Biometrika (2008), **95**, 3, *pp.* 679–694
doi: 10.1093/biomet/asn003

Advance Access publication 4 June 2008
Printed in Great Britain

Improving the efficiency of the log-rank test using auxiliary covariates

BY XIAOMIN LU

Department of Epidemiology and Biostatistics, College of Public Health and Health Professions, University of Florida Gainesville, Florida 32611, U.S.A.
xlu2@phhp.ufl.edu

AND ANASTASIOS A. TSIATIS

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203, U.S.A.
tsiatis@stat.ncsu.edu

SUMMARY

Under the assumption of proportional hazards, the log-rank test is optimal for testing the null hypothesis $H_0 : \beta = 0$, where β denotes the logarithm of the hazard ratio. However, if there are additional covariates that correlate with survival times, making use of their information will increase the efficiency of the log-rank test. We apply the theory of semiparametrics to characterize a class of regular and asymptotically linear estimators for β when auxiliary covariates are incorporated into the model, and derive estimators that are more efficient. The Wald tests induced by these estimators are shown to be more powerful than the log-rank test. Simulation studies are used to illustrate the gains in efficiency.

Some key words: Efficient estimator; Influence function; Log-rank test; Nuisance tangent space; Proportional hazard model; Regular and asymptotically linear estimator.

1. INTRODUCTION

The most commonly used test for treatment differences with censored survival data is the log-rank test (Mantel, 1966; Peto & Peto, 1972). It is well known that, if the survival time T conditional on treatment assignment Z follows the proportional hazards model

$$\lambda_{T|Z}(t|z) = \lambda(t)\exp(\beta z), \tag{1}$$

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, then the log-rank test is optimal for testing for treatment difference, i.e., testing the null hypothesis $H_0 : \beta = 0$.

The log-rank test is equivalent to the partial likelihood score test of Cox (1972, 1975) for the proportional hazards regression model, and is asymptotically equivalent to the Wald test induced by $\hat{\beta}_{\mathrm{PH}}$, the maximum partial likelihood estimator of β. From the theory of semiparametrics, see for example Begun et al. (1983) and Bickel et al. (1993), we know that the maximum partial likelihood estimator is the efficient estimator for β under the assumption of proportional hazards. However, these results assume that only information on time to event, which may be censored, and treatment assignment are available.

In most clinical trials, in addition to the data on survival and censoring times and treatment assignment, auxiliary information is also collected on variables such as age, gender and health conditions that may be important prognostic factors that are correlated with the time to event. Such randomized studies are subject to two main sources of efficiency loss. In addition to a loss of efficiency due to possible censoring, 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. Our idea is to use semiparametric theory to recover information from auxiliary covariates that correlate with survival time without making assumptions beyond those needed for the validity of the log-rank test and the maximum partial likelihood estimator.

Remark 1. Since our primary goal is to estimate the treatment-specific log-hazard-ratio β given by (1) and to test the null hypothesis $\beta = 0$, we will not consider inference that is conditional on the auxiliary covariates; for example, a popular model is the proportional hazards regression model which assumes that

$$\lambda(t|Z, X) = \lambda(t)\exp\left(\beta_* Z + \beta_1^{\mathrm{T}} X\right). \tag{2}$$

In this model, β_* represents the treatment effect conditional on the baseline covariates X. For nonlinear models such as Cox's proportional hazards model, a consistent estimator for β_* will not in general be a consistent estimator for β in (1). We will show that the test for $\beta_* = 0$ will be biased, in not controlling the type I error, and the resulting estimator, although unbiased, will in some cases be inefficient as compared to the unadjusted analysis.

2. MODEL FRAMEWORK AND NOTATION

The data can be summarized as n realizations of independent and identically distributed random vectors $D_i = (U_i, \Delta_i, Z_i, X_i)$ $(i = 1, \ldots, n)$. For the ith individual, $U_i = \min(T_i, C_i)$, where T_i denotes the underlying survival time and C_i denotes the potential censoring time, $\Delta_i = I(T_i \leqslant C_i)$ 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 $X_i = (X_{1i}^{\mathrm{T}}, X_{2i}^{\mathrm{T}})^{\mathrm{T}}$, 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.

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 log-rank test statistic 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 properly designed randomized clinical trials, 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 covariates X_i are also collected on the ith individual, some of which may be correlated with the survival time, so-called prognostic factors. As a result 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 a known π with $0 < \pi < 1$; that is,

$$Z \perp\!\!\!\perp X_1, \quad \mathrm{pr}(Z = 1) = \pi. \tag{3}$$

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

$$C \perp\!\!\!\perp (T, X) \mid Z, \tag{4}$$

and denote the conditional survival distribution of censoring time given Z by $K_{C|Z}(u|Z)$, which is left unspecified. This assumption may be questionable in some situations especially if the belief is that patients who 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 log-rank 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 were 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 (4) will be made.

Without making any additional assumptions, other than those given by (1), (3) and (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 a Wald test, that will be more powerful than the log-rank test to detect proportional hazards alternatives.

3. The class of all semiparametric estimators for β

We first introduce some terminology and results from 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 ith influence function of $\hat{\beta}_n$. The influence function of a regular, see for example Newey (1990) for a definition of the term, and asymptotically linear estimator is unique and determines its asymptotic properties. For example, the asymptotic variance of a regular and asymptotically linear estimator $\hat{\beta}_n$ equals the variance of its influence function.

Influence functions of regular and asymptotically linear estimators can be viewed as vectors in a Hilbert space $\mathcal{H}$ consisting of all zero-mean finite-variance random functions, $h(D)$, of a single observation D, equipped with the covariance inner product $\langle h_1, h_2\rangle = E\{h_1(D)h_2(D)\}$ so that the square of the norm is $\langle h, h\rangle = \text{var}\{h(D)\}$. The key result is that influence functions of regular and asymptotically linear estimators for β must be orthogonal to the so-called nuisance tangent space; for more details see Tsiatis (2006, § 4).

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

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. In particular, η_1 denotes the baseline hazard function $\lambda(t)$ for the proportional hazards model (1). The parameter $\eta_3 = (\eta_{31}, \eta_{32})$ denotes the conditional density of X given Z satisfying assumption (3); that is,

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

where η_{31} denotes an arbitrary conditional density of X_2 given (Z, X_1) and η_{32} denotes an arbitrary marginal density of X_1, which, by assumption, i.e., because of randomization, is independent of Z. The parameter η_4 denotes the class of conditional densities of (T, X) given Z that satisfy the constraints imposed by (1) and (3), namely that

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

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

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

The parameter η_2 denotes an arbitrary conditional density of C given Z.

THEOREM 1. *The nuisance tangent space Λ can be written as a sum of tangent spaces associated with each of the components $\eta_1, \ldots, \eta_4$ making up η; that is,*

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

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

$$\Lambda_1 \subset \{a_1(U, \Delta, Z, X) : a_1 \in H_1\}, \tag{5}$$

$$E\{\Lambda_1|U, \Delta, Z\} = \Lambda_1^*, \tag{6}$$

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 $dN(u) - \lambda_{T|(Z,X)}(u|z, x)Y(u)du$ where $N(u) = I(U \leqslant u, \Delta = 1)$ is the counting process that counts the number of observed deaths up to and including time u and $Y(u) = I(U \geqslant u)$ is the 'at risk' indicator at time u, $\lambda_{T|(Z,X)}(\cdot)$ is the conditional hazard rate of T given (Z, X), $\Lambda_1^* = \{\int a(u)dM(u, Z)$ for all functions $a(u)\}$, $dM(u, Z)$ 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

$$\Lambda_3 \subset \{a_3(U, \Delta, Z, X) : \ a_3 \in H_1, E(a_3|U, \Delta, Z) = 0\}, \tag{7}$$

$$E\{\Lambda_3|Z, X\} = \Lambda_3^*, \tag{8}$$

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

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

$$\Lambda_4 = \{a_4(U, \Delta, Z, X) : a_4 \in H_1, \text{ and } E(a_4|U, \Delta, Z) = 0\}, \tag{9}$$

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

$$\Lambda_2 = \left\{ \int a_2(u, Z)dM_C(u, Z) \text{ for all functions } a_2(u, Z) \right\}, \tag{10}$$

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

All proofs are given in the Appendix.

Remark 2. Note that (5) and (6) 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 $\Lambda^\perp$. The situation is similar for Λ_3.

THEOREM 2. *The orthogonal complement of the nuisance tangent space $\Lambda^\perp$ is given by*

$$\Lambda^\perp = \mathcal{E} + \{\mathcal{R} \oplus \mathcal{C}\},$$

where

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

$$\mathcal{R} = \{(Z - \pi) f(X_1) \textit{ for all functions } f(X_1)\},$$

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

We refer to the space $\mathcal{E}$ as the estimation space because this space contains all the regular and asymptotically linear estimators for β that do not use covariates. The mutually orthogonal spaces $\mathcal{R}$ and $\mathcal{C}$ do involve the covariates and can be used to derive more efficient estimators. The space $\mathcal{R}$ only uses the baseline, pre-randomization, covariates X_1 and the inclusion of this space allows us to retrieve some of the efficiency lost because of 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; the inclusion of this space allows us to gain back some of the efficiency lost because of censoring and is therefore referred to as the censoring space.

Since a regular and asymptotically linear estimator for β must have an influence function that is orthogonal to the nuisance tangent space, we derive that all semiparametric estimators for β can be represented as the solution to the estimating equation

$$\sum_{i=1}^{n} \left[\int \{a(u, Z_i) - \bar{a}(u; \beta)\} dN_i(u) (Z_i - \pi) f(X_{1i}) + \int \{g(u, Z_i, X_i) - \bar{g}(u, Z_i)\} dN_{C_i}(u) \right] = 0, \tag{11}$$

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 $N_C(u) = I(U \leqslant u, \Delta = 0)$ is the counting process that counts the number of observed censored observations up to and including time u. Both the second and the third summands of (11) have mean zero and hence the estimator obtained by solving the estimating equation (11) 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 (11). This suggests that fixing $a(u, Z) = Z$ and judicious choice of $f(X_1)$ and $g(u, Z, X)$ will lead us to more efficient estimators.

Remark 3. The theoretical argument so 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 $t_0, \ldots, t_m$, we clearly cannot observe these values beyond the time that a patient is at risk. Nonetheless, without loss of generality, if we define the function $g(u, Z, X)$ by $g\{u, X_1, X_2(t_0) I(u \geqslant t_0), \ldots, X_2(t_m) I(u \geqslant t_m), Z\}$, a patient who is not at risk at time t_j will not have his/her covariates $X_2(t_j)$ considered in the estimating equation (11), thereby allowing the use of such time-dependent covariates.

4. IMPROVING THE EFFICIENCY OF THE LOG-RANK TEST

4·1. *Estimators for β are more efficient than the maximum partial likelihood estimator*

We begin by restricting our attention to the subclass of regular and asymptotically linear estimators for β, which are the solution to the estimating equation (11) 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 $\mathcal{B}_{a(u,Z)}$ has an influence function proportional to an element in

$$\Lambda^{\perp}_{a(u,Z)} = \Bigg\{ \int \{a(u, Z) - a^*(u; \beta_0)\} dM(u, Z) + (Z - \pi) f(X_1) + \int dM_C(u, Z)[g(u, Z, X) - E\{g(u, Z, X) \mid T \geqslant u, Z\}] : \text{for all functions } f(X_1) \text{ and } g(u, Z, X) \Bigg\}.$$

According to the theory of semiparametrics (Tsiatis, 2006, § 4), the proportionality constant C_a 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

$$e_{a(u,Z)}(D; \beta) = \int \{a(u, Z) - a^*(u; \beta)\} dM(u, Z; \beta), \tag{12}$$

$a^*(u; \beta) = E\{a(u, Z) \exp(\beta Z) Y(u)\}/E\{\exp(\beta Z) Y(u)\}$, and $dM(u, Z; \beta)$ is the martingale increment $dN(u) - \lambda(u) \exp(\beta Z) Y(u) du$. After some algebra, we derive the proportionality constant to be

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

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

THEOREM 3. *Consider the subclass of influence functions proportional to the elements in $\Lambda^{\perp}_{a(u,Z)}$. The efficient influence function for β within the subclass is proportional to*

$$\int \{a(u, Z) - a^*(u)\} dM(u, Z) - (Z - \pi) f_0(X_1) + \int \frac{dM_C(u, Z)}{K_C(u, Z)} [g_0(u, Z, X) - E\{g_0(u, Z, X) \mid T \geqslant u, Z\}],$$

where

$$f_0(X_1) = \frac{1}{\pi(1 - \pi)} E\{(Z - \pi) e_{a(u,Z)}(D; \beta_0) \mid X_1\}, \tag{13}$$

$$g_0(u, Z, X) = E\{e_{a(u,Z)}(D; \beta_0) \mid T \geqslant u, Z, X\} \tag{14}$$

with $e_{a(u,Z)}(D; \beta_0)$ defined by (12).

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 see directly from Theorem 3 that the influence 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 $c_{a(u,Z)}(D) = \int [g_0(u, Z, X) - E\{g_0(u, Z, X) \mid T \geqslant u, Z\}] dM_C(u, Z)/K_C(u, Z)$.

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 asymptotically linear estimator for β within the subclass $\mathcal{B}_{a(u,Z)}$ by solving the estimating equation

$$\sum_{i=1}^{n}\{\hat{e}_{a(u,Z)}(D_i;\beta) - \hat{r}_{a(u,Z)}(D_i) + \hat{c}_{a(u,Z)}(D_i)\} = 0, \tag{15}$$

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 the functions $f_0(\cdot)$ and $g_0(\cdot)$ must be estimated. We suggest the following strategy. Model the conditional expectation $f_0(X_1)$ in (13) as $f(X_1;a) = a^{\mathrm{T}}q(X_1)$ that is linear in a and model $g_0(u,Z,X)$ in (14) as $g(u,Z,X;b) = b^{\mathrm{T}}w(u,Z,X)$, where a and 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 asymptotically linear estimators, which solve the estimating equations

$$\sum_{i=1}^{n}\bigg[\hat{e}_{a(u,Z)}(D_i;\beta) - (Z_i - \pi)f(X_{1i};a) - \int\{g(u,Z_i,X_i;b) - \bar{g}(u,Z_i;b)\}dN_{C_i}(u)/\hat{K}_C(u,Z_i)\bigg] = 0,$$

for all $a \in \mathbb{R}^{r_a}$ and $b \in \mathbb{R}^{r_b}$. Clearly, the maximum partial likelihood estimator is in this class with $a = b = 0$.

We define a_0 and 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 a_0 satisfies $E[\{e_{a(u,Z)}(D;\beta_0) - (Z-\pi)a_0^{\mathrm{T}}q(X_1)\}(Z-\pi)q(X_1)^{\mathrm{T}}a] = 0$ for all $a \in \mathbb{R}^{r_a}$. After some algebra, we derive that $a_0 = [\pi(1-\pi)E\{q(X_1)q(X_1)^{\mathrm{T}}\}]^{-1}E\{q(X_1)(Z-\pi)e_{a(u,Z)}(D;\beta_0)\}$ and, consequently, a_0 is estimated by

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

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 $b_0 = [E\{H_w(u,Z,X)H_w(u,Z,X)^{\mathrm{T}}\}]^{-1}E\{H_w(u,Z,X)e_{a(u,Z)}(D;\beta_0)\}$, where $H_w(u,Z,X) = \int dM_C(u,Z)/K_C(u,Z)[w(u,Z,X) - E\{w(u,Z,X)|T \geqslant u,Z\}]$ and b_0 is estimated by

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

where

$$\hat{H}_w(u,Z_i,X_i) = \int\{dN_{C_i}(u) - \hat{\lambda}_{C|Z}(u|Z_i)duY_i(u)\}\{w(u,Z_i,X_i) - \bar{w}(u,Z_i)\}/\hat{K}_C(u,Z_i), \hat{\lambda}_{C|Z}(u|Z)du,$$

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, $\hat{\lambda}_{C|Z}(u|z) = \sum_i dN_{Ci}(u, z)/\sum_i Y_i(u, z)$, and $\bar{w}(u, Z_i)$ is calculated the same way as $\bar{g}(u, Z_i)$ in the estimating equation (11). We denote the estimated function $f(X_1; \hat{a})$ by $\hat{f}_0(X_1)$ and the estimated function $g(u, Z, X; \hat{b})$ by $\hat{g}_0(u, Z, X)$.

Obviously, from the definitions of $\hat{a}$ and $\hat{b}$, using the estimated functions $\hat{f}_0(\cdot)$ and $\hat{g}_0(\cdot)$ in the estimating equation (15) will result in a more efficient estimator for β than the estimating equation

$$\sum_{i=1}^{n} \hat{e}_{a(u,Z)}(D_i; \beta) = 0, \tag{16}$$

even if the true conditional expectations (13) and (14) 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 (16) becomes the partial likelihood score equation leading to the maximum partial likelihood estimator $\hat{\beta}_{\text{PH}}$ for β without auxiliary covariates. Let $\hat{\beta}_{\text{AUG}}$ be the estimator for β obtained by solving the equation (15) when the fitted functions $\hat{f}_0(\cdot)$ and $\hat{g}_0(\cdot)$ are used.

4·2. *Variance estimator for* $\hat{\beta}_{\text{AUG}}$

Using standard methods for censored survival analysis, we can show that the estimator $\hat{\beta}_{\text{AUG}}$ 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

$$m_2(D; \beta) = e_{a(u,Z)}(D; \beta)|_{a(u,Z)=Z} - (Z - \pi) f(X_1; a_0)$$
$$+ \int \frac{dM_C(u, Z)}{K_C(u, Z)} [g(u, Z, X; b_0) - E\{g(u, Z, X; b_0)|T \geqslant u, Z\}].$$

Therefore, according to the results in Chapter 4 of Tsiatis (2006), we can estimate the variance of $\hat{\beta}_{\text{AUG}}$ by the sandwich variance

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

where

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

Here, $\bar{Z}(u; \beta)$ and $\bar{g}_0(u, Z; \hat{b}_0)$ are calculated as in the estimating equation (11).

5. SIMULATION

We performed a Monte Carlo simulation study to compare the proposed estimator $\hat{\beta}_{\text{AUG}}$ to the maximum partial likelihood estimator $\hat{\beta}_{\text{PH}}$. We considered only one baseline covariate X. In order to ensure that the covariate X was independent of treatment assignment Z but was correlated with 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 means zero, variances 1, and correlation ρ. We then independently generated the treatment indicator Z as $\text{Ber}(\pi)$. 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 $Z = 0, 1$, respectively, was generated independently of (pT, X) from an exponential distribution $C|Z \sim \text{Exp}(c|z)$. 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 $\pi = 0{\cdot}5$, and the correlation in the bivariate normal was taken to be $\rho = 0{\cdot}7$, which resulted in a sample correlation of approximately 0·6 between the survival time T and baseline covariate X conditional on Z. Two values for the proportional hazards regression coefficient were considered, the null value $\beta = 0$ and $\beta = 0{\cdot}25$, and, for each treatment $Z = 0, 1$, the value c for the exponential distribution of the censoring variable was taken to be one of the 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 and 2, we compare the bias, standard error estimate, Monte Carlo standard error, relative efficiency defined as the ratio of the variance estimate and the Monte Carlo variance, type I error rate and the powers 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, namely $\hat{\beta}_*$, the estimator of β_* in model (2), and $\hat{\beta}_{**}$, the estimator of β_* in (2) after adding the quadratic term X^2 in the model. The variances of these two covariate-adjusted estimators were estimated using the standard estimators for the proportional hazards model.

In general, the traditional estimators $\hat{\beta}_*$ and $\hat{\beta}_{**}$, based on the conditional proportional hazards model on X and on X and X^2, are biased estimators of β as we illustrate in Table 2 when

Table 1. *Parameter estimates $\hat{\beta}_{\text{PH}}$, $\hat{\beta}_{\text{AUG}}$, $\hat{\beta}_*$ and $\hat{\beta}_{**}$. The true value is $\beta_0 = 0$, the sample correlation between T and X given Z is about 0·6, and entries in parentheses 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 (0·99)	0·150 (0·97)
		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 (0·99)	0·184 (0·98)
		600	0·116	0·091 (1·63)	0·117 (0·98)	0·117 (0·98)
MCSE	25%	250	0·146	0·118 (1·53)	0·170 (0·74)	0·170 (0·74)
		600	0·095	0·073 (1·67)	0·106 (0·80)	0·107 (0·79)
	50%	250	0·185	0·156 (1·40)	0·210 (0·78)	0·212 (0·76)
		600	0·117	0·095 (1·52)	0·131 (0·80)	0·132 (0·79)
power	25%	250	0·050	0·064	0·082	0·079
		600	0·046	0·053	0·081	0·083
	50%	250	0·052	0·073	0·080	0·084
		600	0·049	0·060	0·077	0·080

$\hat{\beta}_{\text{PH}}$, maximum partial likelihood estimator; $\hat{\beta}_{\text{AUG}}$, our proposed estimator; $\hat{\beta}_*$, estimator for β_* in (2); $\hat{\beta}_{**}$, estimator for β_* in (2) after adding X^2 in the model; Cen, proportion of the data being censored; SE, standard error; MCSE, Monte Carlo standard error.

Table 2. *Parameter estimates $\hat\beta_{\mathrm{PH}}$, $\hat\beta_{\mathrm{AUG}}$, $\hat\beta_{*}$ and $\hat\beta_{**}$. The true value is $\beta_0 = 0{\cdot}25$, the sample correlation between T and X given Z is about $0{\cdot}6$, and entries in parentheses are relative efficiencies with respect to $\hat\beta_{\mathrm{PH}}$*

	Cen	n	$\hat\beta_{\mathrm{PH}}$	$\hat\beta_{\mathrm{AUG}}$	$\hat\beta_{*}$	$\hat\beta_{**}$
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 (0·97)	0·152 (0·96)
		600	0·095	0·073 (1·71)	0·096 (0·98)	0·097 (0·96)
	50%	250	0·183	0·142 (1·65)	0·186 (0·97)	0·187 (0·96)
		600	0·117	0·092 (1·62)	0·118 (0·98)	0·118 (0·98)
MCSE	25%	250	0·147	0·119 (1·52)	0·171 (0·74)	0·171 (0·74)
		600	0·096	0·075 (1·65)	0·106 (0·82)	0·107 (0·81)
	50%	250	0·187	0·157 (1·41)	0·212 (0·78)	0·214 (0·76)
		600	0·119	0·097 (1·52)	0·132 (0·81)	0·133 (0·80)
power	25%	250	0·406	0·593	0·612	0·609
		600	0·710	0·915	0·926	0·925
	50%	250	0·282	0·428	0·438	0·438
		600	0·557	0·753	0·782	0·779

$\hat\beta_{\mathrm{PH}}$, maximum partial likelihood estimator; $\hat\beta_{\mathrm{AUG}}$, our proposed estimator; $\hat\beta_{*}$, estimator for β_* in (2); $\hat\beta_{**}$, estimator for β_* in (2) after adding X^2 in the model; Cen, proportion of the data being censored; SE, standard error; MCSE, Monte Carlo standard error.

$\beta = 0{\cdot}25$. Under the null hypothesis $\beta = 0$, these estimators are unbiased, but the usual estimator for the standard error of these regression estimators is biased, underestimating the true standard error. As a consequence, the Wald tests induced by $\hat\beta_{*}$ and $\hat\beta_{**}$ are also biased with type I error rates around 0·08 at the nominal level 0·05.

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

6. AN EXAMPLE

We applied this methodology to the data from 2139 patients from the AIDS Clinical Trials Group (ACTG) protocol 175 (Hammer et al., 1996), a study that randomized 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 (Zidouvudine 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 decline in their CD4 cell count of at least 50%, an event indicating progression to AIDS, or death. Roughly 76% of the data were censored.

Figure 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 3. For example, the estimate of the log-hazard-ratio between

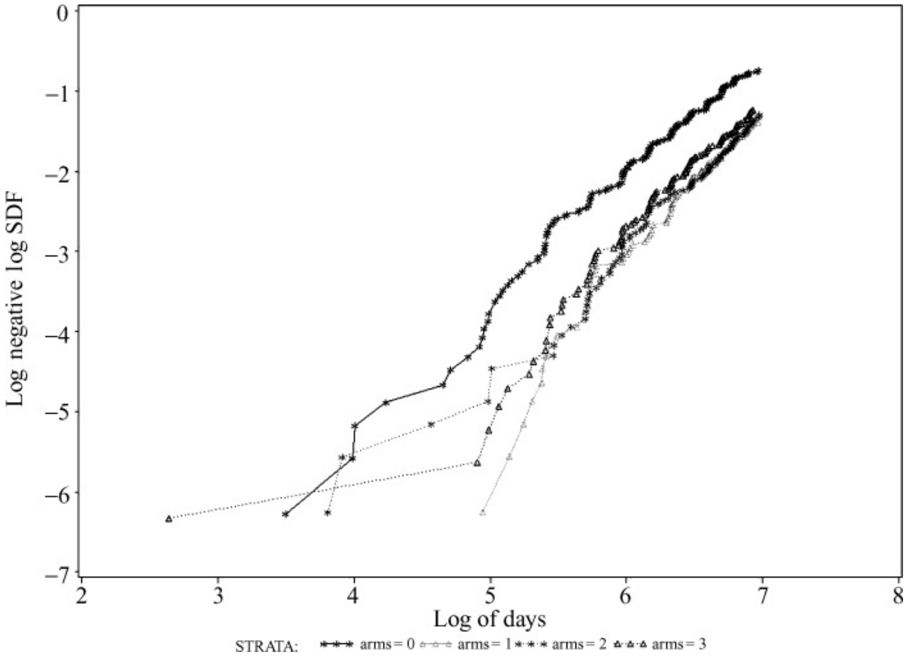

Fig. 1. ACTG 175 data. Log-negative survival functions of time to death for treatment 0, 1, 2 and 3, respectively.

treatment 0 and treatment 1 is $-0{\cdot}703$ and its standard error is $0{\cdot}124$, which is highly significant statistically.

In our analysis, we also considered several prognostic baseline auxiliary covariates and post-treatment covariates. The baseline covariates include CD4, CD8, age in years, weight in kg, history of IV drug use ($0 =$ no), Karnofsky score on a scale of 0–100, Zidouvudine 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 3.

The estimator $\hat{\beta}_{\text{AUG}}$ has estimates of standard errors, $0{\cdot}110$, $0{\cdot}104$ and $0{\cdot}105$, respectively; all are more efficient, roughly, 25%, 36% and 21% more efficient, respectively, than the corresponding maximum partial likelihood estimator for β. These results are consistent with the results from such moderately prognostic auxiliary covariates.

Table 3. *Estimates $\hat{\beta}_{\text{PH}}$ and $\hat{\beta}_{\text{AUG}}$ for the ACTG 175 data*

		Estimate	Standard Error	RE
Treatments 0 and 1	$\hat{\beta}_{\text{PH}}$	$-0{\cdot}703$	$0{\cdot}124$	$1{\cdot}00$
	$\hat{\beta}_{\text{AUG}}$	$-0{\cdot}723$	$0{\cdot}110$	$1{\cdot}25$
Treatments 0 and 2	$\hat{\beta}_{\text{PH}}$	$-0{\cdot}640$	$0{\cdot}121$	$1{\cdot}00$
	$\hat{\beta}_{\text{AUG}}$	$-0{\cdot}555$	$0{\cdot}104$	$1{\cdot}36$
Treatments 0 and 3	$\hat{\beta}_{\text{PH}}$	$-0{\cdot}528$	$0{\cdot}116$	$1{\cdot}00$
	$\hat{\beta}_{\text{AUG}}$	$-0{\cdot}627$	$0{\cdot}105$	$1{\cdot}21$

$\hat{\beta}_{\text{PH}}$, maximum partial likelihood estimator; $\hat{\beta}_{\text{AUG}}$, our proposed estimator; RE, the relative efficiencies with respect to $\hat{\beta}_{\text{PH}}$.

7. CONCLUDING REMARKS

One of the crucial assumptions is that $C \perp\!\!\!\perp (T, X)|Z$ in assumption (4). However, 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

$$\sum_{i=1}^{n} \left[\int \{a(u, Z_i) - \bar{a}(u; \beta)\} dN_i(u) + (Z_i - \pi) f(X_{1i}) + \int \{g(u, Z_i, X_i) - \bar{g}(u)\} dN_{C_i}(u) \right] = 0, \tag{17}$$

where (17) is similar to the estimating equations given in (11) 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 § 4.

ACKNOWLEDGEMENT

The authors thank the Professor D. M. Titterington, an associate editor and a referee for their careful reading and comments. This research was supported by grants from the U.S. National Institutes of Health.

APPENDIX

Technical details

Sketch proof of Theorem 1. The likelihood of the data is given by

$$\begin{aligned} & p(u, \delta, z, x|\beta_0, \eta_1, \eta_2, \eta_3, \eta_4) \\ &\quad = \{p_{(T,X)|Z}(u, x|z; \beta_0, \eta_1, \eta_3, \eta_4)\}^{\delta} \{S_{(T,X)|Z}(u, x|z; \beta_0, \eta_1, \eta_3, \eta_4)\}^{1-\delta} \\ &\qquad \times \{\lambda_{C|Z}(u|z; \eta_2)\}^{1-\delta} \exp\{-L_{C|Z}(u|z; \eta_2)\} p(z), \end{aligned}$$

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 $\eta = (\eta_1, \ldots, \eta_4)$ must satisfy the constraints imposed by (1), (3) and (4), after a little algebra, we have that all the nuisance parameters must satisfy the following constraints (A1) and (A2):

$$\begin{aligned} & \int \{p_{(T,X)|Z}(u, x|z; \beta_0, \eta_1, \eta_3, \eta_4)\}^{\delta} \{S_{(T,X)|Z}(u, x|z; \beta_0, \eta_1, \eta_3, \eta_4)\}^{1-\delta} dx \\ &\qquad = \{\lambda(u; \eta_1) \exp(\beta_0 z)\}^{\delta} \exp\{-L(u; \eta_1) \exp(\beta_0 z)\}, \end{aligned} \tag{A1}$$

$$\sum_{\delta=0}^{1} \int p(u, \delta, z, x|\beta_0, \eta_1, \eta_2, \eta_3, \eta_4)du = p_{X|Z}(x|z; \eta_3). \tag{A2}$$

Following the steps outlined in § 5.2 of Tsiatis (2006), we derive the nuisance tangent space Λ_2 associated with η_2. For this censored survival data, it is convenient to use the counting process and associated martingale process notation defined by Fleming & Harrington (1991):

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

To derive the nuisance tangent space Λ_1, we need to consider any parametric submodel for η_1 in terms of a finite-dimensional parameter γ_1, which satisfies (A1) and (A2). Taking the derivative with respect to γ_1 of the logarithm of both sides of equations (A1) and (A2), and using the fact that $a_1(u, Z, X) = b_1(u, Z, X) - \int_u^\infty b_1(v, Z, X) p_{(T,X)|Z}(v, X|Z; \gamma_{10}) dv / S_{(T,X)|Z}(u, X|Z; \gamma_{10})$ is equivalent to $b_1(u, Z, X) = a_1(u, Z, X) - \int_0^u a_1(u, Z, X) \lambda_{T|(Z,X)}(v|Z, X; \gamma_{10}) dv$, after some algebra, we obtain that

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

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

$$E\left\{ \int a_1(u, Z, X) dM(u, Z, X) \middle| Z, X \right\} = 0. \tag{A3}$$

Since (A3) is automatically true for any function $\int a_1(u, Z, X) dM(u, Z, X)$, we have the result (5).

To derive (6), 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 (A1) and (A2).

Let $F(t|z; \gamma_1)$ be the conditional cumulative distribution function of T given Z and let $F_{X|Z}(x|z)$ 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 proved that

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

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 (7) and (8), and the form of the nuisance tangent space Λ_4 as given by (9).

Sketch proof of Theorem 2. To prove Theorem 2, we need to use the following results.

LEMMA A1.
(i) *If S_1 and S_2 are two linear subspaces in the Hilbert space $\mathcal{H}$, then $(S_1 + S_2)^\perp = S_1^\perp \cap S_2^\perp$.*
(ii) *The entire Hilbert space $\mathcal{H}$, that is, the set of all zero-mean 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 $H_1 = \{\int a(u, Z, X) dM(u, Z, X)$ for all functions $a(u, z, x)\}$, $H_2 = \{\int a(u, Z, X) dM_C(u, Z)$ for all functions $a(u, z, x)\}$, $H_3 = [a(Z, X) : E\{a(Z, X)\} = 0]$.*
(iii) *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 (10), $H_1^* = \{\int a(u, Z) dM(u, Z)$ for any function $a(u, z)\}$, and $H_3^* = [a(Z) : E\{a(Z)\} = 0]$.*

From result (i), we deduce that $\Lambda^\perp = \Lambda_1^\perp \cap \Lambda_2^\perp \cap \Lambda_3^\perp \cap \Lambda_4^\perp$. We begin by deriving $\Lambda_4^\perp$. The linear subspace Λ_4 in (9) can be written as $\Lambda_4 = Q_0 \cap H_1$, where $Q_0 = [a(D) : E\{a(D)|U, \Delta, Z\} = 0]$. Using result (i) 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 zero-mean functions of (U, Δ, Z), which, by result (iii), equals $H_1^* \oplus \Lambda_2 \oplus H_3^*$, and the linear space $H_1^\perp$ can be shown by result (ii) to be equal to $H_2 \oplus H_3$. Since $\Lambda_2 \subset H_2$ and $H_3^* \subset H_3$, we have that $\Lambda_4^\perp = H_1^* + (H_2 \oplus H_3)$.

Since $\Lambda_1 \subset H_1$ and $H_1 \perp (H_2 \oplus H_3)$, we deduce that $\Lambda_1^\perp \cap \Lambda_4^\perp = \Lambda_1^\perp \cap H_1^*$, which consists of all elements $h_1^* = \int a(u, Z) dM(u, Z) \in H_1^*$ such that $E\{h_1^* a^*(D)\} = 0$, for any function $a^*(D) \in \Lambda_1$. Using iterated conditional expectations, we obtain that

$$0 = E\left[\int a(u, Z)dM(u, Z)E\left\{a^*(D)|U, \Delta, Z\right\}\right]$$
$$= E\left\{\int a(u, Z)dM(u, Z)\int a^*(u)dM(u, Z)\right\}$$
$$= E\left\{\int a(u, Z)a^*(u)\lambda_0(u)\exp(\beta_0 Z)Y(u)du\right\} \tag{A4}$$

for all $a^*(u)$, where (A4) follows because the covariance between two martingale processes is equal to the expectation of the predictable covariation process; see Fleming & Harrington (1991). In order for (A4) to hold, the function $a(u, Z)$ must be centred appropriately. As a consequence, we obtain that $\Lambda_1^{\perp} = \mathcal{E} + (H_2 \oplus H_3)$, where

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

Continuing with our argument, because $\Lambda_2 \subset H_2$, and because $\mathcal{E} \perp \Lambda_2$, we deduce that $\Lambda_4^{\perp} \cap \Lambda_1^{\perp} \cap \Lambda_2^{\perp} = \mathcal{E} + \{(H_2 \cap \Lambda_2^{\perp}) \oplus H_3\}$, 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 $\int g^*(u, Z)dM_C(u, Z) \in \Lambda_2$. Following the similar arguments above as in (A4), we deduce that

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

Finally, since $\mathcal{E}, \mathcal{C} \perp \Lambda_3$, 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 $(H_3 \cap \Lambda_3^{\perp})$ by $\mathcal{R}$, consisting of all elements $h_3(Z, X_1, X_2) \in H_3$ that are orthogonal to all elements in Λ_3. After a little algebra, we have that $\mathcal{R} = [a(Z, X_1) : E\{a(Z, X_1)|X_1\} = 0]\cdot$ Since Z is a binary indicator, the space $\mathcal{R}$ can be equivalently written as

$$\mathcal{R} = [(Z - \pi)f(X_1) \text{ for any function } f(X_1)].$$

□

Sketch proof of Theorem 3. We first prove the following lemma.

LEMMA A2. *For any function $b(u, Z)$,*

$$\frac{\Delta\int b(u, Z)dM_T(u, Z)}{K_C(U, Z)} + \int \frac{dM_C(u, Z)}{K_C(u, Z)}E\left\{\int b(u, Z)dM_T(u, Z)\big|T \geqslant u, Z\right\}$$
$$= \int \frac{b(v, Z)}{K_c(v, Z)}dM(v, Z),$$

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

Proof. Since

$$E\{I(T = v)|T \geqslant u, Z\} = I(v \geqslant u)e^{\beta_0 Z}\lambda_0(v)dv\,\mathrm{pr}(T \geqslant v, Z)/\mathrm{pr}(T \geqslant u, Z)$$

and

$$E\{I(T \geqslant v)|T \geqslant u, Z\} = I(v < u) + I(v \geqslant u)\mathrm{pr}(T \geqslant v, Z)/\mathrm{pr}(T \geqslant u, Z),$$

we have that

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

This implies that

$$\int \frac{dM_C(u,Z)}{K_C(u,Z)} E\left\{ \int b(u,Z) dM_T(u,Z) \Big| T \geqslant u, Z \right\}$$
$$= \frac{\Delta - 1}{K_C(U,Z)} \int_0^U b(v,Z) e^{\beta_0 Z} \lambda_0(v) dv + \int_0^U \frac{\lambda_C(u,Z)}{K_C(u,Z)} du \int_0^u b(v,Z) e^{\beta_0 Z} \lambda_0(v) dv$$
$$= \frac{\Delta}{K_C(U,Z)} \int b(v,Z) I(v \leqslant U) e^{\beta_0 Z} \lambda_0(v) dv - \int \frac{b(v,Z)}{K_c(v,Z)} I(v \leqslant U) e^{\beta_0 Z} \lambda_0(v) dv.$$

The last equality follows because

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

This proves the lemma. □

Now we begin to prove Theorem 3. From semiparametric theory (Tsiatis, 2006, § 4), we know that finding the efficient influence function within the subclass is equivalent to finding functions $f_0(X_1)$ and $g_0(u, Z, X)$ such that

$$0 = E\left\{ \left(\int \{a(u,Z) - a^*(u)\} dM(u,Z) - (Z - \pi) f_0(X_1) \right.\right.$$
$$\left. + \int dM_C(u,Z)[g_0(u,Z,X) - E\{g_0(u,Z,X) | T \geqslant u, Z\}] \right) \tag{A5}$$
$$\left.\times \left((Z - \pi) f(X_1) + \int dM_C(u,Z)[g(u,Z,X) - E\{g(u,Z,X) | T \geqslant u, Z\}] \right) \right\},$$

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

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

for any functions $f(X_1)$ and $g(u, Z, X)$.

Using Lemma A2 and the result of Fleming & Harrington (1991) for martingale stochastic integrals, we have that

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

$$= E \int \Big\{ \lambda_C(u, Z) du [g(u, Z, X) - E\{g(u, Z, X) | T \geqslant u, Z\}]$$
$$\times \big(- E \left\{ m_1^F(T, Z) \middle| T \geqslant u, Z, X \right\} + E \left\{ m_1^F(T, Z) \middle| T \geqslant u, Z \right\}$$
$$+ K_C(u, Z)[g_0(u, Z, X) - E\{g_0(u, Z, X) | T \geqslant u, Z\}] \big) \Big\},$$

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

Therefore, equation (A5) is equivalent to

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

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

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

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

Note that $E\{dN(v) | T \geqslant u, Z, X\} = \mathrm{pr}(T = v | T \geqslant u, Z, X) K_C(v, Z) dv$, and $E\{Y(v) | T \geqslant u, Z, X\} = \mathrm{pr}(T \geqslant v | T \geqslant u, Z, X) K_C(v, Z)$, which implies that $E\{M(v, Z) | T \geqslant u, Z, X\} = E\{M_T(v, Z) | T \geqslant u, Z, X\} K_C(v, Z)$. Hence $E\{m_1^F(T, Z) | T \geqslant u, Z, X\} = E\{\int \{a(u, Z) - a^*(u)\} dM(u, Z) | T \geqslant u, Z, X\}$, which completes the proof of Theorem 3. □

REFERENCES

BEGUN, J. M., HALL, W. J., HUANG, W. M. & WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11**, 432–52.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. & WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore: The Johns Hopkins University Press.

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

COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–75.

FLEMING, T. R. & HARRINGTON D. P. (1991). *Counting Processes and Survival Analysis.* New York: John Wiley and Sons.

HAMMER, S. M., KATZENSTEIN, D. A., HUGHES, M. D. ET AL. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200-500 per cubic millimeter. *New Engl. J. Med.* **335**, 1081–90.

MANTEL, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemo. Rep.* **50**, 163–70.

NEWEY, W. K. (1990). Semiparametric efficiency bounds. *J. Appl. Economet.* **5**, 99–135.

PETO, R. & PETO, J. (1972). Asymptotically efficient rank invariant procedures. *J. R. Statist. Soc.* A **135**, 185–207.

TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data.* New York: Springer.

[*Received June* 2006. *Revised February* 2007]