 = asymptomatic). The post-treatment covariates include, CD4 at 20 ± 5 weeks, CD8 at 20 ± 5 weeks, CD4 at 96 ± 5 weeks ($= -1$ if missing), indicator of off-trt before 96 ± 5 weeks (0 = no, 1 = yes) and Missing CD4 at 96 ± 5 weeks (0 = missing, 1 = observed). Here, the post-treatment covariate, CD4 at

Table 2.4: Estimates of $\hat{\beta}_{\text{PH}}$ and $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ on the ACTG 175 data (*RE is the relative efficiencies with respect to $\hat{\beta}_{\text{PH}}$.*)

		Estimates	Standard Errors	RE
Treatment 0 and 1	$\hat{\beta}_{\text{PH}}$	-0.703	0.124	1.01
	$\hat{\beta}_1$	-0.689	0.124	1.00
	$\hat{\beta}_2$	-0.724	0.120	1.07
	$\hat{\beta}_3$	-0.721	0.117	1.12
Treatment 0 and 2	$\hat{\beta}_{\text{PH}}$	-0.640	0.121	1.01
	$\hat{\beta}_1$	-0.638	0.122	1.00
	$\hat{\beta}_2$	-0.617	0.114	1.15
	$\hat{\beta}_3$	-0.590	0.111	1.21
Treatment 0 and 3	$\hat{\beta}_{\text{PH}}$	-0.528	0.116	1.01
	$\hat{\beta}_1$	-0.525	0.116	1.00
	$\hat{\beta}_2$	-0.536	0.111	1.10
	$\hat{\beta}_3$	-0.509	0.109	1.14

96 ± 5 weeks, is a time dependent covariate, and it was missing for some patients because they were not at risk at this particular time, 96 ± 5 weeks.

Applying a stratified proportional hazard regression model for the censoring time and using Forward selection with $\text{sle}=0.05$, we obtained some prognostic covariates for each treatment group. The important prognostic covariates for the censoring time are (AGE, RACE, STRAT, OFFTTRT, MisCD4) for treatment 0, (HOMO, Z30, RACE, CD820, OFFTTRT, MisCD4) for treatment 1, (HOMO, RACE, STRAT, SYMPTOM, OFFTTRT) for treatment 2, and (Z30, GENDER, CD80, OFFTTRT, MisCD4) for treatment 3.

We also applied the similar model for the survival time and found some prognostic covariates for each treatment group, which are (CD40, CD80, GENDER, Z30, CD420, CD496, MisCD4) for treatment 0, (CD40, PREANTI, STRAT, SYMPTOM, CD420, CD496, MisCD4) for treatment 1, (PREANTI, KARNOFSKY, SYMPTOM, CD420, CD496, MisCD4) for treatment 2 and (CD40, SYMPTOM, CD420, CD496, MisCD4) for treatment 3.

As we can see, for each treatment group, there is only one common covariate that is prognostic for both survival time and censoring time. For example, for treatment 0, 1 and 3, MisCD4 is the only common covariate and for treatment 2, SYMPTOM is the common covariate and both MisCD4 and SYMPTOM are binary variables. This gave us a rough sense that the survival time and censoring time is weakly correlated given treatment assignment. And hence

our proposed estimator may be close to the traditional partial likelihood estimator which assumes independence between survival time and censoring time given treatment, which explains why there are no substantial differences between our estimators and the maximum partial likelihood estimator after applying our method to the ACTG 175 data as shown in Table 2.4. Again, to obtain a more efficient estimator, we incorporated some prognostic covariates into the model that include baseline covariates CD4, CD8, AGE, WEIGHT, DRUGS, KARNOFSKY, Z30, SYMPTOM and PREANTI, and post-treatment covariates CD420, CD820, CD496, MisCD4 and OFFTRT. From the results we can see that using auxiliary covariates in the model leads to more efficient estimators ($\hat{\beta}_2$ and $\hat{\beta}_3$) than the estimator ($\hat{\beta}_1$) that does not consider auxiliary covariates in the model.

Chapter 3

Conclusion

It is very common, in clinical trials, that the time-lagged response for some patients is missing because of censoring. If not accounted for correctly, such missing data will lead to biased inferences. In addition, because of limited resources, either money or patients, it is very important that we take advantage of available data to make as much as efficient inferences as is possible. In the first chapter, we discuss the special case where the censoring time is noninformative and the survival time itself is the time-lagged response. Using semiparametric theory, we derive an augmented class of regular and asymptotic linear estimators for the treatment specific log hazard ratio parameter. Auxiliary covariates that are prognostic for the survival time are used to derived estimators that are more efficient then the traditional maximum partial likelihood estimator and the corresponding induced Wald tests are more powerful than the logrank test.

In the second chapter, we allow for censoring that is informative in a manner that can be explained through the observed auxiliary covariates. A general time-lagged response that is not just limited to the survival time are considered. By applying the theory of augmented inverse probability complete case estimators derived by Robins and Rotnitzky ([13]), we characterized a class or regular and asymptotic linear estimators for the treatment effect parameter. The special case where the survival time given treatment follows a proportional hazard model is also discussed. To obtain an unbiased estimator, we have to correctly estimate the survival and hazard functions for the censoring time. Clearly, if the censoring is informative and if our

posited models for censoring time contains the truth, our proposed estimators are unbiased, but the maximum partial likelihood estimator is severely biased for the treatment specific log hazard ratio. Even if we cannot correctly estimate the distribution of censoring time, the simulation results show that our proposed estimators are much less biased than the maximum partial likelihood estimator. Again, incorporating the auxiliary covariates into the model helps us to obtain estimators that are more efficient than the estimators without using auxiliary covariates.

Bibliography

- [1] N. Mantel. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Report*, 50:163–170, 1966.
- [2] R. Peto and J. Peto. Asymptotically efficient rank invariant procedures. *Journal of the Royal Statistical Society A*, 135:185–207, 1972.
- [3] D. R. Cox. Regression models and life tables. *Journal of the Royal Statistical Society B*, 34:187–220, 1972.
- [4] J. M. Begun, W. J. Hall, W. M. Huang, and J. A. Wellner. Information and asymptotic efficiency in parametric-nonparametric models. *Annals of Statistics*, 11:432–452, 1983.
- [5] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. The Johns Hopkins University Press, Baltimore, MD, USA, 1993.
- [6] D. R. Cox. Partial likelihood. *Biometrika*, 62:269–275, 1975.
- [7] A.A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer, New York, NY, USA, 2006.
- [8] W. K. Newey. Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5:99–135, 1990.
- [9] Hammer S. M., Katzenstein D. A., Hughes M. D., Gundaker H., Schooley R. T., haubrich R. H., Henry W. K., Lederman M. M., Phair J. P., Niu M., Hirsch M. S., and for the AIDS Clinical Trials Group Study 175 Study Team Merigan T. C. A trial comparing nucleoside

- monotherapy with combination therapy in hiv-infected adults wit hcd4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, 335:1081–1089, 1996.
- [10] K. J. Anstrom and A. A. Tsiatis. Using propensity scores to estimate causal treatment effects wit hcensored time-lagged data. *Biometrics*, 57:1207–1218, 2001.
 - [11] Y. Huang and T. A. Louis. Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika*, 85:785–798, 1998.
 - [12] X. Lu and A. A. Tsiatis. Improving the efficiency of the logrank test using auxiliary covariates. In Revision *Biometrika*, 2007.
 - [13] J. M. Robins and A. Rotnitzky. Recovery of information and adjustment for dependent censoring using surrogate markers. In N. Jewell, K. Dietz, and V. Farewell, editors, *AIDS Epidemiology-Methodological Issues*, pages 297–331. Birkhauser, 1992.
 - [14] D. B. Rubin. *Multiple imputation for nonresponse in surveys*. Wiley, New York, NY, USA, 1987.
 - [15] M. Zhang, A. A. Tsiatis, and M. Davidian. Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. Submitted to *Biometrics*, 2007.
 - [16] N. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30:89–99, 1974.
 - [17] T. R. Fleming and D. P. Harrington. *Counting Processes and Survival Analysis*. John Wiley and Sons, Newyork, NY, USA, 1991.

Appendices

A.1 Sketch proof of Theorem 1.1

The likelihood of the data is given by

$$\begin{aligned} & p(u, \delta, z, x | \beta_0, \eta_1, \eta_2, \eta_3, \eta_4) \\ &= \{p_{(T,X)|Z}(u, x | z; \beta_0, \eta_1, \eta_3, \eta_4)\}^\delta \{S_{(T,X)|Z}(u, x | z; \beta_0, \eta_1, \eta_3, \eta_4)\}^{1-\delta} \end{aligned} \quad (\text{A.1})$$

$$\begin{aligned} & \times \{\lambda_{C|Z}(u | z; \eta_2)\}^{1-\delta} \exp\{-L_{C|Z}(u | z; \eta_2)\} \\ & \times p(z) \end{aligned} \quad (\text{A.2})$$

where $S_{(T,X)|Z}(u, x | z) = \int_u^\infty p_{(T,X)|Z}(v, x | z) dv$, $\lambda_{C|Z}(\cdot)$ denotes the conditional hazard rate of C given Z , $L_{C|Z}(u) = \int_0^u \lambda_{C|Z}(v | z) dv$ denotes the corresponding cumulative hazard function, and $p(z) = \pi z + (1 - \pi)(1 - z)$.

Since the nuisance parameters $\eta = (\eta_1, \dots, \eta_4)$ must satisfy the constraints imposed by (1.1), (1.3) and (1.4), after a little algebra, we have that all the nuisance parameters must satisfy the following constraints (A.3) and (A.4),

$$\begin{aligned} & \int \{p_{(T,X)|Z}(u, x | z; \beta_0, \eta_1, \eta_3, \eta_4)\}^\delta \{S_{(T,X)|Z}(u, x | z; \beta_0, \eta_1, \eta_3, \eta_4)\}^{1-\delta} dx \\ &= \{\lambda(u; \eta_1) \exp(\beta_0 z)\}^\delta \exp\{-L(u; \eta_1) \exp(\beta_0 z)\}, \end{aligned} \quad (\text{A.3})$$

and that

$$\sum_{\delta=0}^1 \int p(u, \delta, z, x | \beta_0, \eta_1, \eta_2, \eta_3, \eta_4) du = p_{X|Z}(x | z; \eta_3). \quad (\text{A.4})$$

Following the steps outlined in §5.2 of [7], we derive the nuisance tangent space Λ_2 associated with the parameter η_2 . For this censored survival data, it is convenient to use the counting process and associated martingale process notation defined by Fleming and Harrington ([17]). Specifically,

$$\Lambda_2 = \left\{ \int a_2(u, Z) dM_C(u, Z) \text{ for all functions } a_2(u, z) \right\}, \quad (\text{A.5})$$

where $dM_C(u, Z)$ is the martingale increment $dN_C(u) - \lambda_C(u | Z)Y(u)du$.

To derive the nuisance tangent space Λ_1 that is associated with the parameter η_1 , we need to consider any parametric submodel for η_1 in terms of finite dimensional parameters γ_1 which satisfies

$$\begin{aligned} & \int \{p_{(T,X)|Z}(u, x|z; \beta_0, \gamma_1)\}^\delta \{S_{(T,X)|Z}(u, x|z; \gamma_1)\}^{1-\delta} dx \\ &= \{\lambda(u; \gamma_1) \exp(\beta_0 z)\}^\delta \exp\{-L(u; \gamma_1) \exp(\beta_0 z)\}, \end{aligned} \quad (\text{A.6})$$

and

$$\sum_{\delta=0}^1 \int p(u, \delta, z, x | \beta_0, \gamma_1) du = p_{X|Z}(x|z). \quad (\text{A.7})$$

The score vector with respect to γ_1 is obtained by finding the derivative of the log-likelihood in (A.1) with respect to γ_1 , which can be written as $S_{\gamma_1}(U, \Delta, Z, X) = \Delta b_1(U, Z, X) + (1 - \Delta) \int_U^\infty b_1(v, Z, X) p_{(T,X)|Z}(v, X|Z; \gamma_{10}) dv / S_{(T,X)|Z}(U, X|Z; \gamma_{10})$, where $b_1(u, z, x) = \partial \log p_{(T,X)|Z}(u, x|z; \beta_0, \gamma_1) / \partial \gamma_1|_{\gamma_1=\gamma_{10}}$ and γ_{10} is the true value for γ_1 . If we let $a_1(u, Z, X) = b_1(u, Z, X) - \int_u^\infty b_1(v, Z, X) p_{(T,X)|Z}(v, X|Z; \gamma_{10}) dv / S_{(T,X)|Z}(u, X|Z; \gamma_{10})$, after some algebra, we can have $b_1(u, Z, X) = a_1(u, Z, X) - \int_0^u a_1(u, Z, X) \lambda_{T|(Z,X)}(v|Z, X; \gamma_{10}) dv$, and the score vector can be written as a stochastic integral of a martingale counting process; Fleming and Harrington ([17]), $S_{\gamma_1}(U, \Delta, Z, X) = \int a_1(u, Z, X) dM(u, Z, X)$. If we now take the derivative with respect to γ_1 of the logarithm of both sides of equation (A.6) and (A.7), then, after some algebra, we obtain that

$$E \left\{ \int a_1(u, Z, X) dM(u, Z, X) \middle| U, \Delta, Z \right\} = \int a(u) dM(u, Z), \quad (\text{A.8})$$

where $a(u) = \partial \log \lambda(u; \gamma_1) / \partial \gamma_1|_{\gamma_1=\gamma_{10}}$, and that

$$E \left\{ \int a_1(u, Z, X) dM(u, Z, X) \middle| Z, X \right\} = 0. \quad (\text{A.9})$$

From this relationship, we see that any element of the tangent space Λ_1 must satisfy (A.8) and (A.9). Note that (A.9) is automatically true for any function $\int a_1(u, Z, X) dM(u, Z, X)$. Therefore, we have the result in (1.12).

To derive (1.13), it is enough to prove that for any function $\lambda(u; \gamma_1)$, there exists a parametric submodel $p_{(T,X)|Z}(t, x|z; \gamma_1)$ satisfies the conditions in (A.6) and (A.7).

Let $F(t|z; \gamma_1)$ be the conditional cumulative distribution function of T given Z and $F_{X|Z}(x|z)$ be the underlying true cumulative distribution function of X given Z , that is, $F(t|z; \gamma_1) = \int_0^t \lambda(u; \gamma_1) \exp(\beta_0 z) \exp\{-L(u; \gamma_1) \exp(\beta_0 z)\} du$ and $F_{X|Z}(x|z) = \int_{-\infty}^x p_{X|Z}(x'|z) dx'$. Let $F_{(T,X)|Z}(t, x|z)$ and $F_{T|Z}(t|z)$ be the underlying true cumulative distribution function of (T, X) given Z and T given Z , respectively. It can be easily prove that

$$F_{(T,X)|Z} \left[F_{T|Z}^{-1} \{ F(t|z; \gamma_1) | z \}, x | z \right]$$

is the cumulative distribution function of the parametric submodel we need to find.

In a similar way, we can derive the form of the nuisance tangent space Λ_3 as given by (1.14) and (1.15), and the form of the nuisance tangent space Λ_4 as given by (1.16).

A.2 Sketch proof of Theorem 1.2

To prove Theorem 1.2, we need to use the following results,

1. If S_1 and S_2 are two linear subspaces in the Hilbert space \mathcal{H} , then $(S_1 + S_2)^\perp = S_1^\perp \cap S_2^\perp$.
2. The entire Hilbert space \mathcal{H} , that is, the set of all mean zero functions of $D = (U, \Delta, Z, X)$, can be partitioned as a direct sum of mutually orthogonal linear subspaces, namely, $\mathcal{H} = H_1 \oplus H_2 \oplus H_3$, where

$$H_1 = \left\{ \int a(u, Z, X) dM(u, Z, X) \text{ for all functions } a(u, z, x) \right\},$$

$$H_2 = \left\{ \int a(u, Z, X) dM_C(u, Z) \text{ for all functions } a(u, z, x) \right\}, \text{ and}$$

$$H_3 = \left[a(Z, X_1, X_2) : E\{a(Z, X_1, X_2)\} = 0 \right].$$
3. The set of all functions of $a(U, \Delta, Z)$ such that $E\{a(U, \Delta, Z)\} = 0$ is equal to a direct sum of mutually orthogonal linear subspaces, namely, $H_1^* \oplus \Lambda_2 \oplus H_3^*$, where Λ_2 was defined by (1.17), $H_1^* = \{ \int a(u, Z) dM(u, Z) \text{ for any function } a(u, z) \}$, and $H_3^* = [a(Z) : E\{a(Z)\} = 0]$.

From result (a), we deduce that $\Lambda^\perp = \Lambda_1^\perp \cap \Lambda_2^\perp \cap \Lambda_3^\perp \cap \Lambda_4^\perp$. We begin by deriving Λ_4^\perp . The linear subspace Λ_4 , given in (1.16) can also be written as $\Lambda_4 = Q_0 \cap H_1$, where $Q_0 = [a(D) : E\{a(D)|U, \Delta, Z\} = 0]$. Use the result (a) again, we have $\Lambda_4^\perp = Q_0^\perp + H_1^\perp$. The linear space Q_0^\perp can easily be shown to be the space of all mean zero functions of (U, Δ, Z) , which, by result (c), equals $H_1^* \oplus \Lambda_2 \oplus H_3^*$, and the linear space H_1^\perp can be shown by result (b) to be equal to $H_2 \oplus H_3$. Since $\Lambda_2 \subset H_2$ and $H_3^* \subset H_3$, we have that $\Lambda_4^\perp = H_1^* + (H_2 \oplus H_3)$.

Now we deduce $\Lambda_1^\perp \cap \Lambda_4^\perp$, which consists of all elements $h_1^* + h_2 + h_3$ that are orthogonal to Λ_1 , where $h_1^* = \int a(u, Z)dM(u, Z) \in H_1^*$, $h_2 \in H_2$ and $h_3 \in H_3$, that is

$$E\left[\left\{\int a(u, Z)dM(u, Z) + h_2 + h_3\right\}a^*(D)\right] = 0, \quad (\text{A.10})$$

for any function $a^*(D) \in \Lambda_1$. Since $a^*(D) \in H_1$ and $h_2, h_3 \in H_1^\perp$, (A.10) reduces to

$$E\left[\left\{\int a(u, Z)dM(u, Z)\right\}a^*(D)\right] = 0. \quad (\text{A.11})$$

Using iterated conditional expectations, we get that

$$\begin{aligned} & E\left[\int a(u, Z)dM(u, Z)E\left\{a^*(D)|U, \Delta, Z\right\}\right] \\ &= E\left\{\int a(u, Z)dM(u, Z) \int a^*(u)dM(u, Z)\right\} \\ &= E\left\{\int a(u, Z)a^*(u)\lambda_0(u)\exp(\beta_0 Z)Y(u)du\right\} = 0 \text{ for all } a^*(u), \end{aligned} \quad (\text{A.12})$$

where (A.12) follows because the covariance of two martingale processes is equal to the expectation of the predictable covariation process; see Fleming and Harrington ([17]). In order for (A.12) to hold, the function $a(u, Z)$ must be centered appropriately. As a consequence, we obtain that $\Lambda_1^\perp = \mathcal{E} + (H_2 \oplus H_3)$, where

$$\mathcal{E} = \left(\int \left[a(u, Z) - \frac{E\{a(u, Z)\exp(\beta_0 Z)Y(u)\}}{E\{\exp(\beta_0 Z)Y(u)\}}\right]dM(u, Z) \text{ for all } a(u, Z)\right). \quad (\text{A.13})$$

Continuing with our argument: because $\Lambda_2 \subset H_2$, and because \mathcal{E} is orthogonal to Λ_2 , we deduce that

$$\Lambda_4^\perp \cap \Lambda_1^\perp \cap \Lambda_2^\perp = \mathcal{E} + \{(H_2 \cap \Lambda_2^\perp) \oplus H_3\}, \quad (\text{A.14})$$

where we denote the space $(H_2 \cap \Lambda_2^\perp)$ by \mathcal{C} , consisting of all elements $\int g(u, Z, X) dM_C(u, Z)$ that are orthogonal to all elements $\int g^*(u, Z) dM_C(u, Z) \in \Lambda_2$, that is,

$$E \left\{ \int g(u, Z, X) dM_C(u, Z) \int g^*(u, Z) dM_C(u, Z) \right\} = 0 \text{ for all } g^*(u, Z),$$

or equivalently, using results regarding the covariance of martingales, we obtain

$$E \left\{ \int g(u, Z, X) g^*(u, Z) \lambda_{C|Z}(u, Z) Y(u) du \right\} = 0 \text{ for all } g^*(u, Z).$$

After a little algebra, we deduce that

$$\mathcal{C} = \left(\int \left[g(u, Z, X) - E\{g(u, Z, X) | T \geq u, Z\} \right] dM_C(u, Z) \text{ for all } g(u, x, z) \right). \quad (\text{A.15})$$

We next consider $\Lambda_4^\perp \cap \Lambda_1^\perp \cap \Lambda_2^\perp \cap \Lambda_3^\perp$, which consists of all elements $e + c + h_3$ that are orthogonal to all $l_3 \in \Lambda_3$, where $e \in \mathcal{E}$, $c \in \mathcal{C}$ and $h_3 \in H_3$. Standard conditioning arguments can now be used to show \mathcal{E} is orthogonal to Λ_3 , by virtue that elements in \mathcal{E} are functions only of (U, Δ, Z) . Because $\Lambda_3 \subset H_1$ and $\mathcal{C} \subset H_2$ and because $H_1 \perp H_2$, this implies that $\mathcal{C} \perp \Lambda_3$. Therefore, we deduce that

$$\Lambda_4^\perp \cap \Lambda_1^\perp \cap \Lambda_2^\perp \cap \Lambda_3^\perp = \mathcal{E} + \{\mathcal{C} \oplus (H_3 \cap \Lambda_3^\perp)\}$$

where we denote $(H_3 \cap \Lambda_3^\perp)$ as \mathcal{R} , consisting of all elements $h_3(Z, X_1, X_2) \in H_3$ that are orthogonal to all elements in Λ_3 , that is,

$$E \left\{ h_3(Z, X_1, X_2) a_3(D) \right\} = 0 \text{ for all } a_3(D) \in \Lambda_3,$$

or equivalently, using iterated conditional expectations, we obtain that

$$E \left[h_3(Z, X_1, X_2) \{a_{31}(Z, X_1, X_2) + a_{32}(X_1)\} \right] = 0,$$

for all functions $a_{31}(Z, X_1, X_2)$ and $a_{32}(X_1)$ such that $E\{a_{31}(Z, X_1, X_2)|Z, X_1\} = 0$ and $E\{a_{32}(X_1)\} = 0$. After a little algebra, we deduce that

$$\mathcal{R} = \left[a(Z, X_1) : E\{a(Z, X_1)|X_1\} = 0 \right].$$

To derive \mathcal{R} , we note that any function of Z and X_1 can be written as $a(Z, X_1) = Za_1(X_1) + (1 - Z)a_2(X_1)$ for arbitrary functions $a_1(X_1)$ and $a_2(X_1)$, and $E\{a(Z, X_1)|X_1\} = \pi a_1(X_1) + (1 - \pi)a_2(X_1)$ equaling zero implies that $a_2(X_1) = -a_1(X_1)\pi/(1 - \pi)$, which, in turn implies, that $a(Z, X_1) = (Z - \pi)a(X_1)$ for an arbitrary function $a(X_1)$. Consequently,

$$\mathcal{R} = \left\{ (Z - \pi)f(X_1) \text{ for any function } f(X_1) \right\}. \quad (\text{A.16})$$

A.3 Sketch proof of Theorem 1.3

We first prove the following Lemma.

Lemma A.1. *For any function $b(u, Z)$, we have the following equality*

$$\begin{aligned} & \frac{\Delta \int b(u, Z) dM_T(u, Z)}{K_C(U, Z)} + \int \frac{dM_C(u, Z)}{K_C(u, Z)} E \left\{ \int b(u, Z) dM_T(u, Z) | T \geq u, Z \right\} \\ &= \int \frac{b(v, Z)}{K_c(v, Z)} dM(v, Z), \end{aligned}$$

where $dM_T(u, Z) = I(T = u) - \lambda_0(u) \exp(\beta_0 Z) I(T \geq u) du$.

Proof. Since $E\{I(T = v)|T \geq u, Z\} = I(v \geq u)e^{\beta_0 Z} \lambda_0(v) dv P(T \geq v, Z)/P(T \geq u, Z)$ and $E\{I(T \geq v)|T \geq u, Z\} = I(v < u) + I(v \geq u)P(T \geq v, Z)/P(T \geq u, Z)$, we have that

$$E\{dM_T(v, Z)|T \geq u, Z\} = -I(v < u)e^{\beta_0 Z} \lambda_0(v) dv.$$

This implies

$$\begin{aligned}
& \int \frac{dM_C(u, Z)}{K_C(u, Z)} E \left\{ \int b(u, Z) dM_T(u, Z) \middle| T \geq u, Z \right\} \\
&= \frac{\Delta - 1}{K_C(U, Z)} \int_0^U b(v, Z) e^{\beta_0 Z} \lambda_0(v) dv + \int_0^U \frac{\lambda_C(u, Z)}{K_C(u, Z)} du \int_0^u b(v, Z) e^{\beta_0 Z} \lambda_0(v) dv \\
&= \frac{\Delta}{K_C(U, Z)} \int b(v, Z) I(v \leq U) e^{\beta_0 Z} \lambda_0(v) dv - \int \frac{b(v, Z)}{K_c(v, Z)} I(v \leq U) e^{\beta_0 Z} \lambda_0(v) dv
\end{aligned}$$

The last equality is because that $\int_v^U \frac{\lambda_C(u, Z)}{K_C(u, Z)} du = \frac{1}{K_c(U, Z)} - \frac{1}{K_c(v, Z)}$. Hence, we have

$$\begin{aligned}
& \frac{\Delta \int b(u, Z) dM_T(u, Z)}{K_C(U, Z)} + \int \frac{dM_C(u, Z)}{K_C(u, Z)} E \left\{ \int b(u, Z) dM_T(u, Z) \middle| T \geq u, Z \right\} \\
&= \frac{\Delta b(U, Z)}{K_C(U, Z)} - \int \frac{b(v, Z)}{K_c(v, Z)} I(v \leq U) e^{\beta_0 Z} \lambda_0(v) dv \\
&= \int \frac{b(v, Z)}{K_C(v, Z)} dM(v, Z), \text{ for all functions } b(v, Z).
\end{aligned}$$

□

Now we begin to prove Theorem 1.3. From the result of semiparametric theory ([7], chapter 4), we know that find the efficient influence function within the subclass is equivalent to find the projection of $\int \{a(u, Z) - a^*(u)\} dM(u, Z)$ onto the linear subspace

$$\left\{ (Z - \pi) f(X_1) + \int dM_C(u, Z) \left[g(u, Z, X) - E\{g(u, Z, X) | T \geq u, Z\} \right] : \text{for all } f(\cdot) \text{ and } g(\cdot) \right\},$$

which is equivalent to find $f_0(X_1)$ and $g_0(u, Z, X)$ such that

$$\begin{aligned}
0 &= E \left\{ \left(\int \{a(u, Z) - a^*(u)\} dM(u, Z) - (Z - \pi) f_0(X_1) \right. \right. \\
&\quad \left. \left. + \int dM_C(u, Z) \left[g_0(u, Z, X) - E\{g_0(u, Z, X) | T \geq u, Z\} \right] \right) \right. \\
&\quad \left. \times \left((Z - \pi) f(X_1) + \int dM_C(u, Z) \left[g(u, Z, X) - E\{g(u, Z, X) | T \geq u, Z\} \right] \right) \right\},
\end{aligned} \tag{A.17}$$

for any functions $f(X_1)$ and $g(u, Z, X)$. From the assumption of $C \perp\!\!\!\perp (T, X) | Z$, we have that $E\{dN_C(u, Z) | U \geq u, Z, X\} = \lambda_C(u, Z) du$, which implies $E\{dM_C(u, Z) | U \geq u, Z, X\} = 0$.

Therefore, using the law of iterated conditional expectations on $(U \geq u, Z, X)$, it is easy to prove that

$$E\left((Z - \pi)f(X_1) \int dM_C(u, Z) \left[g(u, Z, X) - E\{g(u, Z, X)|T \geq u, Z\}\right]\right) = 0,$$

that is, $(Z - \pi)f(X_1)$ is uncorrelated with $\int dM_C(u, Z) [g(u, Z, X) - E\{g(u, Z, X)|T \geq u, Z\}]$ for any functions $f(X_1)$ and $g(u, Z, X)$.

Using Lemma A.2 and the result of Fleming and Harrington ([17]) for martingale stochastic integrals, we have that

$$\begin{aligned} & E\left\{\left(\int \{a(u, Z) - a^*(u)\}dM(u, Z) + \int dM_C(u, Z) \left[g_0(u, Z, X) - E\{g_0(u, Z, X)|T \geq u, Z\}\right]\right)\right. \\ & \times \left.\left(\int dM_C(u, Z) \left[g(u, Z, X) - E\{g(u, Z, X)|T \geq u, Z\}\right]\right)\right\} \\ = & E \int \left\{ \lambda_C(u, Z) du Y(u) \left[g(u, Z, X) - E\{g(u, Z, X)|T \geq u, Z\}\right] \right. \\ & \times \left. \left(-\frac{\Delta m_1^F(T, Z)}{K_C(U, Z)} + \frac{E\{m_1^F(T, Z)|T \geq u, Z\}}{K_C(u, Z)} + \left[g_0(u, Z, X) - E\{g_0(u, Z, X)|T \geq u, Z\}\right]\right) \right\} \\ = & E \int \left\{ \lambda_C(u, Z) du \left[g(u, Z, X) - E\{g(u, Z, X)|T \geq u, Z\}\right] \right. \\ & \times \left. \left(-E\{m_1^F(T, Z)|T \geq u, Z, X\} + E\{m_1^F(T, Z)|T \geq u, Z\} \right. \right. \\ & \left. \left. + K_C(u, Z) \left[g_0(u, Z, X) - E\{g_0(u, Z, X)|T \geq u, Z\}\right]\right) \right\}, \end{aligned}$$

where $m_1^F(T, Z) = \int \{a(u, Z) - a^*(u)\}K_C(u, Z)dM_T(u, Z)$. For the last equality, we use the law of iterative conditional expectation on $(T \geq u, Z, X)$, and the result that

$$E\{\Delta m_1^F(T, Z)/K_C(U, Z)Y(u)|T \geq u, Z, X\} = E\{m_1^F(T, Z)|T \geq u, Z, X\} \text{ and } E\{Y(u)|T \geq u, Z, X\} = K_C(u, Z).$$

Therefore, the equation (A.17) is equivalent to

$$\begin{aligned}
0 = & E \left\{ \left(E \left[(Z - \pi) \int \{a(u, Z) - a^*(u)\} dM(u, Z) | X_1 \right] - \pi(1 - \pi)f_0(X_1) \right) f(X_1) \right\} \\
& + E \int \left\{ \lambda_C(u, Z) \left[g(u, Z, X) - E\{g(u, Z, X) | T \geq u, Z\} \right] \right. \\
& \times \left(K_C(u, Z) \left[g_0(u, Z, X) - E\{g_0(u, Z, X) | T \geq u, Z\} \right] \right. \\
& \left. \left. - \left[E\{m_1^F(T, Z) | T \geq u, Z, X\} - E\{m_1^F(T, Z) | T \geq u, Z\} \right] \right) \right\},
\end{aligned}$$

for any functions $f(X_1)$ and $g(u, Z, X)$, which implies that

$$f_0(X_1) = \frac{1}{\pi(1 - \pi)} E \left\{ (Z - \pi) \int \{a(u, Z) - a^*(u)\} dM(u, Z) | X_1 \right\},$$

and

$$g_0(u, Z, X) = \frac{1}{K_C(u, Z)} E\{m_1^F(T, Z) | T \geq u, Z, X\}.$$

Note that $E\{dN(v) | T \geq u, Z, X\} = P(T = v | T \geq u, Z, X) K_C(v, Z) dv$, and $E\{Y(v) | T \geq u, Z, X\} = P(T \geq v | T \geq u, Z, X) K_C(v, Z)$, which implies $E\{M(v, Z) | T \geq u, Z, X\} = E\{M_T(v, Z) | T \geq u, Z, X\} K_C(v, Z)$. Hence $E\{m_1^F(T, Z) | T \geq u, Z, X\} = E\{\int \{a(u, Z) - a^*(u)\} dM(u, Z) | T \geq u, Z, X\}$, which completes the proof of Theorem 1.3.

A.4 Proof of theorem 2.1

Using the equality

$$\frac{\Delta}{K_C\{U, Z, X(U)\}} + \int \frac{dM_C\{u, Z, X(u)\}}{K_C\{u, Z, X(u)\}} = 1,$$

the influence function (2.6) can be written as

$$\begin{aligned}
& \frac{\Delta \psi_{YZ}(Y, Z; \beta_0)}{K_C\{U, Z, X(U)\}} + (Z - \pi)f(X_1) \\
& + \int \frac{dM_C\{u, Z, X(u)\}}{K_C\{u, Z, X(u)\}} [L\{u, Z, X(u)\} - (Z - \pi)f(X_1)].
\end{aligned}$$

Since $L(\cdot)$ is an arbitrary function of $\{Z, X_1, X_2^H(u)\}$, we have the class of all observed data influence functions as described in theorem 2.1.

A.5 Proof of corollary 2.2

We first prove the following Lemmas.

Lemma A.2. *For any function $b(u, Z)$, we have the following equality*

$$\begin{aligned} & \int \frac{b(v, Z)}{K_C\{v, Z, X(v)\}} dM(v, Z) \\ &= \frac{\Delta \int b(u, Z) dM_T(u, Z)}{K_C\{U, Z, X(U)\}} + \int \frac{dM_C\{u, Z, X(u)\}}{K_C\{u, Z, X(u)\}} E\left\{ \int b(u, Z) dM_T(u, Z) | T \geq u, Z \right\}, \end{aligned}$$

where $dM_T(u, Z) = I(T = u) - \lambda_0(u) \exp(\beta_0 Z) I(T \geq u) du$.

Proof. Since $E\{I(T = v) | T \geq u, Z\} = I(v \geq u) e^{\beta_0 Z} \lambda_0(v) dv P(T \geq v, Z) / P(T \geq u, Z)$ and $E\{I(T \geq v) | T \geq u, Z\} = I(v < u) + I(v \geq u) P(T \geq v, Z) / P(T \geq u, Z)$, we have that

$$E\{dM_T(v, Z) | T \geq u, Z\} = -I(v < u) e^{\beta_0 Z} \lambda_0(v) dv.$$

This implies

$$\begin{aligned} & \int \frac{dM_C\{u, Z, X(u)\}}{K_C\{u, Z, X(u)\}} E\left\{ \int b(u, Z) dM_T(u, Z) | T \geq u, Z \right\} \\ &= \frac{\Delta - 1}{K_C\{U, Z, X(U)\}} \int_0^U b(v, Z) e^{\beta_0 Z} \lambda_0(v) dv + \int_0^U \frac{\lambda_C\{u, Z, X(u)\}}{K_C\{u, Z, X(u)\}} du \int_0^u b(v, Z) e^{\beta_0 Z} \lambda_0(v) dv \\ &= \frac{\Delta}{K_C\{U, Z, X(U)\}} \int b(v, Z) I(v \leq U) e^{\beta_0 Z} \lambda_0(v) dv - \int \frac{b(v, Z)}{K_C\{v, Z, X(v)\}} I(v \leq U) e^{\beta_0 Z} \lambda_0(v) dv \end{aligned}$$

The last equality is because that

$$\int_v^U \frac{\lambda_C\{u | T \geq u, Z, X(u)\}}{K_C\{u, Z, X(u)\}} du = \frac{1}{K_C\{U, Z, X(U)\}} - \frac{1}{K_C\{v, Z, X(v)\}}.$$

Hence, we have

$$\begin{aligned}
& \frac{\Delta \int b(u, Z) dM_T(u, Z)}{K_C\{U, Z, X(U)\}} + \int \frac{dM_C\{u, Z, X(u)\}}{K_C\{u, Z, X(u)\}} E\left\{ \int b(u, Z) dM_T(u, Z) \middle| T \geq u, Z \right\} \\
&= \frac{\Delta b(U, Z)}{K_C\{U, Z, X(U)\}} - \int \frac{b(u, Z)}{K_C\{u, Z, X(u)\}} I(u \leq U) e^{\beta_0 Z} \lambda_0(u) du \\
&= \int \frac{b(u, Z)}{K_C\{u, Z, X(u)\}} dM(u, Z), \text{ for all functions } b(u, Z).
\end{aligned}$$

□

Applying Ψ_{YZ} in (2.15) to Theorem 2.1 and using Lemma A.2, we have the class of all influence function written as

$$\Psi_{PH}^{obs} = \mathcal{E}_1 + \{\mathcal{R} \oplus \mathcal{C}\} \quad (\text{A.18})$$

where

$$\mathcal{E}_1 = \left[C_a^{-1} \int \frac{a(u) \{Z - Z_T^*(u; \beta_0)\}}{K_C\{u, Z, X(u)\}} dM(u, Z; \beta_0) : \text{for all functions } a(u) \right], \quad (\text{A.19})$$

\mathcal{R} and \mathcal{C} are defined in (2.5) and (2.8), respectively.

Now for any function $W(u, Z)$, if we let the function $a(u) = W(u, 1) - \{W(u, 1) - W(u, 0)\}Z^*(u; \beta_0)$, where $Z^*(u; \beta)$ is given in (2.18), after some simple algebra, we can easily get that $a(u)\{Z - Z_T^*(u; \beta_0)\} = W(u, Z)\{Z - Z^*(u; \beta_0)\}$. Recalculate the proportionality, according to the theory of semiparametrics ([7], chapter 4), is equal to the expectation of the partial derivative of

$$e_{W(u, Z)}(D; \beta_0) = \int W(u, Z) \{Z - Z^*(u; \beta_0)\} / K_C\{u, Z, X(u)\} dM(u, Z; \beta_0)$$

with respect to β evaluated at the true value β_0 , that is, $C_W = [E\{\partial e_{W(u, Z)}(D; \beta_0) / \partial \beta\}]^{-1}$. After some algebra, we derive the proportionality constant C_W is the same as (2.19) and hence, the space \mathcal{E}_1 is the same as \mathcal{E} as described in (2.17).

A.6 Proof of theorem 2.3

Using standard results for the proportional hazards model; see, for example, ([7], Chapter 5.2), the nuisance tangent space associated with the nuisance baseline hazard function $\lambda_{0C}(\cdot, Z)$ is given by

$$\Lambda_{\lambda(\cdot, Z)} = \left[\int a(u, Z) dM_C\{u, Z, X(u)\} \text{ for all } a(u, Z) \right], \quad (\text{A.20})$$

and the nuisance tangent space associated with the nuisance parameters $\alpha = (\alpha_{11}, \dots, \alpha_{22})$ is given by

$$\Lambda_\alpha = \left(\mathbf{B} \int \left[ZX_1, Zg_1\{X_2^H(u)\}, (1-Z)X_1, (1-Z)g_2\{X_2^H(u)\} \right]^T dM_C\{u, Z, X(u)\} : \mathbf{B} \in \mathbb{R}^{1 \times 4} \right). \quad (\text{A.21})$$

Find the projection $\Pi\{\psi(D)|\Lambda_{\lambda_{0C}(\cdot, Z)}^\perp\}$ in (2.25) is equivalent to find the function $a_0(u, Z)$ such that $\psi(D) - \int a_0(u, Z) dM_C\{u, Z, X(u)\}$ is perpendicular to the nuisance tangent space $\Lambda_{\lambda(\cdot, Z)}$, that is

$$E\left(\left[\psi(D) - \int a_0(u, Z) dM_C\{u, Z, X(u)\}\right] \int a(u, Z) dM_C\{u, Z, X(u)\}\right) = 0 \quad (\text{A.22})$$

for all functions $a(u, Z)$. Using the result of Lemma A.2, we have the influence function $\psi(D) \in \Psi_{\text{PH}}^{\text{obs}}$ be written as

$$\psi(D) = \frac{\Delta\psi^F(T, Z)}{K_C\{T, Z, X(T)\}} + \int \frac{dM_C\{u, Z, X(u)\}}{K_C\{u, Z, X(u)\}} E\{\psi^F(T, Z)|T \geq u, Z\},$$

where $\psi^F(T, Z)$ is defined in (2.26). Following the standard results of Fleming and Harrington ([17]) for the covariance of two martingale processes, we have (A.22) equivalent to

$$\begin{aligned} 0 &= E\left(\int \left[-\frac{\Delta\psi^F(T, Z)}{K_C\{T, Z, X(T)\}} + \frac{E\{\psi^F(T, Z)|T \geq u, Z\}}{K_C\{u, Z, X(u)\}} + L\{u, Z, X(u)\} - a_0(u, Z)\right] \right. \\ &\quad \times \left. a(u, Z)\lambda_{C0}(u, Z)H\{u, Z, X(u)\}Y(u)du\right), \end{aligned}$$

for all functions $a(u, Z)$. After a little algebra, we have that

$$\begin{aligned}
a_0(u, Z) &= -\frac{E\left[\psi^F(T, Z)/K_C\{u, Z, X(u)\}H\{u, Z, X(u)\}Y(u)|Z\right]}{E\left[H\{u, Z, X(u)\}Y(u)|Z\right]} \\
&+ \frac{E\left[E\{\psi^F(T, Z)|T \geq u, Z\}/K_C\{u, Z, X(u)\}H\{u, Z, X(u)\}Y(u)|Z\right]}{E\left[H\{u, Z, X(u)\}Y(u)|Z\right]} \\
&+ \frac{E\left[L\{u, Z, X(u)\}H\{u, Z, X(u)\}Y(u)|Z\right]}{E\left[H\{u, Z, X(u)\}Y(u)|Z\right]}.
\end{aligned}$$

Therefore, $\Pi\{\psi(D)|\Lambda_{\lambda_{0C}(\cdot, Z)}^\perp\} = \psi(D) - \int a_0(u, Z)dM_C\{u, Z, X(u)\}$ is identical to (2.25).

Similarly, obtain the projection $\Pi(\psi(D)|\Lambda_\alpha)$ is equivalent to find the function \mathbf{B}_0 such that $\psi(D) - \mathbf{B}_0 \int \mathbf{Q}(\mathbf{u})dM_C\{u, Z, X(u)\}$ is perpendicular to the nuisance tangent space $\mathbf{B} \int \mathbf{Q}(\mathbf{u})dM_C\{u, Z, X(u)\}$, where

$$\mathbf{Q}(\mathbf{u}) = [ZX_1, Zg_1\{X_2^H(u)\}, (1-Z)X_1, (1-Z)g_2\{X_2^H(u)\}]^T. \quad (\text{A.23})$$

After some algebra, we have that

$$\begin{aligned}
\mathbf{B}_0 &= -\frac{E\left[\mathbf{Q}(\mathbf{u})\psi^F(T, Z)/K_C\{u, Z, X(u)\}H\{u, Z, X(u)\}Y(u)\lambda_{C0}(u, Z)\right]}{E\left[\mathbf{Q}(\mathbf{u})\mathbf{Q}(\mathbf{u})^T H\{u, Z, X(u)\}Y(u)\lambda_{C0}(u, Z)\right]} \\
&+ \frac{E\left[E\{\mathbf{Q}(\mathbf{u})\psi^F(T, Z)|T \geq u, Z\}/K_C\{u, Z, X(u)\}H\{u, Z, X(u)\}Y(u)\right]}{E\left[\mathbf{Q}(\mathbf{u})\mathbf{Q}(\mathbf{u})^T H\{u, Z, X(u)\}Y(u)\lambda_{C0}(u, Z)\right]} \\
&+ \frac{E\left[\mathbf{Q}(\mathbf{u})L\{u, Z, X(u)\}H\{u, Z, X(u)\}Y(u)\right]}{E\left[\mathbf{Q}(\mathbf{u})\mathbf{Q}(\mathbf{u})^T H\{u, Z, X(u)\}Y(u)\lambda_{C0}(u, Z)\right]},
\end{aligned}$$

and

$$\Pi(\psi(D)|\Lambda_\alpha) = \mathbf{B}_0 \int \mathbf{Q}(\mathbf{u})dM_C\{u, Z, X(u)\}. \quad (\text{A.24})$$