

Locally Efficient Estimation With Current Status Data and Time-Dependent Covariates

Mark J. VAN DER LAAN and James M. ROBINS

In biostatistical applications, interest often focuses on the estimation of the distribution of a failure time variable T . If one observes only 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 extend the data structure by allowing the presence of a possibly time-dependent covariate process which is observed up till the monitoring time C . We follow the approach of Robins and Rotnitzky by modeling the hazard of C conditional on the failure time variable and the covariate-process (i.e., the missingness or censoring process) under the restriction that the missingness (monitoring) process satisfies coarsening at random. Because of the curse of dimensionality, no globally efficient nonparametric estimators with a good practical performance at moderate sample sizes exist. We introduce an inverse probability of censoring weighted estimator of the distribution and of smooth functionals of this distribution that are guaranteed to be consistent and asymptotically normal if we have available a correctly specified parametric or semiparametric model for the missingness process. Furthermore, given a correctly specified model for the missingness process, we propose a locally efficient one-step estimator whose asymptotic variance attains the efficiency bound, if we correctly specify a lower-dimensional model for the conditional distribution of T given the covariates. The estimator remains consistent and asymptotically normal even if this latter submodel is misspecified. We conclude with a simulation experiment and a data analysis.

KEY WORDS: Asymptotically efficient; Asymptotically linear estimator; Cox proportional hazards model; Current status data; Influence curve.

1. INTRODUCTION

Consider a study in which interest lies in the marginal distribution F of a failure time random variable T . Suppose that T is never observed. Rather, for each individual we observe at a random monitoring (censoring) time C whether T exceeds C . This data structure is called *current status data*.

Previous work and examples of current status data have been provided by Diamond and McDonald (1991), Diamond, McDonald, and Shah (1986), Jewell and Shiboski (1990), Keiding (1991), and Sun and Kalbfleisch (1993), among others. In its nonparametric setting it is also known as interval censoring, case I (Groeneboom and Wellner 1992). Current status data commonly arise in epidemiological investigations of the natural history of disease and in animal tumorigenicity experiments. Jewell, Malani, and Vittinghoff (1994) gave two examples arising from studies of human immunodeficiency virus (HIV) disease.

In each of the aforementioned papers, inferences on the distribution of T were made under the assumptions that (a) T and C are independently distributed and (b) no data are available on additional time-independent or time-dependent covariates. Under (a) and (b), the nonparametric maximum likelihood estimator (NPMLE) of F is the pooled adjacent violators estimator for the estimation of the monotone regression $F(t) = E(\Delta|C = t)$ of Barlow, Bartholomew, Brenner, and Brunk (1972), where $\Delta = I(T \leq C)$ is the current status indicator at time C . The asymptotic distribution of this estimator under (a) has been analyzed by Groene-

boom and Wellner (1992). Under assumptions (a) and (b), the efficiency of the NPMLE of smooth functionals of F (such as its mean and variance) has been proved by Groeneboom and Wellner (1992), Huang and Wellner (1995) and van de Geer (1994). The purpose of this article is to develop methods for the estimation of smooth functionals of F when (b) and/or (a) is violated.

It will be pedagogically advantageous to motivate our methods by considering the following idealized study design for a mouse tumorigenicity experiment in which the outcome of interest is time to the development of that liver adenoma and study mice are randomly allocated to either placebo or active treatment with a suspected tumorigen. In such an experiment it is of interest to estimate the treatment arm specific distributions F of the time T until the development of liver adenoma. Suppose 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 $L = L(\cdot)$ be the entire weight process. We observe the weight process only up to time C ; that is we observe $\bar{L}(C) = \{L(u): 0 < u < C\}$. Because mice with liver adenomas tend to lose weight, $\bar{L}(C)$ and T are correlated. Suppose for the moment that the sacrifice times C are randomly chosen by the investigators, guaranteeing independence of T and C . A major goal of this article is to develop estimators of smooth functionals of F that, by incorporating information on the surrogate marker process $\bar{L}(C)$, are guaranteed both to be more efficient than the NPMLE that ignores data on $\bar{L}(C)$ and to remain consistent and asymptotically normal, whatever the joint distribution of (T, L) . Specifically, we propose a new class of

Mark J. van der Laan is Assistant Professor, School of Public Health, Division of Biostatistics, University of California, Berkeley, CA 94720. James M. Robins is Professor of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA 02115. The authors thank two anonymous referees and the associate editor for their helpful remarks and Alan Hubbard for his work on the simulations and data analysis. The authors also thank Steve Shiboski and Nancy Padian for providing the dataset. This research was supported by a FIRST award from the National Institute of General Medical Sciences National Institutes of Health (NIH) Grant (GM53722), and NIH grant R01-AI32475.

© 1998 American Statistical Association
Journal of the American Statistical Association
June 1998, Vol. 93, No. 442, Theory and Methods

“inverse probability of censoring weighted” (IPCW) estimators that accomplish both goals. In contrast to our weighted estimators, the NPMLE that incorporates data on $\bar{L}(C)$ fails to attain these goals because of the curse of dimensionality (Robins and Ritov 1997); indeed, the NPMLE is not even well defined.

The results described in the previous paragraph required the assumption that the monitoring time C was assigned completely at random. But in realistic settings this will often not be so, in which case the assumption that T and C are independent will also be false. Consequently, the NPMLE that ignores data on L will be inconsistent both for F and for smooth functionals of F . As an example, suppose that in our mouse tumorigenicity experiment the investigators wish, if possible, to sacrifice an animal soon after it has developed a tumor, to obtain more efficient estimates of F . To do so, they decide to increase the probability of sacrificing an animal in the interval $(t, t + \delta t)$ if the animal has begun to lose weight prior to t . Under this design, T and C will be dependent (because they both are correlated with the weight process), but the hazard of censoring (sacrifice) at t given the weight process $\bar{L}(t)$ up to t will not further depend on the unknown failure time T . That is,

$$\lambda(t|T, L) = \lambda(t|\bar{L}(t)), \quad (1)$$

where $\lambda(t|T, L)$ is the hazard of censoring, given the full data (T, L) . Formally, assumption (1) says that the conditional hazard of C at time t , given the full data T, L , is only a function of $(t, \bar{L}(t))$. We show that when T and C are dependent but (1) is true, our inverse probability of censoring weighted estimators of smooth functionals of T will remain consistent and asymptotically normal provided that we can consistently estimate the hazard function $\lambda(t|\bar{L}(t))$. In the idealized experiment described earlier, $\lambda(t|\bar{L}(t))$ will be known by design, because it is under the control of the investigator. But, in many settings $\lambda(t|\bar{L}(t))$ will not be known. For example, in our tumorigenicity experiment, it may happen that a number of animals die of other non-tumor-related diseases prior to sacrifice, at which point they are autopsied and the presence or absence of a liver adenoma is determined. In such animals, their monitoring time is their time of death C . For that case, one cannot be sure that (1) holds. Consequently, the investigator should try to incorporate into $\bar{L}(t)$ any factors that could be correlated with the presence of an underlying adenoma (i.e., correlated with T) that also predict the hazard of death (censoring) at time t . Even if the investigator were successful so that (1) were true, the functional form of $\lambda(t|\bar{L}(t))$ would remain unknown. In this setting it is necessary to posit a statistical model for $\lambda(t|\bar{L}(t))$. In this article, we emphasize modeling $\lambda(t|\bar{L}(t))$ by a time-dependent Cox-proportional hazards model. Our IPCW estimators will be consistent if (1) holds and our Cox proportional hazards model for $\lambda(t|\bar{L}(t))$ is correctly specified.

In current practice, carcinogenicity experiments have few predetermined times of sacrifice (see e.g., Dinse 1988) and do not collect time-dependent measurements (e.g., weight) until sacrifice. In this case, our estimators can still be used

to produce efficient estimates of $F(\cdot)$ at the sacrifice times. Note, however, that such carcinogenicity experiments have an inferior design relative to our idealized design described here because they do not allow nonparametric estimation of smooth functionals of F , use an inefficient sacrificing scheme independent of T , and do not exploit the possibility of using a surrogate process $\bar{L}(C)$ to improve estimation.

The critical assumption (1) is not empirically testable, because T is not observed. The need for untestable assumptions such as (1) and for the correct specification of models for high-dimensional processes, such as $\lambda(t|\bar{L}(t))$, is the rule in analyzing data obtained from observational studies. Our critical assumption (1) is equivalent to the assumption that the censoring (i.e., missingness) mechanism satisfies coarsening at random (CAR) (Robins 1996; Robins and Rotnitzky 1992). Coarsening at random was originally formulated by Heitjan and Rubin (1991) and was generalized by Gill, van der Laan, and Robins (1997) and Jacobsen and Keiding (1994). The usefulness of the CAR assumption in estimation of F in the presence of time-dependent surrogate processes $\bar{L}(t)$ has been argued by Robins and Rotnitzky (1992). Gill et al. (1997) showed that if CAR is the only assumption, then the model for the observed data is nonparametric and all regular asymptotically linear estimators of μ are asymptotically equivalent and efficient. By the curse of dimensionality, this means that estimators with reasonable moderate sample performance do not exist in this unrestricted model. But when $\lambda(t|\bar{L}(t))$ follows a semiparametric model, our proposed estimators perform reasonably in moderate samples and are still efficient at a submodel of interest. It is important to stress again that when the CAR assumption (1) holds but T and C are dependent, no alternatives to the approach described herein are available in the literature.

The article is organized as follows. In Section 2 we formalize our estimation problem. In Section 3 we introduce the class of inverse probability weighted estimators. We then use in Section 4 a preliminary inverse probability of censoring weighted estimator as an initial estimator and propose a one-step update that leads to a locally efficient estimator. Our proposed locally efficient estimator requires that we specify a model for the conditional distribution of T given $\bar{L}(u)$. The one-step estimator is locally efficient in the sense that whenever $\lambda(t|\bar{L}(t))$ is either known or correctly modeled, it attains the semiparametric variance bound for the model if in fact the model for T given $\bar{L}(u)$ is correct and yet remains consistent and asymptotically normal with good efficiency properties even if the latter model is misspecified. Indeed, in the extreme case in which the surrogate process $L(\cdot)$ is time-independent and happens to have correlation 1 with T (which is unknown to the experimenter) and the selected model for T given L is correct, the inability to observe T remarkably results in no loss of information. That is, our proposed locally efficient update of the IPCW estimator is asymptotically equivalent to the estimator obtained from the empirical distribution of T_1, \dots, T_n . Similarly, if $\bar{L}(t)$ happens to have correlation 1 with $I(T \leq t)$ (e.g., in the foregoing example all mice first begin to lose weight immediately on onset of the tumor), then our pro-

posed one-step estimator is asymptotically equivalent to the Kaplan–Meier estimator based on $(T_i \wedge C_i, I(T_i \leq C_i))$, $i = 1, \dots, n$, if the selected model for $F(t|\bar{L}(u))$ is correct. The consistency of our one-step estimator under misspecification of the model for T , given $\bar{L}(u)$, is protected by our assumption of a correctly specified model for the hazard $\lambda(t|T, L) = \lambda(t|\bar{L}(t))$ of censoring given (T, L) . The only price we pay for this robustness is that the calculation and estimation of the asymptotic variance of our one-step estimator requires explicit computation of projections on various Hilbert spaces, analogous to those in Robins (1993, 1996) and Robins and Rotnitzky (1992) for right-censored data.

In Section 5 we present the local efficiency theorem and a consistent estimator for the asymptotic variance of our locally efficient one-step estimator, which we then use to construct Wald confidence intervals for the functional of interest. The proof of the theorem is deferred to our technical report (van der Laan and Robins 1997). In Section 6 we compare our methods to the NPMLE for the marginal model in a simulation experiment. In Section 7 we use our methods to reanalyze data on time to HIV infection in the female partners of HIV-infected males in the California Partners Study. We conclude in Section 8 with a discussion in which we address in particular how one can generalize our methodology to estimate the onset distribution at a given point and to handle discrete monitoring times.

2. A FORMALIZATION OF OUR PROBLEM

For simplicity, we assume that T has support on a finite interval $[0, \tau]$. Let $L = \{L(u): 0 < u < \tau\}$. For each individual, we observe $Y = (C, \Delta = I(T \leq C), \bar{L}(C))$. Given a real-valued function r , we are concerned with nonparametric estimation of functionals $\mu(r) \equiv \int r(t)(1 - F(t)) dt$ of the distribution of F of T , based on n iid observations $Y_i = (C_i, \Delta_i, \bar{L}(C_i))$. We now show that $\mu(r)$ includes the mean and all higher-order moments of T for appropriate choices of $r(\cdot)$. Let $R(x) = \int_0^x r(t) dt$. Notice that by integration by parts, we have

$$\int r(t)(1 - F(t)) dt = R(t)(1 - F(t))|_0^\tau + \int R(t) dF(t).$$

Hence if $\lim_{t \rightarrow \tau} (1 - F(t))R(t)$ is 0, then an estimate of $\int r(1 - F) dt$ provides us with an estimate of $\int R dF$. In particular, if $r(t) = 1$, then $\mu(r) = ET$, and if $r(t) = kt^{k-1}$, then $\mu(r) = ET^k$. Moreover, by setting $r(t) = K(\{t - t_0\}/h)/h$ for some kernel K and bandwidth h , an estimator of $\mu(r)$ provides us with an estimator of $S(t_0) = 1 - F(t_0)$. For notational convenience, we also use L to denote the function $L(\cdot)$. We also use this same shorthand for other functions. In addition, define $X \equiv (T, L)$ and let F_X be the distribution of X . We make no assumptions about this distribution. Let $G(\cdot|X)$ be the conditional distribution function of C given X ; we call it the censoring mechanism or missingness process. We assume that $G(\cdot|X = x)$ is absolutely continuous with respect to the Lebesgue measure

with density $g(c|x)$, which satisfies

$$g(c|(T, L)) = h(c, \bar{L}(c))$$

for some function h of $(c, \bar{L}(c))$. (2)

Let $\lambda(c|x = (t, l))$ be the hazard corresponding to $g(c|x)$: $\lambda(c|x) = g(c|x)/\bar{G}(c|x)$ with $\bar{G}(c|x) = 1 - G(c|x)$. Because $\bar{l}(c)$ determines $\bar{l}(u)$, $u \leq c$, we have that $g(c|x) = \lambda(c|x) \exp(-\int_0^c \lambda(u|x) du)$ is a function of $\bar{l}(c)$ if and only if

$$\lambda(c|X = (t, l)) = m(c, \bar{l}(c))$$

for some function m of $(c, \bar{l}(c))$. (3)

In turn, (3) is equivalent to (1) of Section 1. Condition (2) implies that the censoring mechanism satisfies CAR (i.e., it is noninformative, given the observed covariates). In the appendix of van der Laan and Robins (1997) we prove that (2) is in fact equivalent to the apparently weaker assumption of CAR. Note that if L is time independent and thus is actually uncensored, (i.e., $\bar{L}(C) = L$), then (3) states that C and T are independent, given the covariates L . A special case of CAR is that $g(c|x) = g(c)$ so that T and C are marginally independent. Gill, van der Laan, and Robins (1997) showed that the assumption that the data are CAR, by itself, places no restrictions on the joint distribution of the observed data; that is, the observed data model is completely nonparametric.

As discussed in Section 1, whenever the density $g(c|x)$ is not known by design, it is usually necessary to model it to avoid the curse of dimensionality. We assume a parametric or semiparametric model for $g(c|x)$ or, equivalently, for $\lambda(c|x)$, indexed by a parameter η :

$$g(c|x) = h_\eta(c, \bar{l}(c)), \tag{4}$$

where h_η , $\eta \in \Gamma$, is a parametric or semiparametric parameterization of h . In particular, we emphasize the Cox proportional hazards model as a model for $\lambda(c|x)$. Our inverse probability of censoring weighted estimators require that we obtain an $n^{1/4}$ -consistent estimator of $g(c|x)$. Thus if one assumes a Cox proportional hazard, then the usual discrete estimator of the baseline cumulative hazard must be replaced by a smooth version, as discussed further in Section 4. $g(c|x)$ can be estimated by smoothing the discrete density obtained from the partial likelihood estimators for the Cox-proportional hazards model, as we do in Section 6.

3. INVERSE PROBABILITY OF CENSORING WEIGHTED ESTIMATORS

The key identity that we exploit to construct our IPCW estimators is

$$\begin{aligned} E\left(\frac{(1 - \Delta)r(C)}{g(C|X)}\right) &= EE\left(\frac{(1 - \Delta)r(C)}{g(C|X)} \middle| X\right) \\ &= E\left(\int_0^\tau I(T \geq c)r(c) dc\right) \\ &= \int_0^\tau r(c)(1 - F(c)) dc. \end{aligned}$$

This suggests the following IPCW estimator of $\mu = \int r(t) (1 - F) dt$:

$$\mu_n^0 = \frac{1}{n} \sum_{i=1}^n \frac{I(\Delta_i = 0)r(C_i)}{g_{\eta_n}(C_i|X_i)}, \quad (5)$$

where g_{η_n} is a consistent estimator of g obtained assuming the model $\{g_\eta: \eta \in \Gamma\}$. Recall that by (4), $g_\eta(c|x) = g_\eta(c|L = l) = h_\eta(c, \bar{l}(c))$ and hence is a function of the observed data. By setting $r(t) = K(\{t - t_0\}/h)/h$ for a kernel K and bandwidth h , we obtain the following estimator of $S(t_0) = 1 - F(t_0)$:

$$S_n^0(t_0) = \frac{1}{n} \sum_{i=1}^n \frac{(1 - \Delta_i) \frac{1}{h} K(\{C_i - t_0\}/h)}{g_{\eta_n}(C_i|X_i)}, \quad (6)$$

where K is a kernel and h is a bandwidth that converges to 0 at an appropriate rate.

For example, if $g(c|x) = g(c)$ (i.e., we have independent censoring), and if g_{η_n} is a kernel density estimator of g based on the C_1, \dots, C_n , then F_n^0 is just the regularized MLE in the current status model (without a covariate process) as studied by van der Laan, Bickel, and Jewell (1997). For smooth functionals such as the moments of F , μ_n^0 will be asymptotically linear where an estimator μ_n is asymptotically linear with influence curve IC if it can be approximated by a sum of iid random variables in the sense of

$$\mu_n - \mu = \frac{1}{n} \sum_{i=1}^n IC(C_i, \Delta_i, \bar{L}_i(C_i)) + o_P(1/\sqrt{n}),$$

for some function IC of $(C, \Delta, \bar{L}(C))$ with mean 0 and finite variance. Then $\sqrt{n}(\mu_n - \mu)$ is asymptotically normal with mean 0 and variance given by the variance of the influence curve. Van der Laan (1995) provided conditions under which the estimator μ_n^0 is asymptotically linear. The most important condition is that for every (T, L) and all c in the support $[0, \tau]$ of T with $r(c) > 0$, $g(c|T, L) > \varepsilon > 0$ for some $\varepsilon > 0$. Such a condition is not unexpected, because $F(t) = E(\Delta|C = t)$ if T is independent of C .

4. THE LOCALLY EFFICIENT ONE-STEP ESTIMATOR

In this section we construct a locally efficient one-step estimator by adding to the estimator μ_n^0 (5) an estimate of the empirical mean of the efficient influence function. Thus our first task is to provide a representation of the efficient influence function. This representation has two pieces. The first is given by the influence function of μ_n^0 when using the known $g(c|X)$, which is given by

$$IC_0(C, \Delta, \bar{L}(C)|\mu, G) = \frac{r(C)(1 - \Delta)}{g(C|X)} - \mu. \quad (7)$$

The second piece is a projection of IC_0 on the so-called orthogonal complement to the tangent space defined in Section 5. The projection is a function $IC_{nu}^*(\cdot|F_X, G)$

$(C, \Delta, \bar{L}(C))$:

$$IC_{nu}^* = \int \left\{ \frac{r(u)\{1 - F(u|\bar{L}(u))\}}{g(u|X)} - \frac{1}{G(u|X)} \times \left\{ \int_u^\infty r(t)\{1 - F(t|\bar{L}(u))\} dt \right\} \right\} dM(u), \quad (8)$$

where

$$dM(u) \equiv I(C \in du) - \Lambda(du|X)I(C > u). \quad (9)$$

It is important to emphasize that for any function $H(u, \bar{L}(u))$, the stochastic integral

$$\int H(u, \bar{L}(u)) dM(u) = H(C, \bar{L}(C)) - \int_0^C H(u, \bar{L}(u))\Lambda(du|X)$$

is a function of the observed data, because $\lambda(u|X)$ depends on X only through $\bar{L}(u)$. Similarly, the integrand in (8) depends on X only through $\bar{L}(u)$.

Sometimes, as on the left side of (8), we will suppress the dependence of functions on both F_X, g and the data. In the appendix of van der Laan and Robins (1997), we showed that the efficient influence curve for estimation of μ is

$$IC^*(C, \Delta, \bar{L}(C)|F_X, G, \mu) \equiv IC_0(C, \Delta, \bar{L}(C)|\mu, G) - IC_{nu}^*(C, \Delta, \bar{L}(C)|F_X, G).$$

If L is a vector of time-independent covariates and thus always completely observed (i.e., $\bar{L}(C) = L$), then IC_{nu}^* reduces to

$$IC_{nu}^*(C, \Delta, L|F_X, G) = \frac{r(C)}{g(C|X)} \{1 - F(C|L)\} - \int r(u)\{1 - F(u|L)\} du. \quad (10)$$

Let $IC_{nu}^*(\cdot|F_{X,n}, G_n)$ be an estimator of $IC_{nu}^*(\cdot|F_X, G)$ obtained by substitution of estimators of $F(t|\bar{L}(u))$, $t \geq u$, and $g(c|x)$, where $G_n = G_{\eta_n}$ suppresses in the notation the dependence on the parameter η . Note that IC_{nu}^* depends on G through $g(c|X)$ and through the measure $dM(u)$. Theorem 1 in Section 5 shows that asymptotic normality of the one-step estimator requires that the estimate of $g(c|X)$ in (7)–(8) be at least $n^{+1/4}$ consistent for $g(c|X)$, thus requiring a smooth. We discuss smooth estimation of $g(c|X)$ according to a given model in Section 6. In contrast, Theorem 1 also teaches us that we can estimate $\Lambda(du|X)$ in $dM(u)$ with a discrete estimator of G ; so there is no need to evaluate integrals with respect to continuous measures. In the next section we propose an estimation method for $F(t|\bar{L}(u))$ in (8). The locally efficient one-step estimator is given by

$$\mu_n^1 = \mu_n^0 + \frac{1}{n} \sum_{i=1}^n \left\{ IC_0(C_i, \Delta_i, \bar{L}_i(C_i)|\mu_n^0, G_n) - IC_{nu}^*(C_i, \bar{L}_i(C_i)|F_{X_n}^n, G_n) \right\}, \quad (11)$$

where μ_n^0 is the estimator (5) using $g_n(c|x)$. Let $Pf \equiv \int f dP$ for a probability measure P and measurable function f . Let P_n be the empirical CDF so that $P_n f = 1/n \sum_{i=1}^n f(Y_i)$. Note that in fact $P_n IC_0(\cdot|\mu_n^0, G_n) = 0$, so that we could delete the IC_0 term in (11), as did Robins and Rotnitzky (1992). But, we chose to include the IC_0 term in the representation (11) to show that μ_n^1 is just the classical one-step estimator as defined by Bickel, Klaassen, Ritov, and Wellner, 1993 (p. 395); that is, by its definition, μ_n^1 is the first step in the Newton–Raphson algorithm for solving the estimating equation

$$0 = \frac{1}{n} \sum_{i=1}^n IC^*(C_i, \Delta_i, \bar{L}_i(C_i)|F_X^n, G_n, \mu) \quad (12)$$

for μ , where we chose μ_n^0 as the initial estimator. This follows from the fact that the derivative of the estimating equation (12) with respect to μ equals -1 . Because (12) is linear in μ , the Newton–Raphson algorithm converges at the first step so that μ_n^1 equals the solution of (12). It is easy to understand that one can improve on a given estimating equation for μ , say $0 = \sum_{i=1}^n IC_0(Y_i|\mu, G_n)$ solved by μ_n^0 , by replacing the equation by another estimating equation with smaller variance but the same derivative with respect to μ . The estimating (12) does this by subtracting from IC_0 an estimate of an optimally chosen negatively correlated random variable that is not a function of μ , namely $IC_{nu}^*(Y|F_X^n, G_n)$. This provides a heuristic explanation of the fact that μ_n^1 typically improves on μ_n^0 .

4.1 A Method for Estimation of IC_{nu}^* .

Consider first the case in which L is time independent and thus fully observed. Then IC_{nu}^* is given by (10), and consequently, estimation of IC_{nu}^* requires estimation of $F(u|L_i)$ for observations $i = 1, \dots, n$ for u in the set $C = \{C_1, \dots, C_n\}$ of observed monitoring times.

In a particular application, a natural model for $F(u|L) = P(T < u|L)$ might arise, and one could estimate $F(u|L)$ with the maximum likelihood procedure. Here we provide a generic estimation method based on an additive logistic model. Suppose that L is a k -dimensional vector of covariates, so $L = (L_1, L_2, \dots, L_k) \in \mathbb{R}^k$. By CAR, T and C are independent given L . Hence

$$F(c|L) = P(T < c|C = c, L) = E(\Delta|C = c, L), \quad (13)$$

where again $\Delta = I(T < C)$. Consequently, we can estimate $F(c|L)$ by estimating the regression of Δ on the $k + 1$ -dimensional vector (C, L) . To avoid the curse of dimensionality, we assume a generalized additive logistic regression model,

$$E(\Delta|C = c, X = (t, l)) = \frac{1}{1 + \exp(f_0(c) + f_1(l_1) + \dots + f_k(l_k))},$$

for unknown functions f_0, \dots, f_k , where f_0 is constrained to be monotone. The functions f_0, \dots, f_k can be estimated using the S-PLUS function GAM, which allows the user to specify a parametric model for one or more of the func-

tions f_m if desired (Hastie and Tibshirani 1990). The model considered by Rossini and Tsiatis (1996) is a special case of this model in which f_1, f_2, \dots, f_k are enforced to be linear. Rossini and Tsiatis (1996) considered estimation of the model using a sieve-MLE estimator. In contrast, Huang (1994) modeled the distribution of T given L with a Cox proportional hazards model and carried out estimation via maximum likelihood.

Consider now the case in which $L(u)$ is a time-dependent covariate process and we observe $\bar{L}(C)$. Then IC_{nu}^* is given by (8). Thus it is necessary to estimate $F(t|\bar{L}(u))$ for a given (t, u) with $t \geq u$ and $U \in C$. The estimator that we propose is motivated by the CAR identity

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

with

$$J(u, t) = \frac{\bar{G}(u|X)I(C \geq t)F(t|\bar{L}(t), C = t)}{\bar{G}(t|X)}.$$

First, for fixed t construct summary measures W_1, \dots, W_k (i.e., functions) of $\bar{L}(t)$, which hopefully contain the most relevant information for predicting T . Then model $F(t|\bar{L}(t), C = t) = E(\Delta|C = t, \bar{L}(t))$ with an additive logistic regression model, where the functions $f_m(l_m)$ are now replaced by functions $f_m(w_m)$, and use GAM to obtain an estimator $\hat{F}(t|\bar{L}(t))$. This results in an estimator of $J(u, t)$,

$$\hat{J}(u, t) = \frac{\bar{G}_n(u|X)I(C \geq t)\hat{F}(t|\bar{L}(t))}{\bar{G}_n(t|X)}.$$

Finally, obtain $\hat{F}(t|\bar{L}(u))$ by regressing the $\hat{J}_i(u, t)$ on functions of $\bar{L}(u)$, among subjects with $C > u$.

If the density of C depends only on the time-independent covariates—that is, $g(c|X) = g(c|L(0))$ —then we can simplify the foregoing procedure by noting that

$$F(t|\bar{L}(u)) = E(\Delta|C = t, \bar{L}(u), C > u). \quad (14)$$

In this case we can estimate $F(t|\bar{L}(u))$ with GAM as earlier by regressing Δ on C and summary measures of $\bar{L}(u)$ among the subjects with $C > u$, where the GAM procedure is repeated for each $u \in C$.

5. LOCAL EFFICIENCY RESULT AND CONSTRUCTION OF ASYMPTOTIC CONFIDENCE INTERVALS

When the model for $F(t|\bar{L}(u))$ is misspecified, the asymptotic variance of μ_n^1 will depend on the model for the nuisance parameter $g(c|x)$. Characterizing the form of this dependence requires that we introduce the notion of a tangent space.

Denote the Hilbert space of functions of $(C, \Delta, \bar{L}(C))$ with finite variance and mean 0, endowed with the covariance norm

$$\|v\|_{P_{F_X, G}} \equiv \sqrt{\int v^2 dP_{F_X, G}},$$

by $L_0^2(P_{F_X, G})$. The tangent space $T_1(P_{F_X, G})$ for the parameter F_X is by definition the closure of the linear extension

in $L_0^2(P_{F_X,G})$ of the scores at $P_{F_X,G}$ from correctly specified parametric models for the distribution F_X of X . The tangent space $T_2(P_{F_X,G})$ for the parameter G_η 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 . For convenience, we often denote these tangent spaces by T_1 and T_2 , suppressing the dependence on $P_{F_X,G}$.

The CAR assumption (2) implies that the spaces T_1 and T_2 are mutually orthogonal. Given this fact, it follows from Bickel et al. (1993) that the efficient influence curve is given by

$$IC^* = IC_0 - IC_{nu}^*, \tag{15}$$

where $IC_{nu}^* = \Pi(IC_0|T_1^\perp)$ and T_1^\perp is the orthogonal complement of T_1 . Here $\Pi(\cdot|T_1^\perp): L^2(P_{F_X,G}) \rightarrow L^2(P_{F_X,G})$ is the projection operator on T_1^\perp . We have previously proved (van der Laan and Robins 1997) that the explicit form of IC_{nu}^* is given by (8).

From Bickel et al. (1993), we have that an estimator μ_n is asymptotically efficient if it is asymptotically linear with influence curve IC^* . Theorem 1 shows that if our model $F(t|\bar{L}(u))$ is correctly specified, the one-step estimator μ_n^1 is indeed asymptotically linear with influence curve IC^* and thus is asymptotically efficient. Moreover, μ_n^1 has the additional feature that it remains a consistent and asymptotically normal estimator of μ even when our model for $F(t|\bar{L}(u))$ is misspecified. This is due to the fact that, by (8), $IC_{nu}^* = \int H(u, \bar{L}(u)) dM(u)$ for a particular function H , where the stochastic integral $\int H(u, \bar{L}(u)) dM(u)$ has conditional mean 0, given X , for any function H , because $E(dM(u)|X) = 0$. This explains why μ_n^1 will still be consistent if H is estimated inconsistently.

With these preliminaries, we are now ready to state our main theorem. Before doing so, we note that condition (b) of Theorem 1 is a general empirical process condition. (For empirical process theory, refer to van der Vaart and Wellner 1996.) We decided not to derive more primitive conditions that imply condition (b), because this condition is technical and model dependent. Recall the notation $Pf = \int f(x) dP(x)$.

Theorem 1. We assume that the following conditions hold:

- a. $\tau(c)/g(c|l) > \varepsilon > 0$ a.e. with respect to $dF_{C,L}(c,l)$, for some $\varepsilon > 0$.
- b. $IC_0(\cdot|\mu_n^0, G_n) - IC_{nu}^*(\cdot|F_{X,n}, G_n)$ falls in a $P_{F_X,G}$ -Donsker class with probability tending to 1.
- c. $F_n(t|\bar{L}(u))$ converges uniformly in $(t, \bar{L}(u))$, $t \geq u$ to a $F^1(t|\bar{L}(u))$, and $g_n(c|L)$ converges uniformly in (c, L) to $g(c|L)$, both on the support of $P_{F_X,G}$.
- d. $\sup_{c,L} |g_n(c|L) - g(c|L)| \sup_{t \in [u, \tau], L} |F_n(t|\bar{L}(u)) - F^1(t|\bar{L}(u))| = o_P(1/\sqrt{n})$, (16)

where τ is the endpoint of the support of T and the supremum is taken over the support of $P_{F_X,G}$.

- e. $\Phi(G_n)$ is an asymptotically efficient estimator of $\Phi(G)$ for a model containing the true G with tangent space $T_2(P_{F_X,G})$.

Then μ_n^1 is asymptotically linear with influence curve given by

$$IC(\cdot|F_X^1, G, \mu) = \Pi\{IC_0(\cdot|\mu, G) - IC_{nu}^*(\cdot|F_X^1, G)|T_2^\perp(P_{F_X,G})\}.$$

In particular, if $IC_{nu}^*(\cdot|F_X^1, G) = IC_{nu}^*(\cdot|F_X, G)$, then μ_n^1 is asymptotically efficient.

For the case where our semiparametric model for $g(c|x)$ is characterized by the restriction $g(c|x) = g(c)$ for an unspecified density g , we have for any $V \in L^2(P_{F_X,G})$, $\Pi(V|T_2) = E(V(C, \Delta, \bar{L}(C))|C)$. For the case in which $\bar{L}(u)$ is possibly time-dependent and $g(c|x)$ follows a Cox proportional hazards model with k -dimensional covariate $\mathbf{W}(\bar{L}(u)) = (W_1(\bar{L}(u)), \dots, W_k(\bar{L}(u)))^T \in \mathbb{R}^k$,

$$\lambda(c|x) = \lambda_0(c) \exp(\beta^T \mathbf{W}(\bar{L}(c))), \tag{17}$$

we have

$$\begin{aligned} \Pi(V|T_2) &= \int E(V^*(u, \bar{L}(u))|C = u) dM(u) \\ &+ E \left(\int V^*(u, \bar{L}(u)) dM(u) \right. \\ &\quad \left. \times \left\{ \int \mathbf{W}'(u, \bar{L}(u)) dM(u) \right\}^T \right) \\ &\times \Sigma^{-1} \int \mathbf{W}'(u, \bar{L}(u)) dM(u), \end{aligned} \tag{18}$$

where

$$\begin{aligned} \mathbf{W}'(u, \bar{L}(u)) &\equiv \mathbf{W}(\bar{L}(u)) - E(\mathbf{W}(\bar{L}(u))|C = u), \\ V^*(u, \bar{L}(u)) &\equiv E(V(u, I(u < T), \bar{L}(u)) \\ &\quad - V(C, \Delta, \bar{L}(C))|C > u, \bar{L}(u)), \end{aligned} \tag{19}$$

and the (i, j) th element of the covariance matrix Σ is given by the expectation of

$$\int \int W_i'(u, \bar{L}(u)) W_j'(u, \bar{L}(u)) \lambda(u|L) I(C > u) du.$$

Condition (e) of Theorem 1 will typically be satisfied when $g_n(\cdot|X)$ is a smoothed MLE, where the smooth needs to be chosen so that $g_n(\cdot|X)$ is at least $n^{-1/4}$ consistent. Note that Theorem 1 also gives the limiting distribution of the estimator μ_n^0 : By (8), simply choose the estimator $F_n(t|\bar{L}(u))$ to be identically 1 for all t and u . Then $\mu_n^1 = \mu_n^0$, and $F_X^1(t|\bar{L}(u))$ is also identically 1. Furthermore, note that when in truth $g(c|L) = g(c)$, the marginal censoring model and the Cox proportional hazards model are both true. But the tangent space T_2 associated with the less restrictive Cox proportional hazards censoring model strictly includes the tangent space T_2 associated with the

marginal censoring model. This implies, by Theorem 1, that the estimator μ_n^0 (5) that estimates $g(c|X)$ by imposing the Cox model, estimating the coefficient β by partial maximum likelihood, and estimating the baseline hazard function using a kernel smooth is at least as efficient as the estimator μ_n^0 that estimates $g(c|X)$ by ignoring data on L and applying a kernel density estimator to the data $C_i, i = 1, \dots, n$. As discussed earlier, this latter estimator is asymptotically equivalent to the NPMLE $\hat{\mu}_{NPMLE}$ that ignores data on L . Thus when censoring is independent—that is, $g(c|X) = g(c)$ —we have proved that (as promised in Section 1), we are able to construct an estimator guaranteed to be at least as efficient as $\hat{\mu}_{NPMLE}$ that ignores data on L .

Construction of Wald-Type Confidence Interval. Consider the case where we impose the model $g(c|x) = g(c)$ with g unspecified and we estimate the unknown density $g(c)$ using a kernel density estimator g_n that ignores data on L with appropriately chosen bandwidth. Then the influence curve $IC(\cdot|F_X^1, G, \mu)$ of μ_n^1 can be estimated as the residual from the nonparametric estimation of the regression of $IC_0(Y_i|\mu_n^0, G_n) - IC_{nu}^*(Y_i|F_X^1, G_n)$ on $C_i, i = 1, \dots, n$. The resulting estimator \widehat{IC} of IC is, under the assumptions of Theorem 1, $L^2(P_{F_X, G})$ consistent. Because $\sqrt{n}(\mu_n^1 - \mu)$ is asymptotically normal with mean 0 and variance equal to the variance of $IC(C_i, \Delta_i, \bar{L}_i(C_i)|F_X^1, G, \mu)$, we can construct an asymptotic Wald confidence interval for μ by estimating the variance of $IC(C_i, \Delta_i, \bar{L}_i(C_i)|F_X^1, G, \mu)$ by

$$\frac{1}{n} \sum_{i=1}^n \widehat{IC}^2(C_i, \Delta_i, \bar{L}_i(C_i)),$$

where \widehat{IC} is consistent for IC . If $F_n(t|\bar{L}(u))$ is consistent for $F(t|\bar{L}(u))$, \widehat{IC} will converge to the efficient influence curve. In practice, however, it will not be known whether $F_n(t|\bar{L}(u))$ is consistent. But whether it is or not, the actual coverage rate of our Wald intervals will converge to its nominal rate in large samples. In the case where $g(c|x)$ follows the Cox proportional hazards model (17), it is necessary to estimate $\Pi(V|T_2)$ as given in Theorem 1.

6. SOME SIMULATION RESULTS

In the simulations we refer to the estimator μ_n^0 that estimates $g(c|X)$ by a kernel density estimator applied to the data $C_i, i = 1, \dots, n$ as the estimator “univariate.” We refer to the estimator μ_n^0 based on the Cox proportional hazards model as the estimator “Cox.” In the simulations herein, we smoothed the usual discrete baseline hazard estimator for the Cox model using a Gaussian kernel with an edge

correction and bandwidth selected by cross-validation. The consistency properties of such a kernel smoother have been derived by Andersen, Borgan, Gill, and Keiding (1993). Finally, we denote the one-step estimator μ_n^1 as “1-step.”

In this section we compare the performance of the competing estimators pool adjacent violator (PAV; i.e., the NPMLE $\hat{\mu}_{NPMLE}$ that ignores data on L), Cox, and 1-step in a number of simulation experiments. It is useful to remember that the estimator PAV is consistent only for μ under the assumption that $g(c|X) = g(c)$. In all of our simulation experiments, the functional μ was the truncated mean survival time $\int r(t)(1 - F(t)) dt$, where either $r(t) = I(t \leq \tau)$ or $r(t) = I(\tau_1 < t < \tau_2)$. In all experiments we used a univariate time-independent covariate L that is uniformly distributed on the interval $[-1, 1]$.

The data-generating process of the first simulation experiment was

$$g(c|l) = \frac{\exp(-10 + c)}{(1 + \exp(-10 + c))^2} \exp(1.75l)$$

and

$$F(c|l) = \frac{1}{1 + \exp(-10 + c + 1.75l)}.$$

Under this data-generating mechanism, T and C are marginally dependent but independent given the covariate L . Thus, PAV is inconsistent. Further, the density $g(c|l)$ in this and all subsequent simulations lies in the Cox proportional hazards model (17) with $W(\bar{L}(u)) = L$, which was assumed in computing the estimators Cox and 1-step. We set $r(t) = I(\tau_1 < t < \tau_2)$, where τ_1 and τ_2 are the .1 and .9 quantiles of the marginal distribution of C . We need to do this because we can estimate F well only on an interval where $g(c|l) > \delta > 0$ and a full mean would use all values of F . In our first simulation we compare our proposed estimators of the truncated mean $\mu = \int_{\tau_1}^{\tau_2} (1 - F)(t) dt$ with the PAV estimator that ignores the covariates. We estimated $F(c|l)$ by fitting, by maximum likelihood, the linear logistic model

$$\text{logit}E(\Delta|C = t, L = l) = \alpha_0 + \alpha_1 t + \alpha_2 l, \quad (20)$$

which is correctly specified under our data-generating process. It follows that the 1-step estimator will be semiparametric efficient. Results are given in Table 1. In terms of MSE, PAV is beaten by Cox, which in turn is outperformed by the efficient estimator 1-step.

We remark that the gain of the IPCW estimator relative to the PAV depends on the data-generating distribution and the model chosen for the data-generating mechanism.

In the next simulation study we chose

$$g(c|l) = f(c|l) = \lambda(l) \exp(-\lambda(l)c),$$

where $\lambda(l) = .25 \exp(2l)$. Under this data-generating process, we were able to obtain stable estimators by choosing $r(t) = I(t \leq \tau)$, where τ is now the .95 quantile of the distribution of C . In constructing the 1-step estimator, we again estimated $F(c|l)$ using the logistic model (20). Note that under this second data-generating process, the logistic model (20) is misspecified, so our estimator $F(c|l)$ is

Table 1. MSE for Estimation of $\mu = \int_{\tau_1}^{\tau_2} (1 - F)(t) dt$, $n = 500$ Based on 625 Replicates

Estimator	100 × MSE	Relative MSE
PAV	4.3	1.0
Cox	3.6	1.2
1-step	1.2	3.7

NOTE: C and T depend on each other through L , and we guess the right model for $F(c|L)$.

inconsistent; results are given in Table 2. Again note that Cox and 1-step both outperform the inconsistent estimator PAV. But 1-step no longer is more efficient than Cox, presumably because $F(c|l)$ was badly biased.

In the following simulation, C is distributed uniformly $[0, 10]$ and independently of T . Specifically,

$$g(c|l) = \frac{1}{10}$$

and

$$F(c|l) = \frac{1}{1 + \exp(-30 + 6c + 10l)}$$

We estimated $F(c|l)$ consistently with the linear logistic regression model (20). For this simulation with independent censoring, we add both the "univariate" estimator (which uses a kernel density estimator that ignores data on L for the estimate of $g(c)$) and a one-step estimator that also uses the same kernel density for the estimate of $g(c)$. The results are reported in Table 3.

In this simulation, the Univariate, PAV, and Cox exhibit similar performance, and 1-step outperforms them. These results are in agreement with theoretical asymptotic relative efficiencies (AREs) of the estimators in Table 3, except that asymptotically, Cox is more efficient than PAV or univariate.

7. DATA ANALYSIS

We obtained data on 88 female partners of HIV infected males from the California Partners' Study, an ongoing investigation of heterosexual HIV transmission in partners of infected index cases (Padian et al. 1987; Shiboski and Jewell 1992). Participants are recruited through referrals from physicians, research studies, and local public health departments. Upon entry, serum samples are drawn from study subjects to determine status with regard to HIV infection. In addition, detailed medical, contraceptive, and behavioral histories are obtained. Finally, couples are interviewed to determine the total number of sexual contacts between the partners since the time of infection of the index case. The estimation of the distribution of the time T from infection of the index case until infection of the case's sexual partner is of interest, with T measured as the number of sexual contacts. Data on the following variables for the partners were available:

- serostat: infection indicator of female partner at monitoring (interview) time
- length: monitoring time C , measured by the number of sexual contacts since infection of the index case

Table 2. MSE for Estimation of $\int_0^\tau (1 - F(t)) dt$, $n = 500$
Based on 625 Replicates

Estimator	100 × MSE	Relative MSE
PAV	69.4	1.0
Cox	11.0	6.3
1-step	22.7	3.1

NOTE: C and T depend on each other through Z , and we guess the wrong model for $F(c|Z)$.

Table 3. MSE for Estimation of ET , $n = 500$
Based on 625 Replicates

Estimator	100 × MSE	Relative MSE
PAV	1.2	1.0
Univariate	1.2	.9
Cox	1.2	.9
1-step (Kernel)	.5	2.1
1-step (Cox)	.5	2.1

NOTE: C and T independent, and we guess the right model for $F(c|l)$.

- age: female partner's age in years at the time of recruitment
- bleeding: indicator of reported bleeding during intercourse prior to recruitment (1 = yes)
- stdhx: indicator of history of other STDs (1 = yes)
- nocondom: indicator of any condom use (1 = no).

For estimation of $F(c|L) = E(\Delta|C = c, L)$, we specified a linear logistic regression model with covariates length, bleeding, condoms, STD, bleeding*condoms, and bleeding*STD. A GAM analysis in S-PLUS showed that deviations from this model were not significant. Second, an analysis of the Cox proportional hazards model for C and the covariates showed that the covariates were independent of C . Based on our experience in the simulation experiments we decided to estimate $g(c|x)$ with a kernel density estimator with cross-validated bandwidth.

We estimated the truncated mean $\int_0^\tau (1 - F(t)) dt = \int_0^\tau t dF(t) + \bar{F}(\tau)\tau$ with τ the .975 quantile of the C_i 's, using the NPMLE $\hat{\mu}_{NPMLE}$ that ignores data on the covariates and the one-step estimator. We found $\tau = 2,334.0$, $\hat{\mu}_{NPMLE} = 1,602.8$ and 1-step = 1,579.4. The variance of the influence curve for $\hat{\mu}_{NPMLE}$ and 1-step were estimated as 3,827,670 and 3,551,342. Thus a .95 asymptotic confidence interval for the truncated mean based on the one-step estimator is given by

$$1,579.4 \pm 1.96 \sqrt{3,551,342/88} = 1,579.4 \pm 393.8 \\ = [1,185.6, 1,973.2].$$

When we repeated the analysis with τ now as the .85 quantile of the C_i 's, we obtained $\tau = 980$, $\hat{\mu}_{NPMLE} = 788.7$, and 1-step = 760. The corresponding variances of the influence curves of $\hat{\mu}_{NPMLE}$ and 1-step were 250,794.6 and 215,325.8. A 95% confidence interval based on the one-step estimator is now given by

$$760 \pm 97 = [663, 857].$$

As expected, estimation of the tail created extra variability. Therefore, one obtains more precise estimates of (strongly) truncated means.

8. CONCLUSION AND DISCUSSION

The foregoing simulation experiments show that the practical performance of the proposed one-step estimator is in keeping with its excellent theoretical properties. Further simulation studies by van der Laan and Hubbard (1997) suggest that this performance is maintained even at sample sizes smaller than the ones used in our simulation experiments. Furthermore, van der Laan and Hubbard showed that

this approach provides excellent estimates of the distribution function $F(t)$ by specifying $r(c) = K(\{c-t\}/h)/h$ for some kernel K and bandwidth h . This one-step estimator of $F(t)$ uses the information contained in the covariates L in a locally optimal way and performed remarkably well in simulation experiments. If C is discrete at t , then our one-step estimator still yields excellent estimates of $F(t)$ by setting $r(c) = I(c=t)$ and letting $g_n(t|L)$ be an estimator of the conditional probability that the sacrificing time equals t , given L .

The methods presented here can be also used to construct highly efficient estimators for interval censored data with several monitoring times in the presence of covariates (see van der Laan and Hubbard 1997). Generalization of our methods to estimate regression parameters with current status data in proportional hazards and linear regression models will be reported elsewhere.

[Received October 1995. Revised November 1997.]

REFERENCES

- Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.
- Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972), *Statistical Inference Under Order Restrictions*, New York: Wiley.
- Bickel, P. J., Klaassen, A. J., Ritov, Y., and Wellner, J. A. (1993), *Efficient and Adaptive Inference in Semi-Parametric Models*, Baltimore: Johns Hopkins University Press.
- Diamond, I. D., and McDonald, J. W. (1991), "The Analysis of Current Status Data," in *Demographic Applications of Event History Analysis*, eds. J. Trussell, R. Hankinson, and J. Tilton, Oxford, U.K.: Oxford University Press.
- 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.
- Gill, R. D., van der Laan, M. J., and Robins, J. M. (1997), "Coarsening at Random: Characterizations, Conjectures and Counter-Examples," in *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, eds. D. Y. Lin and T. R. Fleming, New York: Springer-Verlag, pp. 255-295.
- Groeneboom, P., and Wellner, J. A. (1992), *Information Bounds and Non-parametric Maximum Likelihood Estimation*, Boston: Birkhäuser.
- Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models*, London: Chapman and Hall.
- Heitjan, D. F., and Rubin, D. B. (1991), "Ignorability and Coarse Data," *The Annals of Statistics*, 19, 2244-2253.
- Huang, J. (1996), "Efficient Estimation for the Proportional Hazards Model With Interval Censoring," *The Annals of Statistics*, 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," *The 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.
- 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.
- Pollard, D. (1990), *Empirical Processes: Theory and Applications*, Regional Conference Series in Probability and Statistics 2, Hayward, CA: Institute of Mathematical Statistics.
- Ritov, Y., and Wellner, J. A. (1988), "Censoring, Martingales, and the Cox Model," *Contemporary Mathematics*, 80, 191-219.
- Robins, J. M. (1993), "Information Recovery and Bias Adjustment in Proportional Hazards Regression Analysis of Randomized Trials Using Surrogate Markers," in *Proceedings of the Biopharmaceutical Section, American Statistical Association*, pp. 24-33.
- (1996), "Locally Efficient Median Regression With Random Censoring and Surrogate Markers," in *Lifetime Data: Models in Reliability and Survival Analysis*, eds. N. P. Jewell et al., Dordrecht: Kluwer, pp. 263-274.
- 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.
- Robins, J. M., and Rotnitzky, A. (1992), "Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers," *Aids Epidemiology, Methodological Issues*, eds. N. P. Jewell, K. Dietz, and V. T. Farewell, Boston, Birkhäuser pp. 297-331.
- Rossini, A. J., and Tsiatis, A. A. (1996), "A Semiparametric Proportional Odds Regression Model for the Analysis of Current Status Data," *Journal of the American Statistical Association*, 91, 713-721.
- 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.
- 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 de Geer, S. (1994), *Asymptotic Normality in Mixture Models*, University of Leiden, the Netherlands.
- van der Laan, M. J. (1997), "Locally Efficient Estimation With Current Status Data," in *Proceedings of the Biometrics Section, American Statistical Association*, pp. 41-49.
- van der Laan, M. J., Bickel, P. J., and Jewell, N. P. (1997), "Doubly and Singly Censored Current Status Data: Estimation, Regression and Asymptotics," *Scandinavian Journal of Statistics*, 24, 289-308.
- van der Laan, M. J., and Robins, J. M. (1997), "Locally Efficient Estimation With Current Status Data and Time-Dependent Covariates," Technical Report 64, University of California, Berkeley, Division of Biostatistics.
- van der Laan, M. J., and Hubbard, A. (1997), "Estimation With Interval Censored Data and Covariates," *Lifetime Data Analysis* 3, 77-91.
- van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer-Verlag.