# Locally Efficient Estimation of Regression Parameters Using Current Status Data

Chris Andrews, Mark van der Laan, and Jamie Robins*

May 15, 2001

**Abstract**

In biostatistics applications interest often focuses on the estimation of the distribution of a time-variable $T$. If one only observes whether or not $T$ exceeds an observed monitoring time $C$, then the data structure is called current status data, also known as interval censored data, case I. We consider this data structure extended to allow the presence of both time-independent covariates and time-dependent covariate processes which are observed until the monitoring time $C$. We assume that the monitoring process satisfies coarsening at random.

Our goal is to estimate the regression parameter $\beta$ of the regression model $\log(T) = Z^\top \beta + \epsilon$ where the conditional density of the error $\epsilon$ given $Z$ is assumed to have location parameter equal to zero. Because of the curse of dimensionality no globally-efficient non-parametric estimator with good practical performance at moderate sample sizes exists. We present a one-step estimator of the parameters $\beta$ which is guaranteed to be consistent and asymptotically normal if we have correctly specified a parametric or semiparametric model for the monitoring mechanism. Furthermore, our estimator attains the semiparametric efficiency bound for our model if we correctly specify a lower-dimensional model for the conditional distribution of $T$ given the covariates. Our estimator remains consistent and asymptotically normal even if this latter submodel is misspecified. In addition, we present a locally efficient doubly robust estimator which has the additional property that

**** Chris Andrews is Research Fellow of Biostatistics, University of California, Berkeley, CA 94720. Mark van der Laan is Professor of Biostatistics and Statistics, University of California, Berkeley, CA 94720. James Robins is Professor of Biostatistics and Epidemiology, Harvard University, Boston MA. This research was supported in part by National Science Foundation grant DMS 97-05993 and a FIRST award from the National Institute of General Medical Sciences, National Institute of Health.

it remains consistent and asymptotically normal if either the monitoring mechanism or the condtional distribution of $T$ given the covariates is correctly specified. We conclude with a simulation experiment and a data analysis.

**KEY WORDS:** Extended Current Status Data, Asymptotically Linear Estimator, Influence Curve, Asymptotically Efficient, Regression, Coarsening at Random, One-step Estimator.

# 1 Introduction

## 1.1 Regression with Current Status Data

Consider a study in which interest lies in the distribution of a random variable, $T$, that is never observed. Rather, for each individual, we observe at a random monitoring (censoring) time, $C$, whether $T$ exceeds $C$. This data structure $(C, \Delta = I(T \leq C))$ is called *current status data*. Our goal is to estimate the parameter vector $\beta$ of the regression model $T = Z^\top \beta + \epsilon$ where $Z$ is a vector of time-independent covariates, the conditional distribution of the error $\epsilon$ given $Z$ has location parameter equal to zero but has an otherwise non-restricted conditional distribution. In addition to data on $Z$, data on additional time-independent and dependent covariate processes up till time $C$, denoted by $\bar{L}(C) = \{L(s) : s \leq c\}$, may be available which explain any dependence between the time $T$ and the monitoring time $C$ and might be used to improve estimation of $\beta$.

A more classical setting of this estimation problem would be to say that we observe current status data $(C_1, I(T_1 \leq C_1), Z, \bar{L}_1(C_1))$ on a chronological time variable $T_1$, while we are willing to assume that the regression model holds for $T \equiv m(T_1)$, where $m$ is a given monotone transformation. For example, if $m(x) = \ln(x)$, then this regression model includes the accelerated failure time model $\ln T_1 = \beta^\top Z + \epsilon$, $\epsilon$ independent of $Z$, as a submodel. This setting is transformed to our setting by simply replacing the observed data $(C_1, I(T_1 \leq C_1), Z, \bar{L}_1(C_1))$ by the equivalent $(C \equiv m(C_1), I(T \leq C), Z, \bar{L}(C))$, where $L(s) \equiv L_1(m(s))$.

Note that we do not specify a parametric family for the error distribution. Furthermore, we do not assume that the error $\epsilon$ is independent of $Z$. Rather, we only assume that the conditional distribution of $\epsilon$ given $Z$ has a specified location parameter equal to zero. That is, in order to make $\beta$ identifiable, we assume

$$E[K(\epsilon) \mid Z] = E[K(T - Z^\top \beta) \mid Z] = 0 \quad (1)$$

where $K(\cdot)$ is a known, monotone function. If $K(\epsilon) = \epsilon$, then equation (1) implies the conditional mean given $Z$ of the error distribution is zero. However, estimation of the mean is quite difficult with current status data because the distribution of the monitoring mechanism must extend as far as the tails of the distribution of $T$. Thus other measures of center may be advantageous or necessary.

The conditional median model is obtained when $K(\epsilon) = I(\epsilon < 0) - 1/2$. Our estimators require a smoother $K(\cdot)$ than this because the median is not $\sqrt{n}$-estimable. In fact, our proof of the asymptotic properties of our estimators require that $K$ is twice differentiable a.e. (w.r.t. support of $C - Z^\top \beta$). A convenient family is $K(\cdot) = 2\Phi(\cdot) - 1$ where $\Phi$ is a (typically symmetric, mean zero) continuous distribution function. If the mass of $\Phi$ is concentrated near zero, we have a "smoothed median"; if $\Phi$ has large variance, we have a trimmed mean. We propose to choose a $K$ with compact support $[-\tau, \tau]$ for some user supplied $\tau$.

To fix idea, consider the following ideal mouse tumorigenicity experiment designed to investigate the relationship between the time, $T$, until the development of liver adenoma and the dose level, $Z$, of a suspected tumorigen. Suppose study mice are randomly allocated to dose groups and that liver adenomas are never, in themselves, the primary cause of an animal's death. Therefore, each mouse is sacrificed (monitored) at a random time $C$; at autopsy it is determined whether a tumor has developed before $C$. In such studies, it is easy to collect daily measurements of the weight of each mouse prior to sacrifice. Let $L(u)$ be the weight at time $u$ and let $\overline{L} = L(\cdot)$ be the entire weight process. Only the weight process up to time $C$ is observed: $\overline{L}(C) = \{L(u) : 0 < u < C\}$. Thus for each individual $Y = (C, \Delta = I(T \leq C), Z, \overline{L}(C))$ is observed, which we consider as a censored observation of the full data $X = (T, Z, \overline{L})$. Because mice with liver adenomas tend to lose weight, $\overline{L}(C)$ and $T$ are associated.

One reasonable monitoring scheme is to increase the hazard of monitoring shortly after a mouse begins to lose weight. If the time of sacrifice can be made closer to the time of tumor onset then more efficient estimation is possible. This monitoring scheme introduces dependence between $C$ and $T$ and estimators that ignore this dependence will be biased. Collecting information on a surrogate process and allowing the censoring time to depend on it is a superior design to carcinogenicity experiments that require independent censoring.

In the mouse experiment the dependence between $C$ and $T$ is only through the observed covariates. That is, the hazard of censoring at time $t$, given the full (unobserved) data

$X = (T, Z, \overline{L})$, is only a function of $Z$ and the observed portion of the covariate process, $\overline{L}(t)$:

$$\lambda_C(t \mid X) = \lambda_C(t \mid Z, \overline{L}(t)). \tag{2}$$

This implies $G(\cdot \mid X)$, the conditional distribution function of $C$, satisfies coarsening at random (CAR) (Robins, 1993). Coarsening at random (CAR) was originally formulated by Heitjan and Rubin (1991) and generalized by Jacobsen and Keiding (1994) and Gill, *et al.* (1997).

Our proposed one-step estimator of $\beta$ is consistent and asymptotically normal if we succeed in consistently estimating $\lambda_C(\cdot \mid X)$ at a suitable rate under the assumption (2). One such case is the idealized experiment described above where $\lambda_C(t \mid Z, \overline{L}(t))$ is known by design because it is under the control of the investigator (so estimation of $\lambda_C(t \mid Z, \overline{L}(t))$ is not even necessary). In general, a correctly specified semiparametric model which admits a consistent estimator for $\lambda_C(t \mid Z, \overline{L}(t))$ can be used. In this paper, we emphasize modeling $\lambda_C(t \mid Z, \overline{L}(t))$ by a time-dependent Cox proportional hazards model:

$$\lambda_C(t \mid Z, \bar{L}(t)) = \lambda_0(t) \exp(\alpha^\top W(t)), \tag{3}$$

where $W(t)$ is a function of $(Z, \bar{L}(t))$. In van der Laan, Robins (1998) it is explained why modelling the monitoring mechanism under CAR is a sensible approach to fight the curse of dimensionality in high dimensional models, which will not be repeated here. Our model for the observed data distribution is now specified since the observed data distribution $P_{F_X,G}$ of $Y$ is indexed by the full data distribution $F_X$ which needs to satisfy the regression model (1) and the conditional distribution $G(\cdot \mid X)$ which needs to satisfy a semiparametric model such as (3).

To have identifiability of $\beta$, we need to assume the conditional density function $g(\cdot \mid X)$ of the monitoring process is located correctly relative to the support of $T$ and the location parameter $K$. To start with, we assume that the support of $g(\cdot \mid X)$ is an interval, say, $(\alpha_X, \alpha^X)$ and that for $F_X$ a.e. $X = (T, Z, \bar{L})$ either $T \in (\alpha_X, \alpha^X)$ or the support of $K(\cdot - Z^\top \beta)$ is contained in $(\alpha_X, \alpha^X)$: if $K(x)$ has support $[-\tau, \tau]$, then the latter holds if $(Z^\top\beta - \tau, Z^\top\beta + \tau) \subset (\alpha_X, \alpha^X)$. This assumption is enough to define unbiased estimating functions for $\beta$ with nuisance parameter $g$, as shown in section 2. By CAR we have that the left-support point $\alpha_X$ of $g(\cdot \mid X)$ can only depend on the baseline covariates $L(0)$ and also $\alpha^X$ is a function of the observed data $Y$. To prove asymptotic consistency and efficiency results we need to bound away $g(\cdot \mid X)$ away from zero on the set

$$A(Z, L) \equiv \{c : c \in (\alpha_X, \alpha^X), K'(c - Z^\top \beta) > 0, \bar{F}(c \mid Z, \bar{L}(c)) > 0\},$$

where $\bar{F}(c \mid Z, \bar{L}(c)) = P(T > c \mid Z, \bar{L}(c))$. To be specific, we assume that for $F_X$ a.e. $(Z, L)$

$$\inf_{c \in A(Z,L)} g(c \mid X) > \delta > 0 \text{ for some } \delta > 0. \tag{4}$$

Because $g(c \mid X)$ must be bounded away from zero when both $K'(c - Z^\top \beta)$ and $\bar{F}(c \mid \bar{L}(c), Z)$ are non-zero, it is necessary that $K'$ have finite support when $T$ is unbounded. When $T$ has finite support, then $K$ may be the identity function.

If we use trimmed regression in the sense that $K$ has compact support $[-\tau, \tau]$ and $(Z^\top \beta - \tau, Z^\top \beta + \tau) \subset (\alpha_X, \alpha^X)$, then there is no need to monitor the tails of the distribution of $T$ since

$$A(Z, L) = \{c : c \in (Z^\top \beta - \tau, Z^\top \beta + \tau), \bar{F}(c \mid Z, \bar{L}(c)) > 0\}.$$

Our estimators in section 2 and 3 do rely on knowing $K(\alpha_X - Z^\top \beta)$ which equals $K(-\tau)$ under assumption (4) so that there is no need to know $\alpha_X$.

Our one-step estimator also uses an estimator of $F(t \mid Z, \bar{L}(u)) = P(T \le t \mid Z, \bar{L}(u))$ for various $u$'s and $t$. By the curse of dimensionality one will need to specify a lower dimensional working model for this conditional distribution and estimate it accordingly. The resulting one-step estimator is locally efficient in the sense that it is asymptotically efficient for our model if the working model contains the truth and it remains consistent and asymptotically normal otherwise. Thus our estimator uses time-dependent covariate information, such as the weight history of the mouse up till time $u$, to predict the time $T$ till onset thereby recovering information lost due to censoring. To illustrate the potential gain possible, if the weight process perfectly predicts $T$, then our estimator is asymptotically equivalent with the Kaplan-Meier estimator if we specify a correct model for $F(t \mid Z, \bar{L}(u))$.

Current practice is to sacrifice the mice at one point in time. Since our methodology shows that sophisticated mouse experiments can be nicely analyzed, we hope that experiments of the type above will be carried out in the future. In section 5 we will analyze a cross sectional study to estimate the time-till-transmission distribution in a previously analyzed HIV-partner study. In this data analysis we estimate the effects of "History of Sexually Transmitted Disease" and "Condom Use" in a model $\log(T) = \beta Z + \epsilon$, which thus includes the accelerated failure time model as a submodel, while using covariates outside the model to allow for informative censoring and improve efficiency. It is important to note that such an analysis is not possible with any of the existing methods since these methods assume that there are no relevant covariates outside the regression model (see next subsection).

### 1.2 Previous work and comparison with our results

There is a large literature on estimation of the distribution of $T$ with current status data when covariates are absent: Diamond, *et al.* (1986), Jewell and Shiboski (1990), Diamond and McDonald (1991), Keiding (1991), Sun and Kalbfleisch (1993), Groeneboom and Wellner (1992), Jewell *et al.* (1994), van de Geer (1994), Huang and Wellner (1995) and several others. van der Laan, Robins (1997) consider estimation of the distribution of $T$ with current status data in the presence of time-dependent covariate processes and time-independent covariates, using them to improve efficiency and allow for informative monitoring schemes.

Several authors have investigated estimation of regression parameters using current status data, $(C, \Delta)$, together with a time-independent covariate, $Z$ (Rabinowitz, *et al.*, 1995, Rossini and Tsiatis, 1996, Huang, 1997). Rabinowitz, *et al.* (1995) fit an accelerated failure time model $\ln T_1 = \beta^\top Z + \epsilon$ (for a failure time $T_1$) which requires error $\epsilon$ to be independent of the covariates $Z$. Huang (1997) derives an efficient estimator of the regression parameters of the proportional hazards model. Rossini and Tsiatis (1996) assume a semiparametric proportional odds regression model and carry out sieve maximum likelihood estimation. In each case the monitoring time may depend on the covariates of the model, $Z$, but not on additional covariates. Shen (2000) fits a linear regression model with current status data and time-independent covariates. In each of these references all covariates which explain the dependence between $C$ and $T$ must be included in the model for $T$. Because the models are for time-independent covariates only, no time-dependent covariates can be used to explain the dependence between $C$ and $T$. None of these limitations apply to our approach. In addition, our approach provides in general a mapping from full-data estimating functions to observed data estimating functions and thus provides the class of all estimators for any well understood full data model.

We would like to stress the implication of our results for the accelerated failure time model as studied by Rabinowitz, *et al.* (1995). Consider our model with the additional restriction on the regression model that $\epsilon$ is independent of $Z$. By monotone transforming the data (see introduction) it follows that this restricted model generalizes the problem of estimation of

$\beta$ in the accelerated failure time model of Rabinowitz, *et al.* (1995) based on current status data, namely by allowing the presence of additional time-dependent and time-independent covariates. The literature does not provide an estimator in this estimation problem. However, since this restricted model is a submodel of our model our locally efficient one step estimator (e.g. using as working model the accelerated failure time model) yields a closed form consistent and asymptotically normally distributed estimator of the regression parameters in the accelerated failure time model. This one-step estimator will be highly efficient in the accelerated failure time model and will remain consistent and asymptotically normal when the monitoring mechanism depends on the additional (time-dependent) covariates. Furthermore, it will still be consistent if the error distribution is not independent of $Z$, but $E(K(\epsilon) \mid Z) = 0$.

### 1.3 Organization of Paper

The next two sections are the heart of the paper. In section 2 we introduce an initial estimator and develop a one-step adjustment which produces our locally efficient estimator. This section also contains details for implementing the estimators and an introduction to the ideas of efficiency theory and one-step estimation. We also point out that by iterating the one-step procedure to solve the corresponding estimating equation yields a double robust estimator which is consistent if either $F(t \mid Z, \bar{L}(u))$ or $g(\cdot \mid X)$ is correctly estimated. In section 3 (and the appendices) we prove consistency, asymptotic linearity and local efficiency of our one-step estimator. Two simulations which demonstrate some asymptotic and finite sample properties of the estimators are presented in section 4. An analysis of the Califorinia Partners' Study of HIV infectivity is given in section 5 and finally we have some closing remarks.

## 2 Estimation

### 2.1 An Initial Estimator

Suppose we have $n$ independent observations, $Y_i = (C_i, \Delta_i = I(T_i \leq C_i), Z_i, \overline{L}_i(C_i))$. The component $Z_i$ is a vector and the component $\overline{L}_i(C_i)$ may itself have several components each of which may be a time-dependent process or a time-independent covariate. If none of the components of $\overline{L}$ is time dependent, we will indicate this by using $W_i$ rather than $\overline{L}_i(C_i)$.

In section A.1, it is shown that all unbiased estimating functions for $\beta$ in the model for the full data $(T, Z, \overline{L})$ are of the form $D(T, Z) = h(Z)K_\beta(T, Z)$ for some $h(Z)$, where $K_\beta(T, Z) = K(T - Z^T\beta)$. It follows that in the full data model one would define an estimator of $\beta$ by the solution of $0 = \sum_{i=1}^n h(Z_i)K_\beta(T_i, Z_i)$. Consider the following mapping on this set of full data estimating functions:

$$U_G(D)(Y) \equiv \frac{D'(C, Z)\overline{\Delta}}{g(C \mid X)} + D(\alpha_X, Z) = h(Z)\left(\frac{K'_\beta(C, Z)\overline{\Delta}}{g(C \mid X)} + K_\beta(\alpha_X, Z)\right) \tag{5}$$

where $D'$ is the derivative of $D$ with respect to the first argument, $K'_\beta(t, Z)$ is the derivative with respect to $t$ of $K(t - Z^\top\beta)$, $\alpha_X$ is the left endpoint in the support of $g(\cdot \mid X)$ and $\overline{\Delta} = 1 - \Delta$. Under assumption (4) we have that $K_\beta(\alpha_X, Z) = K(-\tau)$ is known.

Because the monitoring process satisfies CAR, the expression on the right is actually only a function of the observed data, $Y$. Belwo we show that this mapping satisfies for any $D(T, Z)$

$$E(U_G(D)(Y) \mid X) = D(X). \tag{6}$$

Therefore, for a given $D(T, Z) = h(Z)K_\beta(T, Z)$ and given consistent estimator $g_n(\cdot \mid X)$, we can use $U_{G_n}(D)$ as an estimating function for $\beta$.

As an initial estimator of the $k$-vector regression parameter $\beta$, we propose the solution, $\beta_n^0$, of the estimating equation corresponding with $h(Z) = Z$

$$\frac{1}{n}\sum_{i=1}^n Z_i\left(\frac{K'_{\beta_n^0}(C_i, Z_i)\overline{\Delta}_i}{g_n(C_i \mid X_i)} + K_{\beta_n^0}(\alpha_{X_i}, Z_i)\right) = 0. \tag{7}$$

Formal conditions for existence (Lemma 7) and $\sqrt{n}$-consistency (Lemma 8) of $\beta_n^0$ are given in Appendix B.

**Implementation:** To obtain this estimate of $\beta$ using equation (7), it is necessary to consistently estimate the conditional density of the censoring mechanism, $g(\cdot \mid X)$, from the data. We elected to use a time-dependent Cox proportional hazards model (3). For this model to estimate consistently the censoring density, the usual step-function estimate of the baseline hazard must be smoothed: e.g. as in Andersen et al. (1993). If we believe the censoring mechanism is independent of all covariates we can use a kernel smoother to estimate $g(\cdot \mid X) \equiv g(\cdot)$ In any case, after $g$ has been consistently estimated, the initial estimate $\beta_n^0$ then quickly can be found by numerical methods (e.g., Newton-Raphson) using equation (7).

**Derivation of (6):** Demonstration of condition (6) for the estimating function to be unbiased illuminates what conditions are necessary on the support of $g(\cdot \mid X)$. For clarity, the support of $g(\cdot \mid X)$ for each $X$, is assumed to be an interval, $(\alpha_X, \alpha^X)$. Because $U_G(D)(Y) = h(Z)U_G(K_\beta)(Y)$, condition (6) holds for all $D$ if it holds for $K_\beta$.

$$
\begin{aligned}
E(U_G(K_\beta)(Y) \mid X) &= \int_{\alpha_X}^{\alpha^X} \left( \frac{K'_\beta(c, Z) I(T > c)}{g(c \mid X)} + K_\beta(\alpha_X, Z) \right) g(c \mid X)\, dc \\
&= \int_{\alpha_X}^{\alpha^X} K'_\beta(c, Z) I(c < T)\, dc + K_\beta(\alpha_X, Z) \\
&= K_\beta(T, Z) - K_\beta(\alpha_X, Z) + K_\beta(\alpha_X, Z) = K_\beta(T, Z).
\end{aligned}
$$

The third equality clearly holds if $T \in (\alpha_X, \alpha^X)$. However, it also holds if $K_\beta(\cdot, Z)$ is constant outside $(\alpha_X, \alpha^X)$. Therefore we recommend using a $K$ which is contant outside $[-\tau, \tau]$ for some $\tau$ so that $K_\beta(\cdot, Z)$ is constant outside the interval $(Z^\top\beta - \tau, Z^\top\beta + \tau)$. Then a sufficient condition for $E(U_G(K_\beta)(Y) \mid X) = K_\beta(X)$ is (4).

### 2.2 The One-Step Estimator

An estimator $\beta_n$ of $\beta$ is asymptotically linear at the observed data distribution $P_{F_X,G}$ with influence curve $IC(Y \mid \beta, F_X, G)$ if $\beta_n - \beta = n^{-1}\sum_{i=1}^n IC(Y_i \mid \beta, F_X, G) + o_P(n^{-1/2})$. A regular estimator attains the semiparametric information bound at $P_{F_X,G}$ if its influence curve at $P_{F_X,G}$ is the so called efficient influence curve, $\ell^*_{\text{eff}}$, which is the solution of a mathematical problem solved in the Appendix. The efficient influence curve is also called the canonical gradient and it is orthogonal to all nuisance scores of $\beta$.

To construct a locally efficient estimator we add to $\beta_n^0$ the empirical mean of an estimate of the efficient influence curve at the true data generating distribution $P_{F_X,G}$:

$$
\beta_n^1 \equiv \beta_n^0 + \frac{1}{n}\sum_{i=1}^n \widehat{\ell^*_{\text{eff}}}(Y_i \mid \beta_n^0). \tag{8}
$$

Here $\beta_n^1$ is just the classical one-step estimator as defined in BKRW (page 395); that is, by its definition, $\beta_n^1$ is the first step in the Newton-Raphson algorithm for solving the estimating equation

$$
0 = \frac{1}{n}\sum_{i=1}^n \widehat{\ell^*_{\text{eff}}}(Y_i \mid \beta) \tag{9}
$$

for $\beta$, where we chose $\beta_n^0$ as the initial estimator.

The set of all influence curves (gradients) of the observed data model defined by the regression model (1) and (2) on $G$ (so no additional model on $g$ is imposed) can be represented as the range of a mapping $D \to IC(D) \equiv IC_0(D) - IC_{nu}(D)$ on the full data influence curves (gradients), which will be specified below. The first term is an influence curve of $\beta$ in the model with censoring density, $g(\cdot \mid X)$, being known and is given by

$$IC_0(Y \mid G, D) \equiv \frac{D'(C,Z)\overline{\Delta}}{g(C \mid X)} + D(\alpha_X, Z) \tag{10}$$

where $D \in \{h(Z)K_\beta(T,Z) : h\}$ is an appropriate influence curve of the full data model.

The second term is the projection of $IC_0$ on the tangent space $\{V \in L_0^2(P_{F_X,G}) : E[V(Y) \mid X] = 0\}$ of the monitoring process only assuming CAR (i.e. (2)). It is given by

$$\begin{aligned} IC_{nu}(Y \mid F, G, D) \quad \equiv \quad & \int_0^\infty \left( \frac{D'(u,Z)\overline{F}(u \mid Z, \overline{L}(u))}{g(u \mid X)} \right. \\ & \left. - \frac{1}{\overline{G}(u \mid X)} \int_u^\infty D'(t,Z)\overline{F}(t \mid Z, \overline{L}(u))dt \right) dM(u) \end{aligned} \tag{11}$$

where $F(\cdot \mid Z, \bar{L}(u))$ is the conditional cumulative distribution of $T$, given $(Z, \bar{L}(u))$ and $dM(u) = I(C \in du) - \Lambda_C(du \mid X)I(C > u)$. For a given cumulative distribution $F$ we defined $\bar{F} = 1 - F$. For convenience, we used shorthand $F$ in $IC_{nu}(Y \mid F, G, D)$ for $F(\cdot \mid Z, \bar{L}(u))$ for various $u$. We define $IC(Y \mid F, G, D) \equiv IC_0(Y \mid G, D) - IC_{nu}(Y \mid F, G, D)$.

If $\overline{L} = W$ is time independent, then

$$\begin{aligned} & IC_{nu}(Y \mid F, G, D) \\ = \quad & \frac{D'(C,Z)\overline{F}(C \mid Z, W)}{g(C \mid X)} - \int_0^\infty D'(u,Z)\overline{F}(u \mid Z, W)du \end{aligned} \tag{12}$$

in which case

$$IC(Y \mid F, G, D) = \frac{D'(C,Z)}{g(C \mid X)}(F(C \mid Z, W) - \Delta) + E[D(T,Z) \mid Z, W]. \tag{13}$$

The efficient influence curve for estimation of $\beta$ in the observed model, $\ell^*_{\text{eff}}$, equals $IC(Y \mid F, G, D_{opt})$ where the optimal $D_{opt}$ is derived in Appendix A; see Theorem 2. We have $D^*_{\text{opt}}(T,Z) \equiv c^{-1}_{\text{opt}} h_{\text{opt}}(Z)K_\beta(T,Z)$, where $c_{\text{opt}}$ and $h_{\text{opt}}$ are defined in the next expression for the efficient influence curve:

$$\begin{aligned} & \ell^*_{\text{eff}}(Y \mid F, G, h_{\text{opt}}, c_{\text{opt}}, \beta) \\ \equiv \quad & IC(Y \mid F, G, D^*_{\text{opt}}) \\ = \quad & c^{-1}_{\text{opt}} h_{\text{opt}}(Z) IC(Y \mid F_X, G, K_\beta) \\ = & \left[ E\left( Z h_{opt}(Z) U_G(K'_\beta) \right) \right]^{-1} \frac{Z E(K'_\beta \mid Z)}{\phi(Z)} IC(Y \mid F_X, G, K_\beta) \end{aligned} \tag{14}$$

where $h_{opt}(Z) = (ZE(K'_\beta \mid Z))/\phi(Z)$ with $\phi(Z) = E(IC(Y \mid F, G, K_\beta)^2 \mid Z)$. If $\overline{L} = W$ is time-independent, then $\phi$ simplifies to the expression

$$\phi(Z) = E\left( E(K_\beta \mid Z, W)^2 + \int \frac{F(t \mid Z, W)\overline{F}(t \mid Z, W){K'_\beta}^2(t, Z)}{g(t \mid X)} dt \,\middle|\, Z \right). \tag{15}$$

and if $(Z, \overline{L}) = Z$, then it simplifies further to

$$\phi(Z) = \int \frac{F(t \mid Z)\overline{F}(t \mid Z){K'_\beta}^2(t, Z)}{g(t \mid X)} dt. \tag{16}$$

In the one-step estimator we estimate $\ell^*_{\text{eff}}(Y \mid F, G, h_{\text{opt}}, c_{\text{opt}}, \beta)$ by substitution of estimators $F_n, G_n, h_n, c_n, \beta^0_n$ for $F, G, h_{\text{opt}}, c_{\text{opt}}, \beta$, respectively. Here

$$c_n = -\frac{1}{n} \sum_{i=1}^n h_n(Z_i) Z_i^\top U_{G_n}(K'_{\beta^0_n})(Y_i). \tag{17}$$

Thus

$$\beta^1_n = \beta^0_n + \frac{1}{n} \sum_{i=1}^n \widehat{\ell^*_{\text{eff}}}(Y_i \mid F_n, G_n, h_n, c_n, \beta^0_n).$$

Note that $c_n$ plays the role of the derivative matrix of the estimating equation (9) in $\beta$ as needed in the first step of Newton-Raphson algorithm.

### 2.3 Implementation Issues

The following estimation method can always be used to compute the one-step estimator. As illustrated in Example 2, more specific information about the structure of the model can improve efficiency for finite samples.

Estimation of the efficient influence curve (14) involves estimation of $(F, G, h_{opt}, c_{opt}, \beta)$. The initial estimate of $\beta$ and the estimate of the censoring density $g$ were discussed in section 2.1. We now discuss each of the other three in turn and how to compute the one-step estimator $\beta^1_n$.

***$F_n$* for time-independent case.** If $\overline{L} = W$ is time-independent, then $IC_{nu}$ is given by equation (12). The following identity can be used to estimate $F(\cdot \mid Z, W)$:

$$F(t \mid Z, W) = E[I(T \leq t) \mid Z, W] = E[\Delta \mid C = t, Z, W] \tag{18}$$

where $\Delta = I(T \leq C)$. The second equality follows from CAR.

The proposed submodel can be chosen to be a highly parametric model or a flexible semiparametric model. The former leads to an efficient estimator in fewer circumstances.

Nonetheless, the finite sample performance of a parametric model is comparable if not superior to a semiparametric model because it recognizes the main effects of the covariates and is more stable where the data are sparse. This comparison is made in the second example in the simulation section.

One possible semiparametric model for $F(\cdot \mid Z, W)$ is a logistic generalized additive model.

$$\begin{aligned} F(t \mid Z, W) &= E[\Delta \mid C = t, Z = (Z_1, \ldots, Z_k), W = (W_1, \ldots, W_l)] \\ &= \frac{\exp(f_C(t) + f_{Z_1}(Z_1) + \cdots + f_{W_l}(W_l))}{1 + \exp(f_C(t) + f_{Z_1}(Z_1) + \cdots + f_{W_l}(W_l))}. \end{aligned} \tag{19}$$

The Splus function `gam` with `family=binomial(link=logit)` produces an $F_n$ based on the observed data $\{Y_i\}_{i=1}^n$. Of course, the probit model (`family=binomial(link=probit)`) can also be used. Furthermore, some or all of the general functions $f_C$, $f_{Z_1}$, $\ldots$, $f_{W_l}$, can be replace by more parametric polynomials.

The factor $IC(Y_i \mid F, G, K_\beta)$ in equation (14) can now be estimated for each $Y_i$ using the expressions for $IC_0(Y_i \mid G_n, K_{\beta_n^0})$ and $IC_{nu}(Y_i \mid F_n, G_n, K_{\beta_n^0})$ given in (10) and (12). If $K'_{\beta_n^0}(\cdot, Z_i)$ is zero except on the interval $(Z_i^\top \beta_n^0 - \tau, Z_i^\top \beta_n^0 + \tau)$, the integral in equation (12) is easily approximated (e.g., the trapezoidal rule).

***$F_n$* for time-dependent case.** If $\overline{L}$ is time-dependent, $IC_{nu}$ must be estimated directly from equation (11). It is necessary to estimate $F(t \mid Z, \overline{L}(u))$ for a given $(t, u)$ with $t \geq u$. First consider the case where the density of $C$ depends only on the time-independent covariates (even though $F(t \mid Z, \overline{L}(u))$ may depend on the time-dependent covariates). Then we proceed using the CAR-identity

$$F(t \mid Z, \overline{L}(u)) = E(\Delta \mid C = t, Z, \overline{L}(u), C \geq u). \tag{20}$$

To avoid the curse of dimensionality, for each $u$ we replace $\overline{L}(u)$ by a vector of summary measures, $W_u(\overline{L}(u))$, which hopefully captures the most relevant information for predicting $T$. Now, for each $u$, we can estimate $F(\cdot \mid Z, \overline{L}(u)) \approx F(\cdot \mid Z, W_u)$ by the GAM in equation (19). The model is fit using data $Y_i$ for which $C_i \geq u$ (i.e., those observations for which $\overline{L}(u)$ is observed).

For the general case where the censoring mechanism also depends on the time-dependent covariate, the identity (20) is not guaranteed by CAR. However, wWe proceed in estimating $F(t \mid Z, \overline{L}(u))$ in two stages by noting that

$$F(t \mid Z, \bar{L}(u)) = E(F(t \mid Z, \bar{L}(t)) \mid Z, \bar{L}(u), C > u),$$

where, by CAR, $F(t \mid Z, \bar{L}(t)) = E(\Delta \mid C = t, Z, \bar{L}(t))$. Thus we can estimate $F(t \mid Z, \bar{L}(t))$ by fitting the GAM-model (19) with covariates $t, Z$ and covariates extracted from $\bar{L}(t)$for each $t$. Now for each $u$, regress $\hat{F}(t \mid Z_i, \bar{L}_i(t))$ on $Z_i$ and covariates extracted from $\bar{L}_i(u)$, using individuals for which $C_i \geq u$. The Splus function `gam` with `family=quasi(link=logit, variance=constant)` can be used to fit a logistic GAM model.

***$h_{\text{opt}}$*.** The vector-valued function $h_{\text{opt}}(Z)$ is proportional to $Z$. The constant of proportionality is the ratio of

$$E(K'_\beta \mid Z) = E\left(\frac{K''(C - \beta^\top Z)(1 - \Delta)}{g(C \mid X)} \mid Z\right) + K'(\alpha_X - \beta^\top Z) \tag{21}$$

and $\phi(Z)$. Using $\beta_n^0$ and $g_n(\cdot \mid X)$ to obtain an observed outcome, the expression (21) can be estimated by regressing an observed outcome on $Z$. The function $\phi(Z)$ can be estimated in several ways depending on the number and type of covariates are available. In general, $\phi(Z)$ is the conditional expectation given $Z$ of $IC^2(Y \mid \beta, F, G, K_\beta)$. An estimate of $IC$ has already been computed and its square can be regressed on $Z$ in some parametric or semiparametric method (e.g., splines, gam, running medians).

Although this regression method can always be used, in some cases $\phi(Z)$ has other expressions with more structure which can be exploited. In particular, if there are no covariates other than $Z$, then $\phi$ is given by equation (16) and can be estimated by substitution of an estimator of $F(t \mid Z) = E(\Delta \mid C = t, Z)$.

If $\overline{L} = W$ is time independent, $\phi(Z)$ is given by equation (15) which can thus be estimated by substitution of an estimator of $F(t \mid Z, W) = E(\Delta \mid C = t, Z, W)$. Equation (37) will be more accurate than equation (35) but potentially more computationally intensive. In example 2 the assumptions on $T \mid Z$ and $W \mid T, Z$ imply a distribution on $W \mid Z$ which can be exploited in computing $\phi(Z)$.

***$c_{\text{opt}}$*.** The normalizing matrix $c_{\text{opt}}$ is given by equation (41). The expectation can be estimated by the empirical mean. Each factor inside the expectation has already been estimated to obtain the estimate of $h_{\text{opt}}$.

# 3 Asymptotic Efficiency Theorem for One-Step Estimator

An estimator $\beta_n$ of $\beta$ is asymptotically linear at $P_{F_X,G}$ with influence curve $IC(Y \mid \beta, F_X, G)$ if $\beta_n - \beta = n^{-1} \sum_{i=1}^n IC(Y_i \mid \beta, F_X, G) + o_P(n^{-1/2})$. From Bickel, *et al.* (1993) we have that an estimator is asymptotically efficient if it is asymptotically linear with influence curve the so called efficient influence curve, $\ell^*_{\text{eff}}$, which is orthogonal to all nuisance parameters. The efficient influence curve is also called the canonical gradient and it is computed in the Appendix. It is given by $\ell^*_{\text{eff}}(Y \mid F, G, h_{opt}, c, \beta)$ as defined in section 2.

Theorem 1 below shows that if our model $F(t \mid Z, \overline{L}(u))$ is correctly specified, the one-step estimator $\beta_n^1$ is indeed asymptotically linear with influence curve $\ell^*_{\text{eff}}$ and thus is asymptotically efficient. Moreover, $\beta_n^1$ has the additional feature that it remains a consistent and asymptotically normal estimator of $\beta$ even when the model for $F(t \mid Z, \overline{L}(u))$ is misspecified. This is due to the fact that $IC_{nu} = \int H(u, Z, \overline{L}(u))dM(u)$ for a particular $H$ (equation (11)) and that for any function $H$, $\int H(u, Z, \overline{L}(u))dM(u)$ has mean zero, given $X$, because $E(dM(u) \mid X) = 0$. This explains why $\beta_n^1$ is consistent even if $F$ is estimated inconsistently.

This protection from model misspecification of $F$ follows from the general representation of $\ell^*_{\text{eff}}$ developed by Robins and Rotnitzky (1992) and further developed in van der Laan, Gill, Robins (2000). For further details about computing this representation we refer to the Appendix.

When the model for $F(t \mid Z, \overline{L}(u))$ is misspecified, the influence curve of $\beta_n^1$ depends on the model for the nuisance parameter $g(c \mid X)$. Characterization of this dependence requires we introduce the notion of a tangent space. Denote by $L_0^2(P_{F_X,G})$ the Hilbert space of functions of $(C, \Delta, Z, \overline{L}(C))$ with finite variance and mean zero endowed with the covariance inner product $< v_1, v_2 >_{P_{F_X,G}} \equiv \sqrt{\int v_1 v_2 dP_{F_X,G}}$. The tangent space $T_1 = T_1(P_{F_X,G})$ for the parameter $F_X$ is, by definition, the closure in $L_0^2(P_{F_X,G})$ of the linear extension of the scores at $P_{F_X,G}$ from correctly specified parametric models for the distribution $F_X$. The tangent space $T_2 = T_2(P_{F_X,G})$ for the parameter $G$ is the closure of the linear extension in $L_0^2(P_{F_X,G})$ of the scores at $P_{F_X,G}$ from all correctly specified parametric submodels (i.e., submodels of the assumed semiparametric model) for the distribution $G$.

With these preliminaries, we are nearly ready to state our main theorem. Before doing so, we note that condition (2) in the theorem below is a general empirical process condition. For empirical process theory we refer to van der Vaart and Wellner (1996). We decided not to derive more primitive conditions that imply condition (2) because it is technical and model dependent. Condition (1) assures the initial estimator exists and is $\sqrt{n}$-consistent and that the structural condition (23) as needed in the proof holds. Condition (3) requires that $g_n$ converges uniformly to $g$ over a set $A$ and that $F_n(t \mid Z, \bar{L}(u))$ converges uniformly to something (not necessarily the truth) over a set $B$, where $A$ and $B$ are intersections of the support of $g$ and $K$: in other words, one only needs convergence over sets at which $F$ and $g$ are identifiable (under condition (1)). In addition, condition (3) requires that the product of the rates is $o_P(1/\sqrt{n})$. Condition (4) requires that one uses an efficient procedure for estimation of the monitoring mechanism $g(c \mid X)$ such as a maximum likelihood estimator.

**Theorem 1** *Assume*

- *Conditions of lemma 1 hold.*
- $\ell^*_{\text{eff}}(\cdot \mid F_n, G_n, h_n, c_n, \beta^0_n)$ *is contained in a* $P_{F_X,G}$*-Donsker class with probability tending to one.*
- *For some* $F^\dagger$ *we have that*

$$\begin{aligned} r_{1n} &\equiv \sup_A |g_n(u \mid X) - g(u \mid X)| \to 0 \\ r_{2n} &\equiv \sup_B \left| F_n(t \mid Z, \overline{L}(u)) - F^\dagger(t \mid Z, \overline{L}(u)) \right| \to 0 \\ \sqrt{n} r1n r_{2n} &\to 0, \end{aligned}$$

*where the uniform convergence statements need to hold in probability over the sets*

$$\begin{aligned} A &\equiv \{(u, Z, \bar{L}(u)) : K'(u - Z^\top \beta) > 0, \max(\bar{F}^\dagger, \bar{F}_n, \bar{F})(u \mid Z, \bar{L}(u)) > 0\} \\ B &\equiv \{(t, Z, \bar{L}(u)) : t \geq u, K'(t - Z^\top \beta) > 0, \max(g_n, g)(u \mid Z, \bar{L}(u)) > 0\}. \end{aligned}$$

- $\Phi(G_n) \equiv E(\ell^*_{\text{eff}}(Y \mid F^\dagger, G_n, h, c, \beta))$ *is an efficient estimator of* $\Phi(G)$.

*Then* $\beta^1_n \equiv \beta^0_n + n^{-1} \sum_{i=1}^n \ell^*_{\text{eff}}(Y_i \mid F_n, G_n, h_n, c_n, \beta^0_n)$ *is asymptotically linear with influence curve*

$$\Pi \left[ \ell^*_{\text{eff}}(\cdot \mid F^\dagger, G, h, c, \beta) \middle| T_2^\perp(P_{F_X,G}) \right], \tag{22}$$

*where* $T_2(P_{F_X,G})$ *is the tangent space for the chosen CAR-model containing the true* $G$ *and the matrix* $c$ *is the limit of* $c_n$*. Furthermore, if* $h = h_{\text{opt}}$ *and* $F^\dagger = F$ *(so that* $\ell^*_{\text{eff}}(\cdot \mid F^\dagger, G, h, c, \beta) = \ell^*_{\text{eff}}(\cdot \mid F, G, h_{\text{opt}}, c_{\text{opt}}, \beta)$*), then* $\beta^1_n$ *is asymptotically efficient.*

**Proof:** This general proof is analogue to the proof of the result of van der Laan and Robins (1997) for the estimation of smooth functionals of $F_T$. We have

$$\begin{aligned} \beta^1_n - \beta &= \beta^0_n - \beta + (P_n - P_{F_X,G}) \ell^*_{\text{eff}}(Y \mid F_n, G_n, h_n, c_n, \beta^0_n) \\ &+ P_{F_X,G} \ell^*_{\text{eff}}(Y \mid F_n, G, h_n, c_n, \beta^0_n) + P_{F_X,G}\{\ell^*_{\text{eff}}(Y \mid F_n, G_n, h_n, c_n, \beta^0_n) - \ell^*_{\text{eff}}(Y \mid F_n, G, h_n, c_n, \beta^0_n)\}, \end{aligned}$$

where we used the notation $Pf \equiv \int f(y)dP(y)$. The right-hand side is a sum of four terms. Condition (2) and (3) imply that (see e.g. van der Laan, Robins, 1997) the second term equals

$$(P_n - P_{F_X,G})\ell^*_{\text{eff}}(Y \mid F^\dagger, G, h, c, \beta) + o_P(n^{-1/2}).$$

Lemma 1 below shows that under condition (1) we have that the third term equals

$$\beta - \beta^0_n + o_P(n^{-1/2}).$$

Condition (3) implies that the fourth term equals

$$\Phi(G_n) - \Phi(G) + o_P(n^{-1/2}).$$

Condition (4) implies that $\beta^1_n$ is asymptotically linear with influence curve $\ell^*_{\text{eff}}(Y \mid F^\dagger, G, h, c, \beta) + IC_{nuis}(Y)$, where $IC_{nuis}(Y)$ is the influence curve of $\Phi(G_n)$ as an estimator of $\Phi(G)$. Finally, as in van der Laan, Robins (1998) the fact that $IC_{nuis}$ is actually an efficient influence curve in a model for $G$ with tangent space $T_2$ (by (4)) implies that this influence curve can be represented as a projection given by (22). □

**Lemma 1** *Let* $h_n$ *be given,* $\beta^0_n$ *the initial estimator defined in section 2,* $c_n = -\frac{1}{n}\sum_{i=1}^n h_n(Z_i)Z_i^\top U_{G_n}(K'_{\beta^0_n})(Y_i)$.

*Suppose that the following conditions on the true data generating distribution hold:* $(\partial/\partial\beta)E(ZU_G(K_\beta)(Y))$ *is invertible at the true value of* $\beta = \beta_0$*,* $g'$ *and* $K''$ *are bounded above, condition (4) holds for* $\beta \in N(\beta_0)$*, where* $N_{\beta_0}$ *is an arbitrarily small neighborhood of* $\beta_0$.

*In addition, we make the following consistency assumptions:* $\parallel g_n - g \parallel_{\infty,A} = O_P(n^{-1/4})$, $\parallel G_n - G \parallel_{\infty,A} = O_P(n^{-1/2})$, $h_n(Z)$ *converges uniformly to an arbitrary* $h(Z)$ *with* $c(h) \equiv E(h(Z)Z^\top U_G(K'_\beta))$ *invertible.*

*Then* $c_n \equiv c(h_n)$ *converges to* $c(h) = -E(h(Z)Z^\top U_G(K'_\beta)(Y))$ *and (under no conditions on* $F_n$*)*

$$E\left(\ell^*_{\text{eff}}(\cdot \mid F_n, G, h_n, c_n, \beta^0_n)\right) = \beta - \beta^0_n + o_P(n^{-1/2}) \tag{23}$$

**Proof.** See the proof of lemma 9 in Appendix B.

### 3.1 Construction of Confidence Intervals

A confidence region for the parameter vector $\beta$ or individual confidence intervals for each regression parameter can be constructed by estimating the covariance matrix of the efficient influence function, $\ell^*_{\text{eff}}$. If the model for $F(t \mid Z, \overline{L}(u))$ is correctly specified, the vector $\sqrt{n}(\beta^1_n - \beta)$ is asymptotically distributed $N(0, \text{COV}(\ell^*_{\text{eff}}))$ because the projection operator in expression (22) is the identity operator in this case. Thus an asymptotic 95% confidence region for $\beta$ is

$$\left\{\beta \in \mathbb{R}^k \mid (\beta^1_n - \beta)^\top \widehat{\Sigma}^{-1}(\beta^1_n - \beta) \leq \frac{k}{n} F_{0.95,k,\infty}\right\}$$

(e.g., Morrison, 1990) where $\widehat{\Sigma}$ is the empirical variance of the estimated efficient influence function,

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^n \left(\widehat{\ell^*_{\text{eff}}}(Y_i) - \frac{1}{n}\sum_{i'=1}^n \widehat{\ell^*_{\text{eff}}}(Y_{i'})\right)\left(\widehat{\ell^*_{\text{eff}}}(Y_i) - \frac{1}{n}\sum_{i'=1}^n \widehat{\ell^*_{\text{eff}}}(Y_{i'})\right)^\top,$$

where we define $\widehat{\ell^*_{\text{eff}}}(Y) \equiv \ell^*_{\text{eff}}(Y \mid F_n, G_n, h_n, c_n, \beta^0_n)$. A 95% confidence interval for a single parameter is $\beta^1_n \pm 1.96\widehat{\sigma}/\sqrt{n}$, where $\widehat{\sigma}$ is the appropriate diagonal element of $\widehat{\Sigma}$.

If the model for $F(t \mid Z, \overline{L}(u))$ is misspecified the above confidence intervals are conservative. The true variance of the estimator is given by the variance of expression (22) which is smaller than the variance of $\ell^*_{\text{eff}}$. We refer to Robins and van der Laan (1997) and van der Laan, Robins (2001) for exact expressions when the model for the monitoring process is either Cox proportional hazards or independence. However, unless $F$ is very poorly specified, the conservative intervals will be fairly accurate.

### 3.2 A doubly robust estimator.

Given estimates $F_n, G_n$ of the nuisance parameters $F, G$ and a choice $h_n, c_n$ for the full data estimating function, consider the estimator $\beta_n$ solving

$$0 = \frac{1}{n}\sum_{i=1}^n \ell^*_{\text{eff}}(Y_i \mid F_n, G_n, h_n, c_n, \beta).$$

Recall that our one-step estimator is just the first step of the Newton-Raphson algorithm for solving this estimating equation for $\beta$, where we chose $\beta^0_n$ as initial estimator. Due to the orthogonality of $\ell^*_{\text{eff}}(Y \mid F, G, h, c, \beta)$ w.r.t. $G$, we actually have the following double-robustness property w.r.t. the nuisance parameters $F, G$ of this estimating function $\ell^*_{\text{eff}}(Y \mid F, G, h, c, \beta)$:

$$E\ell^*_{\text{eff}}(Y \mid F_1, G_1, h, c, \beta) = 0 \text{ if either } F_1 = F \text{ or } G_1 = G.$$

This means that, in fact, under regularity conditions, $\beta_n$ will be consistent and asymptotically linear if either the model for $F(t \mid Z, \bar{L}(t))$ is correctly specified or the model for $G$ is correctly specified. If one wants to obtain a confidence region for $\beta$ based on $\beta_n$ in this more nonparametric model than the one we assumed in this paper (i.e. that the model for $G$ is correctly specified), then we recommend to use the nonparametric bootstrap. This double robustness property of $\beta_n$ implies that, in practice, a minor misspecification of the model for $G$ can be corrected by doing a good job in modelling $F(t \mid Z, \bar{L}(t))$ and vica versa.

**Data adaptive selection of location parameter.** The regression parameter $\beta$ represents the effect of $Z$ on the location parameter identified by $K$. Thus the choice of location parameter affects immediately the interpretation of $\beta$ and could therefore just be subject-matter driven. However, one might also decide to choose the location parameter which is best identifiable from the data. Suppose that we choose a location parameter $K_\tau$ with compact support $[-\tau, \tau]$ which (e.g) approximates the median for $\tau \to 0$ and approximates the mean for $\tau \to \infty$. In that case, we propose to calculate $\widehat{\Sigma}$ for a range of $\tau$'s and select the $\tau$ which minimizes this estimated variance of the efficient influence curve. This corresponds with choosing the location parameter which results in the smallest confidence bands.

## 4 Simulations

Two simulation studies are presented to illustrate the applicability and efficiency of these methods. Example 1 demonstrates that the asymptotic properties of the one-step estimator apply to a dataset of moderate size. The superiority of the one-step estimator over the initial estimator is also shown. The effects of an additional time-independent covariate, $W$, and the submodel selected for $F(\cdot \mid Z, W)$ are considered in Example 2.

The function $K$ we use in these simulations is a smoothed truncated mean given by

$$K(t) = \begin{cases} -\tau & t < -\tau \\ t + \frac{\tau}{\pi}\sin(\frac{\pi t}{\tau}) & -\tau \leq t \leq \tau \\ \tau & \tau < t \end{cases} . \tag{24}$$

with $\tau = 3$. $K$ has two continuous derivatives, both of which are zero outside the interval $(-\tau, \tau)$.

### 4.1 Example 1: No Unmodeled Covariate.

The data generating distribution has $\beta = (\beta_0, \beta_1) = (0, 1)$, $Z_0 \equiv 1$ (intercept), $Z_1 \sim N(0, 1)$, $T \mid Z \sim N(Z_1, 1)$, and $C|Z \sim N(Z_1, 1)$. The observed data is $(C, \Delta, Z)$. The general method of estimation described in section 2.3 was used with the following specifics. The censoring distribution was estimated via linear regression of $C$ on $Z$ with independent normal error. The distribution of $T \mid Z$ was estimated using equation (18) and a generalized linear model with probit link. $h_{\text{opt}}$ was computed after approximating the integrals $E(K_{\beta_n^0}(T, Z) \mid Z) = -\int K''_{\beta_n^0}(t, Z)F(t \mid Z)dt$ and the expression (16) for $\phi(Z)$ by Simpson's Rule with 20 intervals. The results in Table 1 are based on 1000 repetitions.

The one-step estimator is efficient in this example because the submodel chosen for $F$ is correct. In finite samples we estimate the efficiency by comparing the variance of the estimator with the variance of the efficient influence curve. Similarly we estimate the efficiency of the one-step estimator relative to the initial-estimator. Results for the parameter $\beta_1$ at three sample sizes are given in Table 1 (similar patterns are seen for $\beta_0$). The asymptotic efficiency is evident in both of the larger samples.

| Estimator | Sample Size | Asymptotic Relative Efficiency | Relative Efficiency (baseline=Initial) |
|---|---|---|---|
| Initial ($\beta_{n,1}^0$) | 250 | 0.53 | 1 |
| | 500 | 0.76 | 1 |
| | 1000 | 0.81 | 1 |
| One-step ($\beta_{n,1}^1$) | 250 | 0.67 | 1.2 |
| | 500 | 1.05 | 1.3 |
| | 1000 | 1.05 | 1.3 |

Table 1: Comparison of initial and one-step estimators for simple linear regression example. One-step estimator is asymptotically efficient and appears to be fully efficient even for moderate sample sizes.

### 4.2 Example 2: Unmodeled Covariate

Suppose in addition to $Z$, another covariate $W$ has been collected which is associated with $T$. Our method uses the information contained in the covariate to improve the estimate of $\beta$. The strength of the relationship between $T$ and $W$ is one factor which determines how much our one-step estimator can improve the initial estimator which does not use $W$. In this example we consider three covariates: $W_1 = T$, $W_2 = T +$ small error, and $W_3 = T +$ large error. The first corresponds to a perfect surrogate for $T$, the second to a good predictor of $T$, and the third to a poor predictor of $T$.

The degree to which we will be able to exploit the information in $W$ also depends on the submodel we select for $F(\cdot \mid Z, W)$. It is frequently wise to be optimistic and select a small submodel; for example, a generalized linear model often outperforms a generalized additive model if linearity is at all reasonable. In this example we consider two one-step estimators. The first estimator is the generic method described in Section 2.3. The assumed model for $F(\cdot \mid Z, W)$ is correct for each of the three covariates. Thus $\beta_n^1$ is asymptotically efficient in each case.

The second one-step estimator assumes $W$ is a perfect surrogate for $T$ and thus "estimates" $F(t \mid W, Z)$ with $I(W < t)$. This is correct in the first scenario because $W_1 = T$ but not correct for the second or third case. Under the assumption $W = T$, one could directly estimate $\beta$ by linear regression: $W = Z^\top \beta + \epsilon$. This direct linear regression method is optimal in case 1 where $W_1 = T$. However, in the other two cases, this estimator is inconsistent. On the other hand, our estimator is consistent in each of the three cases and is asymptotically equivalent with this direct linear regression method if $W = T$.

The simulation results are presented in Table 2. The initial estimator is exactly the estimator in the previous example. It does not use the information provided by the covariate $W$ and thus is not nearly efficient. If $W$ is very informative, as in the first two cases, the variance bound is less than half the variance of the initial estimator.

The generic one-step estimator is efficient, but for samples with $N = 1000$ the variance bound is about 10% smaller than the variance of the estimator. The special one-step estimator which assumes $W = T$ reaches the efficiency bound (and then some) when $W$ is very informative. When $W$ is a poor predictor of $T$, the performance of this estimator suffers as should be expected because the assumption $W = T$ is bad. The variance of the special estimator is larger than the generic estimator in this case.

**Details.** The data generating distribution has $Z_0 \equiv 1$ (intercept), $Z_1 \sim N(0, 1)$, $T \mid Z \sim N(Z_1, 1)$, $W_1 \mid T, Z = T$, $W_2 \mid T, Z \sim N(T, 0.1^2)$, $W_3 \mid T, Z \sim N(T, 1.0^2)$, and $C \mid Z, W \sim N(Z_1, 1)$. The general method of estimation described in section 2.3 was used with

| Estimator | Available Covariate | Asymptotic Relative Efficiency | Relative Efficiency (baseline=Initial) | Relative Efficiency (baseline=Generic) |
|---|---|---|---|---|
| Initial ($\beta^0_{n,1}$) | $W_1 = T$ | 0.40 | 1 | |
| | $W_2$ | 0.44 | 1 | |
| | $W_3$ | 0.78 | 1 | |
| Generic ($\beta^1_{n,1}$) | $W_1 = T$ | 0.90 | 2.20 | 1 |
| | $W_2$ | 0.93 | 2.12 | 1 |
| | $W_3$ | 0.94 | 1.19 | 1 |
| Special ($\beta^{1*}_{n,1}$) | $W_1 = T$ | 1.03 | 2.32 | 1.15 |
| | $W_2$ | 1.08 | 2.48 | 1.08 |
| | $W_3$ | 0.90 | 1.15 | 0.96 |

Table 2: Comparison of (the variances of) the initial estimator and two one-step estimators. The generic one-step estimator is efficient in each case. The special one-step estimator assumes $W = T$ and is therefore efficient only in case 1. The generic one-step estimator has not reached the (asymptotic) efficiency bound in this simulation ($N = 1000$) but the special one-step estimator has in the first two cases in which $W$ is a perfect or good predictor of $T$.

the following specifics to compute the generic one-step estimator. The censoring distribution was estimated via linear regression of $C$ on $Z$ with independent normal error. The distribution of $T \mid Z, W$ was estimated using equation (18) and a generalized linear model with probit link. From the data model it can be shown that $f_{W|Z}$ can be estimated consistently with linear regression with normal errors in each of the three cases. This estimate can be used to more accurately estimate

$$\phi(Z) = \int f_{W|Z}(w \mid Z) \left[ E(K_{\beta^0_n} \mid Z, W)^2 + \int \frac{F(t \mid Z, w)\overline{F}(t \mid Z, w) K'_{\beta^0_n}(t, Z)}{g(t \mid Z, w)} dt \right] dw$$

(see equation (37)).

The special one-step estimator based on the assumption $W = T$ is easier to compute because the assumption implies $F(t \mid Z, W) = I(t > W)$, $E(K'_{\beta^0_n} \mid Z, W) = K'_{\beta^0_n}(W, Z)$, and

$$\phi(Z) = \int f_{W|Z}(w \mid Z) K_{\beta^0_n}(w, Z)^2 dw.$$

Results in Table 2 are based on 1000 repetitions.

## 5 California Partners' Study

The methods described in this paper were applied to a dataset extracted from the California Partners' Study. Each case consists of a monogamous heterosexual couple in which the male is HIV-positive due to a prior sexual contact. The "failure time variable" on which current status data is available is the time (in months) until infection of the female partner. Several time-independent covariates are available including an indicator of condom use (never=1, ever=0), an indicator of bleeding (ever=1, never=0), an indicator of a sexually transmitted disease (STD) history in the female (ever=1, never=0), an estimate of the rate of sexual contact (contacts per month), and the age of the female (years). There are 87 subjects with complete information on these five covariates. More detailed descriptions of the data are available in Padian, *et al.* (1987), Shiboski and Jewell (1992), Jewell and Shiboski (1990), and Padian, *et al.* (1997).

Our ultimate goal is to estimate the regression parameters in the model $T = Z^{\top}\beta + \epsilon$, where $T$ is the log of the transmission time. Define the following notation: $Z_0 \equiv 1$ is the intercept, $Z_1 = I(\text{No condom use})$, $Z_2 = I(\text{STD History})$, $Z_3 = Z_1 Z_2$. We expect the coefficients of $Z_1$ and $Z_2$ to be negative, indicating these risk factors lower the expected time until transmission of the disease. We include the interaction term because the effect of STD history may not be observed if condoms are used.

Before estimating $\beta$, we must model the censoring mechanism. The distribution of $C$ may be dependent on the covariates in the model and possibly other external to the regression model. Several classes of models for the conditional distribution of C given covariates are feasible including simple linear regression and Cox proportional hazards. In each of these classes the only significant dependence is between the monitoring time and $Z_1$. As noted in the introduction, it may be safer to include more rather than fewer covariates and to specify a semi-parametric rather than parametric model to protect against dependence between $T$ and $C$ as much as possible. With that in mind we chose to use the Cox proportional hazards model and to include all five covariates mentioned in the paragraph describing the dataset.

With a model for the censoring mechanism in hand, we proceed to computing an initial estimate of $\beta$ based on equation (7). The length of the support window of $K'$ can be varied (as can the functional form of $K$) to obtain results for a range of estimators from smoothed median regression to trimmed mean regression. Table 3 displays how the initial and one-step estimates depend on the selection of the window length. For the analysis of log transmission time the

| $\tau$ | Parameter | $\beta_n^0$ | $\beta_n^1$ |
|---|---|---|---|
| 0.17 | $Z_0$ | 4.43 | 4.42 |
| | $Z_1$ | -0.52 | -0.49 |
| | $Z_2$ | -0.27 | -0.26 |
| | $Z_3$ | 0.15 | 0.26 |
| 0.21 | $Z_0$ | 4.43 | 4.43 |
| | $Z_1$ | -0.53 | -0.50 |
| | $Z_2$ | -0.29 | -0.26 |
| | $Z_3$ | 0.17 | 0.30 |
| 0.25 | $Z_0$ | 4.44 | 4.43 |
| | $Z_1$ | -0.54 | -0.50 |
| | $Z_2$ | -0.31 | -0.27 |
| | $Z_3$ | 0.20 | 0.42 |
| 0.29 | $Z_0$ | 4.45 | 4.44 |
| | $Z_1$ | -0.56 | -0.51 |
| | $Z_2$ | -0.33 | -0.28 |
| | $Z_3$ | 0.24 | 0.45 |

Table 3: Dependence of Estimates on Window Length. $K'$ is zero outside $Z^\top\beta \pm \tau$. If $\tau$ is larger than 0.3, this window extends beyond the support of $g_n$. If $\tau$ is smaller then 0.15, the initial estimator has numerous solutions.

estimates do not change substantially with $\tau$. In a similar analysis of the untransformed transmission time, the estimates changed due to the right skewness of the the distribution. For example, the intercept, which represents the time until infection in pairs with neither risk factor, was largest for large $\tau$ and smallest for small $\tau$. A wide window indicates the tail of the distribution will have an effect while a small window indicates only the center of the data is measured.

For $\tau = 0.25$ the initial estimator is $\beta_n^0 = (4.44, -0.54, -0.31, 0.24)$; that is, the conditional log time until infection is centered at $T = 4.44 - 0.54Z_1 - 0.31Z_2 + 0.24Z_3$.

The remaining item is to compute the one-step estimator. The covariates in this data set are time-independent so equation (12) applies. The cumulative distribution function

$F(t|Z, W)$ was estimated using the generalized additive model as in equation (19) with $Z = (Z_0, Z_1, Z_2, Z_3)$ and logit link function. The indicator of bleeding and the age of the female were used as covariates outside the regression model (that is, $W$). Adjusting for these covariates is not possible with any other technique in the literature. The one-step estimator is $\beta_n^1 = (4.43, -0.50, -0.27, 0.42)$; that is, the conditional log time until infection is centered at $T = 4.43 - 0.50Z_1 - 0.27Z_2 + 0.42Z_3$.

The individual standard errors of the coefficients of the two main effect indicator variables are 0.19 and 0.11, respectively. Thus the indicators of no condom use and of STD history are significant (0.01¡p¡0.02) factors in predicting the log time until transmission. The coefficient of the interaction is not statistically significant.

## 6 Discussion

We provided locally efficient estimators or regression coefficients based on current status data with time-dependent covariates with a general linear regression failure-time model, $T = Z^\top\beta + \epsilon$, where the distribution of the error term has conditional location parameter equal to zero. Although the curse of dimensionality prevents a globally efficient estimator, the proposed one-step estimator attains the efficiency bound at a user-supplied submodel of interest and is consistent and asymptotically normal over the whole model.

Another advantage of this locally efficient one-step estimation approach is that the censoring process need not be independent of the failure time; only coarsening at random is required. Unlike other regression estimation approaches, the one-step estimator allows the effects of other unmodeled covariates to be incorporated in a very general way. Thus if a surrogate covariate for $T$ is available, it may be used to improve the estimation of the regression parameters even though the surrogate is not included in the model. Furthermore, the unmodeled covariates may even be time-dependent processes.

The one-step estimator exists in closded form and has been implemented with generally available software. It was shown in simulations to perform according to its asymptotic theoretical properties in finite samples and was applied to data from the California Partners' Study.

# A The efficient influence curve.

## A.1 Orthogonal complement of nuisance tangent space in full data model.

The nuisance tangent space consists of scores of parametric submodels through $F_X$ with $\beta = \beta_0$ fixed. Let $L_0^2(F_X)$ denote the space of (vector-valued) functions $D$ of $X$ with mean zero and finite variance. Consider parametric submodels of the form

$$dF_X^{(D,s)}(X) = (1 + s^\top D(X))dF_X(X),$$

where the vector $s$ is the parameter and $D \in L_0^2(F_X)$:

To ensure $dF_X^{(D,s)}$ is a model of the form (1), it is necessary that the conditional mean given $Z$ of $K_{\beta_0}(T, Z)$ be zero. Because $E(K_{\beta_0} \mid Z) = 0$ under the true distribution, it follows that

$$E^{(D,s)}(K_{\beta_0} \mid Z) = 0 \iff E(DK_{\beta_0} \mid Z) = 0.$$

Thus the nuisance tangent space of the full data model is

$$\Lambda^F = \{D \in L_0^2(F_X) : E(DK_{\beta_0} \mid Z) = 0\}. \tag{25}$$

Define the score of $\beta$ by $S_\beta(X) \equiv (\partial/\partial\beta) \log dF_X(X)$. In our case,

$$S_\beta(X) = \frac{\partial}{\partial\beta} \log f(t \mid Z) = \frac{\partial}{\partial\beta} \log f_{\epsilon|Z}(t - Z^\top\beta \mid Z) = -Z\frac{f'(t \mid Z)}{f(t \mid Z)}.$$

The efficient score $S_{eff}^F$ of $\beta$ is the projection of $S_\beta$ onto the orthogonal complement of the nuisance tangent space, $\Pi[S_\beta \mid \Lambda^{F,\perp}]$. The full data efficient influence curve or canonical gradient is then given by $c^{-1}S_{eff}^F$, where $c$ is the normalizing constant $E(S_{eff}^{F,\top} S_{eff}^F)$ The variance of a regular asymptotically linear estimator of $\beta$ is bounded below by the variance of the efficient influence curve, which equals the inverse of the variance matrix of the efficient score.

**Lemma 2** *The orthogonal complement in $L_0^2(F_X)$ of the nuisance tangent space, $\Lambda^F$, is*

$$\Lambda^{F,\perp} = \{D \in L_0^2(F_X) : D(X) = h(Z)K_\beta(T, Z) \text{ for some } h(Z) \in L^2(F_Z)\}. \tag{26}$$

*The projection of $D \in L_0^2(F_X)$ onto $\Lambda^{F,\perp}$ is*

$$\Pi[D \mid \Lambda^{F,\perp}] = \frac{E(DK_\beta \mid Z)}{E(K_\beta^2 \mid Z)} K_\beta. \tag{27}$$

*If $f(\cdot \mid Z)$ is absolutely continuous on $\mathbb{R}$, then the efficient score for $\beta$ is*

$$S^F_{\text{eff}}(T, Z, \overline{L}) = \frac{-ZK_\beta}{E(K^2_\beta \mid Z)} \int_{\text{supp } f} f'(t \mid Z) K_\beta(t, Z) dt. \tag{28}$$

*If $K$ also is absolutely continuous on $\mathbb{R}$, then the efficient score for $\beta$ is*

$$S^F_{\text{eff}}(T, Z, \overline{L}) = \frac{ZE(K'_\beta \mid Z)}{E(K^2_\beta \mid Z)} K_\beta. \tag{29}$$

**Proof:** The operator $\Pi[\cdot \mid \Lambda^{F,\perp}]$ given in equation (27) is such that that $D - \Pi[D \mid \Lambda^{F,\perp}] \in \Lambda^F$ and $D - (D - \Pi[D \mid \Lambda^{F,\perp}]) \perp \Lambda^F$. This proves that $D - \Pi[D \mid \Lambda^{F,\perp})$ is the projection of $D$ on $\Lambda^F$ and thus that $\Pi[D \mid \Lambda^{F,\perp}]$ equals the projection of $D$ onto $\Lambda^{F,\perp}$. The form of $\Lambda^{F,\perp}$ is now apparent from the projection operator.

The efficient score for $\beta$ is the projection of $S_\beta$ onto $\Lambda^{F,\perp}$:

$$\Pi[S_\beta \mid \Lambda^{F,\perp}] = \frac{-ZK_\beta}{E(K^2_\beta \mid Z)} \int_{\text{supp } f} f'(t \mid Z) K_\beta(t, Z) dt = \frac{ZK_\beta}{E(K^2_\beta \mid Z)} E(K'_\beta \mid Z).$$

The second equality is obtained via integration by parts if $K$ is absolutely continuous. □

The substantial advantage of equation (29) over equation (28) is that the derivative of $f$ no longer needs to be estimated. The stated continuity condition on $f$ can be relaxed. Suppose $T \mid Z$ is exponential ($f$ has discontinuity at 0). Then (integration by parts)

$$\int_0^\infty f'(t \mid Z) K_\beta(t, Z) dt = 0 - f(0 \mid Z) K_\beta(0, Z) - E(K'_\beta \mid Z).$$

All the following derivations hold if the above expression replaces $-E(K'_\beta \mid Z)$ in equation (29). However for simplicity we continue for $f$ absolutely continuous on $\mathbb{R}$.

## A.2 The orthogonal complement of the nuisance tangent space of $\beta$ in Observed Data Model

The orthogonal complement $\Lambda^\perp$ of the nuisance tangent space of $\beta$ in observed data model (only assuming CAR on $G$) can be represented as the range of a mapping, $IC$, defined on the orthogonal complement $\Lambda^{F,\perp}$ of the nuisance tangent space of $\beta$ in the full data model. Let $P_{F_X,G}$ be the distribution of the observed data, $Y$, and $T_{\text{CAR}}(G) \subset L^2_0(P_{F_X,G})$ be the space of all functions of $Y$ with conditional mean zero given $X$. The latter space is the nuisance tangent space for $G$ only assuming CAR. By van der Laan and Robins (1997) and Robins and Rotnitzky (1992) we have at $P_{F_X,G}$ that $IC(D) \equiv U_G(D) - \Pi[U_G(D) \mid T_{\text{CAR}}]$ is an element of $\Lambda^\perp$ for each $D(T, Z) \in \Lambda^{F,\perp}$ where

$$U_G(D)(C, \Delta, Z, \overline{L}(C)) = \frac{D'(C, Z)\overline{\Delta}}{g(C \mid X)} + D(\alpha_X, Z)$$

and

$$\begin{aligned}
&\Pi[U_G(D) \mid T_{\text{CAR}}](C, \Delta, Z, \overline{L}(C)) \\
&= \int \left( \frac{D'(u, Z)\overline{F}(u \mid Z, \overline{L}(u))}{g(u \mid X)} - \frac{1}{\overline{G}(u \mid X)} \int_u D'(t, Z)\overline{F}(t \mid Z, \overline{L}(u))dt \right) dM(u)
\end{aligned}$$

where $dM(u) = I(C \in du) - \Lambda_C(du \mid X)I(C > u)$ and $\Lambda_C$ is the cumulative hazard of censoring. In Robins and Rotnitzky (1992) and van der Laan, Gill, Robins (2000) it is also shown there exists a $D^*_{\text{opt}}$ which is mapped to the efficient influence curve (canonical gradient), $\ell^*_{\text{eff}}$, of the observed data model. The estimating equation in the full data model, $D_{\text{opt}} \in \Lambda^{F,\perp}$, which corresponds with the efficient score $IC(D_{\text{opt}})$ of $\beta$ in the observed data model is the solution of the following mathematics problem:

$$\begin{aligned}
D_{\text{opt}}(T, Z) &\in \Lambda^{F,\perp} \\
\Pi[m^{-1}(D_{\text{opt}}) \mid \Lambda^{F,\perp}] &= S^F_{\text{eff}},
\end{aligned} \tag{30}$$

where $m : L^2_0(F_X) \to L^2_0(F_X)$ is the nonparametric information operator $m = g^\top g = E(E(\cdot \mid Y) \mid X)$. The score operator $g : L^2_0(F_X) \to L^2_0(F_Y)$ is defined by $E(\cdot \mid Y)$ and its adjoint, $g^\top : L^2_0(F_Y) \to L^2_0(F_X)$ is $E(\cdot \mid X)$.

In section A.4 we first solve for $D_{\text{opt}}$. Subsequently, by standardizing $D_{\text{opt}}$ with a matrix $c_{\text{opt}}$ we obtain the optimal $D^*_{\text{opt}} = c^{-1}_{\text{opt}} D_{\text{opt}}$ which yields the efficient influence curve $\ell^*_{\text{eff}} = IC(D^*_{\text{opt}})$ in the observed data model.

## A.3 Useful Properties of $m$

Recall $X = (T, Z, \bar{L})$.

**Lemma 3** *Suppose $D_A$ and $D_B$ are any two functions in the range of $m : L^2_0(F_X) \to L^2_0(F_X)$. Then*

$$E(m^{-1}(D_A)D_B \mid Z) = E(gm^{-1}(D_A)gm^{-1}(D_B) \mid Z).$$

**Proof:** The basic fact to be shown is that

$$\left\langle m^{-1}(D_A), D_B \right\rangle = \left\langle gm^{-1}(D_A), gm^{-1}(D_B) \right\rangle$$

where the inner product is expectation conditional on $Z$. Properties of adjoints gives the result immediately by moving the first $g$ to the other side of the inner product and noting that $g^\top gm^{-1}$ is the identity. □

**Lemma 4** *Let $h_a, h_b \in L^2(F_Z)$ (thus only depends on $X$ through $Z$) and $D \in L_0^2(F_X)$. Then*

$$m(h_a D + h_b) = h_a m(D) + h_b$$

*and, if $D \in range(m) \subset L_0^2(F_X)$,*

$$m^{-1}(h_a D + h_b) = h_a m^{-1}(D) + h_b.$$

**Proof:** The result for $m$ follows directly from its definition and properties of conditional expectations. Then the property for $m^{-1}$ follows. □

**Lemma 5** *Suppose $D_a, D_b \in L_0^2(F_X)$ and $D_a = m(D_b)$. Then $E(D_b \mid W) = E(D_a \mid W)$.*

**Proof:**

$$E(D_a \mid Z) = E(m(D_b) \mid Z) = E(E(E(D_b \mid C, \Delta, Z, \overline{L}(C)) \mid T, Z, \overline{L}) \mid Z) = E(D_b \mid Z)$$

The final equality is the result of the law of iterated expectation. □

The information operator, $m$, can be written more explicitly for current status data when all covariates are uncensored. This is the case if, for example, the covariates are time independent.

**Proposition 1** *Suppose censored data $Y = (C, \Delta, W)$ is observed on the full data $X = (T, W)$ and $D \in L_0^2(F_X)$. Then*

$$\begin{aligned} & m(D)(X) \\ = \;& E\left( \frac{\int^C D(t,W)dF(t \mid W)\Delta}{F(C \mid W)} + \frac{\int_C D(t,W)dF(t \mid W)\overline{\Delta}}{\overline{F}(C \mid W)} \middle| T, W \right) \\ = \;& \int_T \frac{\int^c D(t,W)dF(t \mid W)}{F(c \mid W)} dG(c \mid X) + \int^T \frac{\int_c D(t,W)dF(t \mid W)}{\overline{F}(c \mid W)} dG(c \mid X). \end{aligned} \tag{31}$$

**Proof:** The expressions follow directly from the definition of $m$. □

The information operator, $m$, can be inverted in closed form for current status data when all covariates are time-independent.

**Proposition 2** *Suppose censored data $Y = (C, \Delta, W)$ is observed on the full data $X = (T, W)$. Let $D \in L_0^2(F_X)$ be twice differentiable with respect to $T$. Then $D \in range(m)$ and*

$$m^{-1}(D)(X) = E(D(T,W) \mid W) - \frac{1}{f(T \mid W)} \frac{\partial}{\partial t} \left( \frac{F(t \mid W)\overline{F}(t \mid W)D'(t,W)}{g(t \mid X)} \right)_{t=T}. \tag{32}$$

**Proof:** Define $D_a(T,W) \equiv D(T,W) - E(D(T,W) \mid W)$ which has conditional mean zero given $W$. The operator $m^{-1}$ is linear (Lemma 4) so

$$m^{-1}(D)(T,W) \quad = \quad m^{-1}(D_a)(T,W) + E(D(T,W) \mid W). \tag{33}$$

Define $D_b(T,W) \equiv m^{-1}(D_a)(T,W)$. Differentiate the expression for $D_a = m(D_b)$ given by equation (31) with respect to $T$. By Lemma 5, $E(D_b(T,W) \mid W) = 0$, which implies $\int_T D_b(t,W)dF(t \mid W) = -\int^T D_b(t,W)dF(t \mid W)$. Thus we can solve for $\int^T D_b(t,W)dF(t \mid W)$. Differentiate again with respect to $T$ and divide by $f(T \mid W)$ to show

$$D_b(T,W) \quad = \quad \frac{-1}{f(T \mid W)} \frac{\partial}{\partial t} \left( \frac{F(t \mid W)\overline{F}(t \mid W) D_a'(t,W)}{g(t \mid X)} \right)_{t=T}.$$

Note $D_a' = D'$ and then substitute $D_b$ into equation (33). □

### A.4 Optimal Estimating Equation, $D_{\text{opt}}$

We wish to solve equation (30) for $D_{\text{opt}}$ for our observed data regression model. Although we succeeded in solving for $m^{-1}$ explicitly even for general $\overline{L}$, we prefer a simpler alternative based on $gm^{-1}$ when $\overline{L}$ involves censored, time dependent covariates (equation (35)); the explicit expression for $m^{-1}$ happens to be very lengthy. If $\overline{L}$ contains only time independent covariates then more explicit formulae for $D_{\text{opt}}$ are attractive (equations (36) and (37)).

**Theorem 2** *Suppose $K$ is twice differentiable and assume (4. Then*

$$D_{\text{opt}}(X) \equiv h_{\text{opt}}(Z) K_\beta(T,Z) \equiv \frac{Z E(K_\beta' \mid Z)}{\phi(Z)} K_\beta(T,Z) \tag{34}$$

*where, for general $\overline{L}$,*

$$\phi(Z) = E(IC(Y \mid F, G, K_\beta)^2 \mid Z). \tag{35}$$

*Consider the case where $(Z, \overline{L}) = Z$ (i.e., no covariates other than those modeled). Then*

$$\phi(Z) = \int \frac{F(t \mid Z)\overline{F}(t \mid Z) {K_\beta'}^2(t,Z)}{g(t \mid X)} dt. \tag{36}$$

*Consider the case where $(Z, \overline{L}) = (Z, W)$ with $W$ time independent but not modeled. Then*

$$\phi(Z) = E\left( E(K_\beta \mid Z, W)^2 + \int \frac{F(t \mid Z,W)\overline{F}(t \mid Z,W){K_\beta'}^2(t,Z)}{g(t \mid X)} dt \,\middle|\, Z \right). \tag{37}$$

**Proof:** From the representation of $\Lambda^{F,\perp}$ given in equation (26), we have $D_{\text{opt}}(T,Z) = h_{\text{opt}}(Z)K_\beta(T,Z)$ for some $h_{\text{opt}}(Z)$. Under condition (4 all elements in this proof will actually be elements of the Hilbert space $L_0^2(P_{F_X,G})$ so that projections are well defined. Lemma 4 implies $m^{-1}(D_{\text{opt}})(X) = h_{\text{opt}}(Z)m^{-1}(K_\beta)(X)$. From equation (27), the projection onto $\Lambda^{F,\perp}$ is

$$\Pi[m^{-1}(D_{\text{opt}}) \mid \Lambda^{F,\perp}] = h_{\text{opt}}(Z)\frac{E(m^{-1}(K_\beta)(X)K_\beta \mid Z)}{E(K_\beta^2 \mid Z)}K_\beta. \tag{38}$$

Rewrite equation (30) substituting equations (29) and (38):

$$h_{\text{opt}}(Z)\frac{E(m^{-1}(K_\beta)(X)K_\beta \mid Z)}{E(K_\beta^2 \mid Z)}K_\beta = \frac{ZE(K'_\beta \mid Z)}{E(K_\beta^2 \mid Z)}K_\beta.$$

Solving for $h_{\text{opt}}$ gives

$$h_{\text{opt}}(Z) \quad = \quad \frac{ZE(K'_\beta \mid Z)}{E(K_\beta m^{-1}(K_\beta)(X) \mid Z)}. \tag{39}$$

Thus $D_{\text{opt}}$ has the structure of equation (34) where $\phi(Z)$ is the denominator of equation (39). Apply Lemma 3 with $D_a = D_b = K_\beta$ and use that for any $D \in L^2(F_X)$, $gm^{-1}(D) = IC(D)$ (Gill, *et al.*, 1998, Robins and Rotnitzky, 1992) to get equation (35).

Now suppose $(Z, \overline{L}) = Z$. Expand the denominator of equation (39) using Proposition 2. The first term, $E(K_\beta \mid Z)$, is zero for our model.

$$\begin{aligned} &E(K_\beta(X)m^{-1}(K_\beta)(X) \mid Z) \\ &\quad = \quad -E\left(\frac{K_\beta(T,Z)}{f(T \mid Z)}\frac{\partial}{\partial t}\left(\frac{F(t \mid Z)\overline{F}(t \mid Z)K'_\beta(t,Z)}{g(t \mid X)}\right)_{t=T}\middle| Z\right) \\ &\quad = \quad -\int_{\text{supp}f} K_\beta(t,Z)\frac{\partial}{\partial t}\left(\frac{F(t \mid Z)\overline{F}(t \mid Z)K'_\beta(t,Z)}{g(t \mid X)}\right)dt. \end{aligned} \tag{40}$$

Integration by parts gives

$$E(K_\beta(X)m^{-1}(K_\beta)(X) \mid Z) = \int \frac{F(t \mid Z)\overline{F}(t \mid Z){K'_\beta}^2(t,Z)}{g(t \mid X)}dt.$$

Now suppose $\overline{L} = W$ is time independent. Because $m$ and $m^{-1}$ are determined by the data structure and not the regression model, Lemma 4 and 5 and Propositions 1 and 2 still hold when $Z$ is renamed $(Z, W)$. By the law of iterated expectation

$$E(K_\beta(X)m^{-1}(K_\beta)(X) \mid Z) = E(E(K_\beta(X)m^{-1}(K_\beta)(X) \mid Z, W) \mid Z).$$

The interior expectation is similar to equation (40). We proceed slightly differently only because $E(K_\beta \mid Z, W)$ may not be zero.

$$
\begin{aligned}
& E(K_\beta(X) m^{-1}(K_\beta)(X) \mid Z, W) \\
&= E(K_\beta(X) E(K_\beta(X) \mid Z, W) \mid Z, W) \\
&\quad - E\left( \frac{K_\beta(T,Z)}{f(T \mid Z, W)} \frac{\partial}{\partial t} \left( \frac{F(t \mid Z, W) \overline{F}(t \mid Z, W) K'_\beta(t, Z)}{g(t \mid X)} \right)_{t=T} \Bigg| Z, W \right) \\
&= E(K_\beta(X) \mid Z, W)^2 + \int \frac{F(t \mid Z, W) \overline{F}(t \mid Z, W) {K'_\beta}^2(t, Z)}{g(t \mid X)} dt.
\end{aligned}
$$

Similarly to the previous case, the final equality follows via integration by parts. Take expectations conditional on $Z$ to obtain equation (37). □

#### A.4.1 Normalization of $D_{\text{opt}}$

Above we solved for $D_{\text{opt}}(T, Z) = h_{\text{opt}}(Z) K_\beta(T, Z)$. We have that $IC(D_{\text{opt}})$ equals the efficient influence curve up till a normalizing matrix $c(h_{\text{opt}})$. Let $D^*_{\text{opt}} = c(h_{\text{opt}})^{-1} h_{\text{opt}}(Z) K_\beta(T, Z)$ be such that $IC(D^*_{\text{opt}})$ equals the efficient influence curve. Then because $IC(D^*_{\text{opt}})$ is an influence curve we have $-E((\partial/\partial\beta) IC(c(h_{\text{opt}})^{-1} D_{opt})$ is the $k \times k$ identity matrix. Thus $c(h_{\text{opt}}) = -E(\frac{\partial}{\partial\beta} IC(D_{\text{opt}}))$. In general, we have $c(h) = -E((\partial/\partial\beta) IC(D))$, where $D = h(Z) K_\beta(T, Z)$. The result of Lemma 6 is that $c(h) = E(h(Z) Z^\top U_G(K'_\beta))$. In particular,

$$
c_{\text{opt}} = c(h_{\text{opt}}) = E\left( Z Z^\top \frac{E(K'_\beta \mid Z) U_G(K'_\beta)}{\phi(Z)} \right). \tag{41}
$$

**Lemma 6** *The following equality holds for* $D(X) = h(Z) K_\beta(T, Z)$*:*

$$
c(h) \equiv -E\left( \frac{\partial}{\partial\beta} IC(D) \right) = E(h(Z) Z^\top U_G(K'_\beta)). \tag{42}
$$

**Proof:** From equations (10) and (11) it can be seen that $IC(h(Z) K_\beta) = h(Z) IC(K_\beta)$. Furthermore, $E(IC(K_\beta)(Y) \mid Z) = E(U_G(K'_\beta)(Y) \mid Z) = E(K_\beta(\epsilon) \mid Z) = 0$. Thus

$$
\begin{aligned}
c(h) &\equiv -E\left( \frac{\partial}{\partial\beta} IC(D) \right) \\
&= -E\left( \frac{\partial}{\partial\beta} \left[ h(Z) IC(K_\beta) \right] \right) \\
&= -E\left( \left( \frac{\partial}{\partial\beta} h(Z) \right) IC(K_\beta) - h(Z) \left( \frac{\partial}{\partial\beta} IC(K_\beta) \right) \right) \\
&= 0 - E\left( h(Z) \left( IC(\frac{\partial}{\partial\beta} K_\beta) \right) \right)
\end{aligned}
$$

$$
\begin{aligned}
&= E\left(h(Z)Z^\top IC(K'_\beta)\right) \\
&= E\left(h(Z)Z^\top U_G(K'_\beta)\right).
\end{aligned}
$$

□

# B Consistency of the initial estimator and verification of the structural condition for theorem 1

In this section we give conditions under which the initial estimator, $\beta_n^0$, exists and is $\sqrt{n}$-consistent (Lemma 7 and 8). Further conditions (Lemma 9) insure the structural condition $E(\ell^*_{\text{eff}}(Y \mid F_n, G, h_n, c_n, \beta_n^0)) = \beta - \beta_n^0 + o_P(n^{-1/2})$. These results complete the proof of Theorem 1.

In this section we use the following estimating equation notation: for a random quantity $X$, $PX = E(X)$, $P_nX = n^{-1}\sum_{i=1}^n X_i$, and $H(\cdot)$ is a vector-valued estimating equation.

**Lemma 7** *Define $H(P, g, \beta) \equiv E(ZU_G(K_\beta)(Y))$. Suppose $(\partial/\partial\beta)H(P, g, \beta)$ is invertible at the true value of $\beta = \beta_0$ and that $H(P_n, g_n, \beta)$ is continuous in $\beta$ in a neighborhood $N_{\beta_0}$ of $\beta_0$. (If $K'$ is continuous and $g_n \neq 0$ at the observed monitoring times, this continuity condition holds.) Assume condition (4 and that $ZU_{G_n}(K_\beta)(Y)$ is in a $P_{F_X,G}$-Donsker class with probability tending to one. Assume $\| g_n - g \|_{\infty,A} \to 0$ in probability, where $A = \{(c, Z, \bar{L}(c)) : K'(c - Z^\top\beta) > 0, \bar{F}(c \mid Z, \bar{L}(c)) > 0\}$. Asume the mean and covariance matrix of the covariate $Z$ exist. Then there exists a solution $\beta = \beta_n^0$ in $N_{\beta_0}$ of $H(P_n, g_n, \beta_n^0) = 0$ (equation (7)) with probability tending to one.*

**Proof:** First note $H(P, g, \beta_0) = 0$. The invertibility condition implies existence of $\beta_1$ and $\beta_2$ in $N_{\beta_0}$ such that for each component, $H(P, g, \beta_1) < 0$ and $H(P, g, \beta_2) > 0$. Under the regularity condition, sample values converge to true values with probability one. Therefore $H(P_n, g_n, \beta_1) < 0$ and $H(P_n, g_n, \beta_2) > 0$ with probability tending to one. The continuity condition implies existence of a solution in $N_{\beta_0}$ of $H(P_n, g_n, \beta_n^0) = 0$ (equation (7)) with probability tending to one. □

**Lemma 8** *Assume the conditions of Lemma 7. Suppose that $\| g_n - g \|_{\infty,A} = O_P(n^{-1/4})$, and that $\| G_n - G \|_{\infty,A} = O_P(n^{-1/2})$. Finally assume $g'$ and $K''$ are bounded above and condition (4) holds for $\beta \in N(\beta_0)$. where $N_{\beta_0}$ is an arbitrarily small neighborhood of $\beta_0$. . Then $\beta_n^0 - \beta = O_P(n^{-1/2})$.*

**Proof:** Recall from the previous lemma, $H(P,g,\beta) \equiv E(ZU_G(K_\beta)(Y))$ and $H(P,g,\beta_0) = H(P_n,g_n,\beta_n^0) = 0$. Thus

$$\begin{aligned} & H(P,g,\beta_n^0) - H(P,g,\beta_0) \\ = \;& H(P,g,\beta_n^0) - H(P,g_n,\beta_n^0) + H(P,g_n,\beta_n^0) - H(P_n,g_n,\beta_n^0). \end{aligned} \tag{43}$$

The difference of the final two terms is an empirical process indexed by a random function which falls with probability tending to 1 in a $P_{F_X,G}$-Donsker class by the regularity condition of Lemma 7. Thus this difference is $O_P(n^{-1/2})$.

The difference of the first two terms on the right hand side is

$$\begin{aligned} & H(P,g,\beta_n^0) - H(P,g_n,\beta_n^0) \\ = \;& E\left(ZK'_{\beta_n^0}(C,Z)\overline{\Delta}\left(\frac{1}{g}-\frac{1}{g_n}\right)\right) \\ = \;& E\left(ZK'_{\beta_n^0}(C,Z)\overline{\Delta}\left(\frac{g_n-g}{gg_n}-\frac{g_n-g}{g^2}+\frac{g_n-g}{g^2}\right)\right) \\ = \;& -E\left(ZK'_{\beta_n^0}(C,Z)\overline{\Delta}\frac{(g_n-g)^2}{g^2g_n}\right) + E\left(E\left(ZK'_{\beta_n^0}(C,Z)\overline{\Delta}\frac{g_n-g}{g^2} \mid X\right)\right). \end{aligned}$$

The first term is $O_P(n^{-1/2})$ because $g$ and $g_n$ are bounded away from zero by condition (4, $K'$ is bounded above, $Z$ has finite mean, and $||g_n-g||_{\infty,A} = O_P(n^{-1/4})$.

Continuing with inside expectation of the second term,

$$\begin{aligned} & E\left(K'_{\beta_n^0}(C,Z)\overline{\Delta}\frac{g_n-g}{g^2}\middle| X\right) \\ = \;& \int_{\alpha_X}^T K'_{\beta_n^0}(c,Z)\frac{g_n-g}{g}dc \\ = \;& \int_{\alpha_X}^T \frac{K'_{\beta_n^0}(c,Z)}{g}d(G_n-G) \\ = \;& (G_n-G)\frac{K'_{\beta_n^0}(c,Z)}{g}\Bigg|_{\alpha_X}^T - \int_{\alpha_X}^T (G_n-G)\frac{d}{dc}\frac{K'_{\beta_n^0}(c,Z)}{g}dc \end{aligned}$$

This is bounded in absolute value by

$$\begin{aligned} \| G_n - G \|_{\infty,A} \Bigg( & \frac{K'_{\beta_n^0}(T,Z)}{g(T|X)} + \frac{K'_{\beta_n^0}(\alpha_X,Z)}{g(\alpha_X|X)} \\ & + \left|\int_{\alpha_X}^T \frac{K''_{\beta_n^0}(c,Z)}{g(c|X)} - \frac{K'_{\beta_n^0}(c,Z)g'(c|X)}{g^2(c|X)}dc\right|\Bigg) \end{aligned}$$

The bounds on $g$, $g'$, and $K''$ imply this term is the same order as $\| G_n - G \|_{\infty,A}$, which is $O_P(n^{-1/2})$.

The left side of equation (43) is $[(\partial/\partial\beta)H(P,g,\beta)](\beta_n^0-\beta)+o_P(\beta_n^0-\beta)$. So

$$\beta_n^0-\beta=\left[\frac{\partial}{\partial\beta}H(P,g,\beta)\right]^{-1}\left(O_P(n^{-1/2})+o_P(\beta_n^0-\beta)\right).$$

This can only hold if $\beta_n^0-\beta=O_P(n^{-1/2})$. □

**Lemma 9** *Assume the conditions of Lemma 8. Suppose $h_n(Z)$ converges uniformly to some $h(Z)$. Define*

$$c_n=-\frac{1}{n}\sum_{i=1}^{n}h_n(Z_i)Z_i^\top U_{G_n}(K'_{\beta_n^0})(Y_i). \tag{44}$$

*Then $c_n$ converges to $c=-E(h(Z)Z^\top U_G(K'_\beta)(Y))$. If $c$ is invertible then*

$$E\left(c_n^{-1}h_n(Z)IC(Y\mid\beta_n^0,F_n,G,K_{\beta_n^0})\right)=\beta-\beta_n^0+o_P(n^{-1/2}). \tag{45}$$

**Proof:** Write $c_n=H(\beta_n^0,g_n,h_n,P_n)$ and $c=H(\beta,g,h,P)$. Then

$$\begin{aligned}c_n-c &= H(\beta_n^0,g_n,h_n,P_n)-H(\beta_n^0,g_n,h_n,P)\\ &\quad +H(\beta_n^0,g_n,h_n,P)-H(\beta_n^0,g_n,h,P)\\ &\quad +H(\beta_n^0,g_n,h,P)-H(\beta_n^0,g,h,P)\\ &\quad +H(\beta_n^0,g,h,P)-H(\beta,g,h,P).\end{aligned}$$

The first difference goes to zero by the regularity condition. The second difference goes to zero by the uniform convergence of $h_n$ to $h$. The third difference goes to zero because $g_n$ converges to $g$ and both are bounded from zero on $\Gamma_Z$. The final difference goes to zero because $\beta_n^0$ converges to $\beta$ (Lemma 8) and $K'$ is continuous.

If $c_n$ is not invertible, define $c_n^{-1}=I$. However, if $c$ is invertible, $c_n^{-1}$ will exist with probability tending to one.

We can simplify the expression on the left side of equation (45) because the projection term, $IC_{nu}$, has conditional expectation zero given $Z$ and the first term, $IC_0$, has conditional expectation $K_{\beta_n^0}$ given $X$:

$$\begin{aligned}&E\left(c_n^{-1}h_n(Z)IC(Y\mid\beta_n^0,F_n,G,K_{\beta_n^0})\right)\\ &= E\left(c_n^{-1}h_n(Z)(E(IC_0(Y\mid G,K_{\beta_n^0})\mid X)-E(IC_{nu}(Y\mid\beta_n^0,F_n,G,K_{\beta_n^0})\mid Z))\right)\\ &= E(c_n^{-1}h_n(Z)K_{\beta_n^0}(X)).\end{aligned} \tag{46}$$

The Taylor expansion of $K_{\beta_n^0}(X)$ about $\beta$ is

$$K_{\beta_n^0}(X)=K_\beta(X)-K'_\beta(X)Z^\top(\beta_n^0-\beta)+o_P(K'_\beta(X)Z^\top(\beta_n^0-\beta)).$$

Because $E(K_\beta \mid Z) = 0$, expression (46) equals

$$-E(c_n^{-1}h_n(Z)K'_\beta(X)Z^\top)(\beta_n^0 - \beta) + o_P\left(E\left(c_n^{-1}h_n(Z)K'_\beta(X)Z^\top\right)(\beta_n^0 - \beta)\right).$$

Invertibility of $c$ and uniform convergence of $h_n$ implies $c_n^{-1}h_n$ converges uniformly to $c^{-1}h$ and the above quantity equals

$$-E(c^{-1}h(Z)K'_\beta(X)Z^\top)(\beta_n^0 - \beta) + o_P(E(c^{-1}h(Z)K'_\beta(X)Z^\top)(\beta_n^0 - \beta)).$$

By the definition of $c$, this is $\beta - \beta_n^0 + o_P(\beta - \beta_n^0)$. □

In particular, if $h(Z) = h_{\text{opt}}(Z) = ZE(K'_\beta \mid Z)/\phi(Z)$, then $c = c_{\text{opt}} = E(ZZ^\top E(K'_\beta \mid Z)^2/\phi(Z))$ is nonsingular. Write $c = \int zz^\top f^*(z)dz$ where $f^*$ is the density of $Z$ times $E(K'_\beta \mid Z)^2/\phi(Z))$. Because the latter factor is positive, the function $f^*$ is proportional to a density. Thus $c = E^*(ZZ^\top)$ and is invertible if the components of $Z$ are not linearly dependent. This argument also proves that if one estimates $c_{\text{opt}}$ by substitution of estimators of $E(K_\beta \mid Z)$ and $\phi$, then the resulting estimator $c_n$ converges to a $c$ which is invertible.

## References


Bickel, P.J., Klaassen, A.J., Ritov, Y. and Wellner, J.A. (1993), *Efficient and adaptive inference in semi-parametric models*, Johns Hopkins University Press, Baltimore.

Diamond, I.D. McDonald, J.W. and Shah, I.H. (1986), Proportional hazards models for current status data: application to the study of differentials in age at weaning in Pakistan, *Demography* **23**, 607–620.

Diamond, I.D. and McDonald, J.W. (1991), The analysis of current status data, *Demographic Applications of Event History Analysis*, J. Trussell, R. Hankinson, and J. Tilton (eds.), Oxford: Oxford University Press.

van de Geer, S. (1994), *Asymptotic normality in mixture models,* preprint University of Leiden, the Netherlands.

Gill, R.D., van der Laan, M.J. and Robins, J.M. (1997), Coarsening at Random: Characterizations, Conjectures and Counter-examples, *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, 255-295, Springer, D.Y. Lin and T.R. Fleming (eds).

Groeneboom, P. and Wellner, J.A. (1992), *Information bounds and nonparametric maximum likelihood estimation*, Birkhäuser Verlag.

Heitjan, D.F. and Rubin, D.B. (1991), Ignorability and coarse data, *Annals of Statististics* **19**, 2244–2253.

Huang, J. (1996), Efficient estimation for the proportional hazards model with interval censoring, *Annals of Statististics*, **24**, 540–568.

Huang, J. and Wellner, J.A. (1995), Asymptotic normality of the NPMLE of linear functionals for interval censored data, case I, *Statistica Neerlandica* **49**, 153–163.

Jacobsen, M. and Keiding, N. (1995) , Coarsening at random in general sample spaces and random censoring in continuous time, *Annals of Statistics*, **23**, 774–786.

Jewell, N.P. Malani, H.M. and Vittinghoff, E. (1994), Nonparametric estimation for a form of doubly censored data with application to two problems in AIDS, *Journal of the American Statistical Association*, **89**, 7–18.

Jewell, N.P. and Shiboski, S.C. (1990), Statistical analysis of HIV infectivity based on partner studies, *Biometrics*, **46**, 1133-1150.

Keiding, N. (1991) Age-specific incidence and prevalence (with discussion), *Journal of the Royal Statistical Society Ser. A*, **154**, 371–412.

van der Laan, M.J. and Robins, J.M. (1997), Locally Efficient Estimation with Current Status Data and Time-Dependent Covariates. *Journal of the American Statistical Association*, **93**, No. 442, 693–701.

van der Laan, M.J., Gill, R.D. and Robins, J.M. (2000), *Locally efficient estimation with censored data and high dimensional covariate processes*, Technical report # 85, Division of Biostatistics, University of California, Berkeley.

Morrison D. (1990), *Multivariate Statistical Methods, 3rd ed.*, McGraw-Hill Publishing Co., San Francisco.

Padian, N., Marquis, L., Francis, D.P., Anderson, R.E., Rutherford, G.W., O'Malley , P.M. and Winkelstein, W. (1987), Male-to-Female transmission of Human Immunodeficiency Virus, *Journal of the American Medical Association*, **258**, 788–790.

Padian, N., Shiboski, S.C., Glass, S.O. and Vittinghoff, E. (1997), Heterosexual transmission of HIV in northern California: results from a ten year study, *American Journal of Epidemiology*, **146**, 350–357.

Rabinowitz, D., Tsiatis, A. and Aragon, A. (1995), Regression with interval-censored data, *Biometrika*, **82**, 501–513.

Robins, J.M. (1993), Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers, *Proceedings of the Biopharmaceutical Section, American Statistical Association*, 22–33.

Robins, J.M. and Rotnitzky, A. (1992), Recovery of information and adjustment for dependent censoring using surrogate markers, *Aids Epidemiology, Methodological issues*, Birkhäuser.

Robins, J. M. and Ritov, Y. (1997), Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models, *Statistics in Medicine*, **16**, 285–319.

Shiboski, S.C. and Jewell, N.P. (1992), Statistical Analysis of the time dependence of HIV infectivity based on partner study data, *Journal of the American Statistical Association*, **87**, 360–372.

Shen, X. (2000), Linear Regression With Current Status Data, *Journal of the American Statistical Association*, **95**, number 451, 842–852.

Sun, J. and Kalbfleisch, D. (1993), The analysis of current status data on point processes, *Journal of the American Statistical Association*, **88**, 1449–1454.

van der Vaart, A.W. and Wellner, J.A. (1996), *Weak convergence and empirical processes*, Springer Verlag.