Locally Efficient Estimation of a Multivariate Survival Function in Longitudinal Studies

Mark J. van der Laan, Alan E. Hubbard, and James M. Robins

We consider estimation of the joint distribution of multivariate survival times $\mathbf{T} = (T_1, \ldots, T_k)$, which are subject to right-censoring by a common censoring variable C. Two estimators are proposed: an initial inverse-probability-of-censoring weighted (IPCW) estimator and a one-step estimator. Both estimators incorporate information on available time-independent and time-dependent prognostic factor (covariate) data. The IPCW estimator is consistent and asymptotically normal (CAN) under coarsening at random (CAR) and a correct specification of a model for the hazard of censoring given the past covariate and failure data. The one-step estimator is a locally efficient doubly robust estimator. That is, (i) it is CAN under the assumption of CAR and either (but not necessarily both) correct specification of a model for the hazard of censoring given the past or correct specification of a model for the conditional distribution of $\mathbf{T}$ given past failure and covariate information, and (ii) it is efficient when both these models are correctly specified. The proposed methodology does not require that the time variables $T_1, \ldots, T_k$ be ordered, although our methods cover this important special case. In particular, our estimators can be used to estimate the gap time distributions associated with an ordered series of events. The proposed methodology is an improvement over currently available approaches in a number of ways. Specifically, when censoring and failure are dependent because the hazard of censoring depends on both past failure and covariate history, our one-step estimator is the only estimator with the double-robustness property. When censoring can be assumed to be independent of the failure and covariate processes, our locally efficient one-step estimator, unlike the maximum likelihood estimator (MLE) of van der Laan but like the estimators of Dabrowska, Prentice and Cai, and Bickel, does not require smoothing and so will perform well in moderate size samples even if k is large, say 7 or 8; furthermore, unlike all previous estimators, our estimator exploits the information available in past covariate as well as failure history and so will be efficient (nearly efficient) even when the components of $\mathbf{T}$ are highly dependent, whenever the specified model for the conditional distribution of $\mathbf{T}$ given past failure and covariate information is correct (nearly correct). We examine the finite sample performance of our estimators in a simulation study. Finally, we apply our estimators to data on time to wound excision and time to wound infection in a population of burn victims.

KEY WORDS: Asymptotically efficient; Asymptotically linear estimator; Cox proportional-hazards model; Influence curve; Multivariate right-censored data.

1. INTRODUCTION

This article discusses locally efficient one-step estimation of multivariate survival functions, $S(\mathbf{t}) = P(\mathbf{T} > \mathbf{t})$, of a multivariate time variable, for example, $\mathbf{T} = (T_1, \ldots, T_k)$. For example, consider a study in which HIV-infected subjects have been randomized to several treatment groups. One might wish to compare the multivariate treatment specific survival functions of (1) time to treatment failure (as measured by viral load), (2) time to AIDS diagnosis, and (3) time until death. Wei, Lin, and Weissfeld (1989) considered estimation of multivariate survival time distribution of tumor recurrence among patients with bladder cancer (see also Andrews and Herzberg 1985). In this setting, a researcher may also wish to estimate a functional of the joint distribution, such as the distribution of the gap time, $T_2 - T_1$, between the first two recurrences (Gill, van der Laan, and Robins 1997; Wang and Wells 1998; Lin, Sun, and Ying 1999).

As another example, Klein and Moeschberger (1997) and Ichida, Wassell, and Keller (1993) considered estimation of the joint distribution of time to wound excision and time to wound infection in a population of burn victims. In Section 7, we use our methods to reanalyze these data.

Let C be the right-censoring time common to all $T_1, \ldots, T_k$, for example, the time from treatment to end of follow-up in a clinical trial or the time at which a subject drops out of a study. The estimators we discuss allow, under certain assumptions, that censoring can be informative. Each subject is observed until $\widetilde{T} = \min(T, C)$, where $T \equiv \max(T_1, \ldots, T_k)$. Let $\overline{L}(\widetilde{T}) = (L(s) : 0 \le s < \widetilde{T})$ represent a (possibly) time-dependent covariate process observed until $\widetilde{T}$. Let $\widetilde{T}_j \equiv \min(T_j, C)$, $j = 1, \ldots, k$. For each subject, the researcher observes the following data structure:

$$Y = \left(\widetilde{T}_1, \ldots, \widetilde{T}_k, \Delta_1 = (T_1 < C), \ldots, \Delta_k = (T_k \le C), \overline{L}(\widetilde{T})\right).$$

We do not require that time variables $T_1, \ldots, T_k$ be ordered, although our general procedure covers this important special case. Table 1 presents this notation as it relates to the burn victim study.

There is an alternative representation of the observed data that motivates the application of previously developed methods. Specifically, the data structure Y can be represented in terms of the process $X(\cdot) = (I(T_1 > \cdot), \ldots, I(T_k > \cdot), \overline{L}(\cdot))$ as

$$Y = \left(\widetilde{T} = C \wedge T, \Delta(\widetilde{T}) = I(\widetilde{T} = T), \overline{X}(\widetilde{T})\right),$$

where $\overline{X}(t) = (X(s) : 0 \le s < t)$.

That is, the observed data Y can be represented as the full-data process $X = \overline{X}(T)$ right censored by C. This data structure was extensively investigated by Robins and Rotnitzky (1992). Thus, we apply their methodology to the problem of multivariate right-censored data with a common censoring time C.

Mark J. van der Laan is Professor of Biostatistics, Division of Biostatistics, University of California, Berkeley, CA 94720 (E-mail: *laan@stat.berkeley.edu*). Alan E. Hubbard is Professor of Biostatistics, Division of Biostatistics, University of California, Berkeley, CA 94720 (E-mail: *hubbard@stat.berkeley.edu*). James M. Robins is Professor of Biostatistics, Department of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA 02115 (E-mail: *robins@hsph.harvard.edu*). This research was supported by a FIRST award from the National Institute of General Medical Sciences, National Institute of Health (GM53722).

Journal of the American Statistical Association
June 2002, Vol. 97, No. 458, Theory and Methods

Table 1. Data in the Burn Victim Study

Variable name	*Symbol*
Time to excision or censoring	$\tilde{T}_1$
Time to infection or censoring	$\tilde{T}_2$
Indicator excision is observed	Δ_1
Indicator infection is observed	Δ_2
Gender	$L_1(0)$
Race	$L_2(0)$
Percentage of body surface burned	$L_3(0)$
Categorical variable indicating location of burn	$L_4(0)$
Indicator head is burned	$L_5(0)$
Type of burn (chemical, scald, electrical, or flame)	$L_5(0)$

Source: Ichida et al. (1993).

Let $\mu = EB$ be the parameter of interest, where $B = b(X)$ is a function of X and EB its expectation. If, for given $\mathbf{t} \in \mathbb{R}^k$, one takes $B = I(\mathbf{T} > \mathbf{t})$, then $\mu = S(\mathbf{t})$. Likewise, if, for given $t > 0$, $B = I(T_2 - T_1 > t)$, then $\mu = P(T_2 - T_1 > t)$ is the first gap time distribution evaluated at t. This article discusses the estimation of μ based on n iid observations $Y_1, \ldots, Y_n$. The focus of Robins and Rotnitzky (1992) was on estimation of parameters related to the distribution of T, whereas we consider estimation of parameters of the distribution of X. However, the general results contained in their Theorem 4.3 allow us to apply Robins and Rotnitzky's (1992) methodology to the estimation of the mean of a functional $b(X)$. Bang and Tsiatis (2000) and Strawderman (2000) also exploited Robins and Rotnitzky's Theorem 4.3 in an analogous fashion.

The distribution of Y can be indexed by the distribution F_X of the full data X and the conditional survival function $G(c|X) \equiv P(C \geq c|X)$ of C, given X. As in Robins and Rotnitzky (1992), we will assume that the conditional distribution G of C, given X, satisfies coarsening at random (CAR). Coarsening at random was originally formulated by Heitjan and Rubin (1991) and generalized by Jacobsen and Keiding (1995) and Gill et al. (1997). For our particular data structure, we have that the censoring mechanism satisfies CAR if and only if, for $c < T$,

$$\lambda_C(c|X) = m(c, \overline{X}(c)) \quad \text{for some function } m, \qquad (1)$$

where $\lambda_C(\cdot|X)$ is the hazard of C, given X, among subjects $T > \cdot$. Note that this allows the hazard of censoring at c to be a function of the observed part (up until time c) of $(T_1, T_2, \ldots)$ and of time-dependent covariates $\overline{L}(c)$. This implies that, even in the absence of relevant covariates, $L(\cdot)$, CAR permits the censoring at time c to depend statistically on failure times observed before c. This differs from existing estimates of bivariate survival, (discussed later), which assume independent censoring. Gill et al. (1997) and Robins, Rotnitzky, and Scharfstein (1999) proved that, in the absence of further assumptions, the truth of the CAR assumption (1) cannot be empirically tested from observed data $Y_1, \ldots, Y_n$. Since (1) cannot be guaranteed to hold even approximately and is not subject to empirical test, it might be useful to investigate the sensitivity of one's inferences concerning μ to violations of (1) through a formal sensitivity analysis. Robins et al. (1999) and Scharfstein and Robins (2000) developed relevant sensitivity analysis methodology.

Identifying the mean μ of a functional $B = b(X)$ from the observed data Y requires that, for each subject, there is a positive probability of observing B. More precisely, we can refine the definition of our data (relative to the parameter of interest) by noting that the relevant information B on a parameter of interest can be observed for a subject, even if the subject is censored before all the failure times are observed. For instance, if estimating $P(\mathbf{T} > \mathbf{t})$, then all the relevant information has been observed by $\max(\mathbf{t})$. Because the information on each subject is nondecreasing with time [i.e., $\overline{X}(t)$ is a function of $\overline{X}(u)$ for $t < u$], it follows that there is an earliest time $V = v(X)$ such that $B = b(X)$ is a function of $\overline{X}(V)$. Using the preceding example, if $B = I(\mathbf{T} > \mathbf{t})$, then $V = T \wedge t^*$, where $t^* = \max(t_1, \ldots, t_k)$. Now let $\Delta \equiv I(B \text{ is observed})$ be the indicator of B being observed (formally, $\Delta = 1$ if B can be calculated from Y). It follows that $\Delta = 1$ if and only if $C > V$. Define $\Pi_G(X) = P(\Delta = 1|X)$; that is, $\Pi_G(X)$ is the conditional survivor function $G(c|X)$ evaluated at $c = V$. Thus, to ensure identification of $\mu = EB$, we assume that, with probability 1, the censoring mechanism satisfies

$$\Pi_G(X) = G(V|X) > \delta \quad F_X\text{-a.e. for some } \delta > 0. \qquad (2)$$

Under CAR, $G(V|X)$ is a function of the observed data Y for subjects with $\Delta = 1$.

Due to the curse of dimensionality, it is impossible to construct reasonable estimators that are asymptotically efficient at all laws allowed by the nonparametric CAR model that only imposes the CAR assumptions (1) and (2). In particular, the nonparametric maximum likelihood estimator of μ will often be inconsistent or even undefined at moderate sample sizes (Robins and Ritov 1997). Furthermore, Gill et al. (1997) showed that, in this model, all *regular* asymptotically linear estimators of any parameter μ of the full-data distribution F_X are asymptotically equivalent and efficient. It follows that, in the nonparametric CAR model, there exist neither efficient nor inefficient estimators of μ that will perform well at all laws allowed by the model. Thus, the only way to obtain practical estimators is to impose additional dimension-reducing modeling assumptions on either G or F_X. For example, as further reviewed later, the previous inefficient estimators of a bivariate survival function based on bivariate right-censored data assume independent censoring, which is a more stringent restriction on $G(c|X)$ than CAR (since, in contrast to CAR, the hazard of censoring at t is not allowed to depend on past failure times). Our approach will be to posit lower dimensional working models for G and for the function $Q(u) \equiv E(B|\overline{X}(u), T \geq u)$ of the full-data distribution F_X. Our inverse-probability-of-censoring weighted (IPCW) estimators will be consistent and asymptotically normal (CAN) if the model for G is correct. Our closed-form locally efficient one-step estimator will be a CAN estimator of μ provided at least one of these working models is correct. This property has been referred to as *double robustness* or, equivalently, *double protection* (Robins, Rotnitzky, and van der Laan 1999; Robins and Rotnitzky 2001; van der Laan and Zhuo 2001).

The censoring mechanism is modeled by estimating the function $m(c, \overline{X}(c))$ in (1). Though the results in this article hold for any choice of parametric or semiparametric model for $\lambda_C(\cdot|X)$, we chose to emphasize the Cox regression model

$$\lambda_C(c|X) = \lambda_0(c)\exp(\alpha^{\mathrm{T}} W(c)), \tag{3}$$

where $\lambda_0(c)$ is an unspecified baseline hazard function, α is an unknown k-dimensional vector of coefficients, and $W(c)$ is a known k-dimensional time-dependent vector function of $\overline{X}(c) = (X(s) : 0 \le s < c)$. Specifically, $W(c)$ is a vector of covariates (or functions of covariates) constructed from the relevant history of the subject observed up to time c. For instance, in the study of recurrent tumors, $W(c)$ might include the size of the last tumor and the time elapsed since the removal of this tumor. Note that, if censoring is independent of X, that is, $\lambda(c|X) = \lambda_0(c)$, then model (3) is correctly specified with $\alpha = 0$. Estimation of model (3) can be conducted with standard Cox model software. Discussion of modeling and estimation of $Q(u) = E(B|\overline{X}(u), \widetilde{T} \ge u)$ will be deferred to Section 4.

We now review previous proposals for estimation of censored multivariate survival functions under the assumption of independent censoring. Because the nonparametric maximum likelihood and self-consistency principle (Efron 1967; Turnbull 1976) do not lead to a consistent estimator of continuous right-censored multivariate survival data, most proposed estimators are explicit representations of the multivariate survival function in terms of distribution functions of the observed data (see Tsai, Leurgans, and Crowley 1986; Dabrowska 1988, 1989; Burke 1988; the Volterra estimator of P. J. Bickel in Dabrowska 1988; Prentice and Cai 1992a, b). These explicit estimators are generally inefficient, but their influence curves can be explicitly calculated and so asymptotic confidence intervals are relatively easy to compute (see Gill 1992; Gill, van der Laan, and Wellner 1995). van der Laan (1996a) showed that a modified nonparametric maximum likelihood estimator (NPMLE) of the bivariate survival function (without covariates), which requires discretization of the data, is asymptotically efficient. The preceding methods allow but do not require that all failure times be censored by a common variable C. Gill et al. (1997), Wang and Wells (1998), and Lin et al. (1999) assumed both that the failure times are ordered, that is, $T_1 < \cdots < T_k$ with probability 1, and that, as in the present article, all are right censored by the same censoring variable C. In Section 3, we demonstrate that the estimator of Lin et al. (1999) is an IPCW estimator as proposed by Robins and Rotnitzky (1992) and defined for general CAR-censored data models in Gill et al. (1997). Not only are the preceding explicit estimators inefficient, but, even when CAR holds, they will be inconsistent when censoring and failure are dependent [i.e., $\alpha \ne 0$ in (3)], because the hazard of censoring depends on past failure or prognostic factor history.

1.1 Organization of the Article

In Section 2, we consider the special case in which $T_1 < \cdots < T_k$ with probability 1, no covariates are available, and there is independent censoring by a common censoring variable C. We prove that the NPMLE is inconsistent and that repairing it will involve multivariate smoothing so that the practical performance of the NPMLE will be reasonable only at large sample sizes and relatively small dimensions of $\mathbf{T}$. The discussion of the NPMLE motivates the need for an alternative estimator. In Section 3, we study an IPCW estimator. In Section 4, we introduce a locally efficient one-step estimator and derive an easy-to-compute confidence interval for this estimator. Section 5 provides a methodology for estimation of the conditional expectation $Q(u) = E(B|\overline{X}(u), \widetilde{T} \ge u)$. Results of a simulation study and data analysis are discussed in Sections 6 and 7; proofs of our main results are deferred to the Appendix.

2. INCONSISTENCY OF THE NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATOR

Consider the special case in which C and T are independent, no covariates are available, and $T_1 < \cdots < T_k$ with probability 1 as considered by Gill et al. (1997), Lin et al. (1999), and Wang and Wells (1998). In the case of ordered time variables, it is useful to rewrite the data structure:

$$Y = \left(\widetilde{T}_1 = T_1 \wedge C, \ldots, \widetilde{T}_k = T_k \wedge C,\ \xi \equiv \sum_{j=1}^{k} I(C > T_j)\right).$$

It is straightforward to show that the likelihood of one observation Y is given by

$$\begin{aligned} &L(F|\widetilde{T}_1, \ldots, \widetilde{T}_k, \xi) \\ &\quad = S_1(\widetilde{T}_1)^{I(\xi=0)}\, dF_1(\widetilde{T}_1)^{I(\xi>0)} \\ &\qquad \times S_{2|1}(\widetilde{T}_2|\widetilde{T}_1)^{I(\xi=1)}\, dF_{2|1}(\widetilde{T}_2|\widetilde{T}_1)^{I(\xi>1)} \cdots \\ &\qquad \times S_{k|1,\ldots,k-1}(\widetilde{T}_k|\widetilde{T}_1, \ldots, \widetilde{T}_{k-1})^{I(\xi=k-1)} \\ &\qquad \times dF_{k|1,\ldots,k-1}(\widetilde{T}_k|\widetilde{T}_1, \ldots, \widetilde{T}_{k-1})^{I(\xi=k)}, \end{aligned}$$

where $S_{j|1,\ldots,j-1}$ and $F_{j|1,\ldots,j-1}$ denote the conditional survival and distribution functions of T_j, given $T_1, \ldots, T_{j-1}$, $j = 1, \ldots, k$. Thus, the likelihood for n iid observations is factorized in separate likelihoods for F_1, $F_{2|1}(\cdot|\widetilde{T}_1), \ldots, F_{k|1,\ldots,k-1}(\cdot|\widetilde{T}_1, \ldots, \widetilde{T}_{k-1})$. The likelihood for $F_{j|1,\ldots,j-1}(\cdot|\widetilde{T}_1, \ldots, \widetilde{T}_{j-1})$ is identical to the likelihood for univariate right-censored data on T_j restricted to the subsample for which $(T_1, \ldots, T_{j-1})$ is observed and $(T_1, \ldots, T_{j-1}) = (\widetilde{T}_1, \ldots, \widetilde{T}_{j-1})$. If the distribution of $\mathbf{T}$ were discrete and $P(\xi = k|\mathbf{T}) > 0$ F_X-a.e., then the subsample would consist of several observations and the nonparametric maximum likelihood estimator of $F_{j|1,\ldots,j-1}(\cdot|(\widetilde{T}_1, \ldots, \widetilde{T}_{j-1}))$ would be the Kaplan–Meier estimator based on this subsample. However, if $\mathbf{T}$ is continuous, then each of these subsamples has only one observation so that the NPMLE of $F_{j|1,\ldots,j-1}(\cdot|\widetilde{T}_1, \ldots, \widetilde{T}_{j-1})$ is the Kaplan–Meier estimator based on a single observation. Since the Kaplan–Meier estimator based on a single observation fails to converge in probability, the unmodified NPMLE is an inconsistent estimator of a continuous multivariate distribution F.

The obvious modification of this NPMLE is to estimate $F_{j|1,\ldots,j-1}(\cdot|\widetilde{T}_1, \ldots, \widetilde{T}_{j-1})$ with the Kaplan–Meier estimator based on the subsample for which $(T_1, \ldots, T_{j-1})$ is observed and $(T_1, \ldots, T_{j-1})$ is "close" to $(\widetilde{T}_1, \ldots, \widetilde{T}_{j-1})$, $j = 1, \ldots, k$. This might still be a reasonable estimator for $k = 2$, being asymptotically equivalent to the modified NPMLE of van der Laan (1996a), which also uses smoothing. As k becomes larger, however, one will need to smooth over large k-dimensional volumes to ensure that the subsample used by the Kaplan–Meier estimator consists of a reasonable number of observations. Smoothing over large volumes in order to reduce the finite sample variance results in a severely biased estimator. In other words, the finite sample performance of such a smoothed NPMLE will only be reasonable for huge sample sizes. Finally, because of the even greater sample partitioning required, the smoothed NPMLE for the data structure that also includes a covariate process $L(t)$ suffers even more dramatically from the curse of dimensionality. Thus, there is a need for estimators that are efficient at user-supplied lower dimensional working models but remain CAN under the sole assumption of independent censoring. Later we will discuss how to construct such *locally efficient* estimators.

3. INVERSE PROBABILITY OF CENSORING WEIGHTED ESTIMATORS

In this section, we discuss a general IPCW estimator of $\mu = EB$ for CAR models subject to right-censoring first proposed by Robins and Rotnitsky (1992). This estimator will function as an initial estimator for our locally efficient one-step estimator discussed later. The motivation for the estimator comes from the fact that, under (2),

$$E\left\{\frac{\Delta B}{G(V|X)}\right\} = E(B) = \mu, \tag{4}$$

where again $\Delta = I(B$ is observed) and V is the earliest time at which B is observed. Note that this identity follows directly from

$$E(\Delta|X) = G(V|X).$$

The identity (4) suggests the following ad hoc estimator

$$\mu_n^0 = \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i B_i}{G_n(V_i|X_i)}, \tag{5}$$

where G_n is an estimator of G, assuming the given model (3). By the coarsening at random assumption (1), $G(V|X) = \exp(-\int_0^V \lambda_C(u|X)\,du)$ is only a function of $Y = (\widetilde{T} = C \wedge T, \Delta = I(T \leq C), \overline{X}(\widetilde{T}))$ so that our estimator μ_n^0 indeed only depends on $Y_1, \ldots, Y_n$. If one assumes the Cox regression model (3), then one can use standard software to obtain the maximum (partial) likelihood estimator of the regression coefficients α and of the baseline hazard.

If we take $B = I(\mathbf{T} > \mathbf{t})$, then $\mu = S(\mathbf{t})$ and

$$\mu_n^0 \equiv S_n^0(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n}\frac{I(\mathbf{T}_i > \mathbf{t})\Delta_i}{G_n(T_i \wedge t|X_i)}.$$

If T_1 and T_2 are ordered ($T_1 < T_2$ with probability 1) and $B = I(T_2 - T_1 > t)$, then $\Delta = I(C > T_1 + t)$ and $V = T_1 + t$. Furthermore, $\mu = P(T_2 - T_1 > t)$. Thus, $\mu_n^0 = (1/n)\sum_i I(T_{2i} - T_{1i} > t)\Delta_i / G_n(T_{1i} + t|X_i)$.

If one is willing to assume that C is completely independent of X [i.e., $\alpha = 0$ in (3)], then one can consistently estimate G with the Kaplan–Meier estimator, G_n^{KM}, for censoring based on the n observations of $(\widetilde{T} = C \wedge T, I(C < T))$, where T now plays the role of the censoring variable for C. The estimator μ_n^0 of $\mu = P(T_2 - T_1 > t)$ with G_n^{KM} substituted for G_n is precisely the estimator of $\mu = P(T_2 - T_1 > t)$ proposed by Lin et al. (1999). Our general asymptotics theorem in the Appendix proves that, under specified regularity conditions, μ_n^0 with G_n estimated according to a submodel of CAR is asymptotically linear with influence curve

$$\text{IC}_0'(Y|G,\mu) \equiv \text{IC}_0 - \Pi(\text{IC}_0|T_G), \tag{6}$$

and asymptotic variance $\text{VAR}(\text{IC}_0) - \text{VAR}(\Pi(\text{IC}_0|T_G))$, where IC_0 is the influence curve of μ_n^0 when $G_n = G$ is known and $\Pi(\cdot|T_G)$ is the projection operator onto the tangent space generated by G under the assumed submodel of CAR. Since the tangent space T_G for model (3) with α regarded as unknown contains the space T_G for the "independent censoring" submodel of (3) that sets α to 0, it follows that the asymptotic variance of the estimator of Lin et al. is never less than and almost always greater than that of our IPCW estimator μ_n^0. If the assumption of independent censoring is false, the estimator of Lin et al. is inconsistent, whereas (5) will be consistent under CAR if (3) is correct.

The estimated variance of (5) can be calculated two ways: by bootstrapping or by consistently estimating the projection, $\Pi(\text{IC}_0|T_G)$, and thus the influence curve of the IPCW estimator. If G is estimated using the Cox regression (3), then the projection is (van der Laan and Robins 2001)

$$\begin{aligned}\Pi(\text{IC}_0|T_G) &= E(\text{IC}_0 S_\alpha^{\mathsf{T}})E(S_\alpha S_\alpha^{\mathsf{T}})^{-1}S_\alpha \\ &\quad - \int \frac{E\{B\exp(\alpha W(u))\Delta/G(T|X)\}}{E\{I(C>u)\exp(\alpha W(u))\}}\,dM(u),\end{aligned} \tag{7}$$

where

$$\text{IC}_0(Y|G,\mu) \equiv \frac{B\Delta}{G(V|X)} - \mu, \tag{8}$$

S_α is the efficient score of the Cox regression coefficients, α,

$$S_\alpha = \int\left\{W(u) - \frac{E\{W(u)I(C>u)\exp(\alpha W(u))\}}{E\{I(C>u)\exp(\alpha W(u))\}}\right\}dM(u),$$

and

$$dM(u) \equiv I(C \in du, \Delta = 0) - \Lambda_C(du|X)I(\widetilde{T} > u) \tag{9}$$

is the martingale $dA(u) - E(dA(u)|\mathcal{F}(u))$ of the counting process $A(u) = I(C \leq u, \Delta = 0)$ w.r.t. the history $\mathcal{F}(u) = (\overline{X}(u), \overline{A}(u))$ and $\Lambda_C(u|X)$ is the cumulative hazard of censoring at u, given the full data X. The influence curve (6) can be estimated by plugging in estimates of the censoring distribution, EB, and the required expectations. Because with Cox regression the estimated censoring distribution only has mass at the observed censoring times, the integrals in (7) reduce to simple sums. The estimated variance of $\sqrt{n}(\mu_n^0 - \mu)$ is then

$$\frac{1}{n}\sum_i \text{IC}_0'(Y_i|G_n, \mu_n^0)^2.$$

Note that the estimate of the variance of μ_n^0 can easily be generalized to estimating the variance-covariance of a vector of $\mu_n^0 = \mu_{n1}^0, \mu_{n2}^0, \ldots, \mu_{nk}^0$, corresponding to a vector of parameters, $\mathbf{EB} = EB_1, EB_2, \ldots, EB_k$. Specifically, if $\mathbf{IC}'_0 \equiv \mathrm{IC}'_0(Y_i|G_n, \mu^0_{n1}), \ldots, \mathrm{IC}'_0(Y_i|G_n, \mu^0_{nk})$ is the vector of the k influence curve estimates for the ith subject, then the estimated variance-covariance of μ^0_n is simply the method-of-moments estimator of $E[\{\mathbf{IC}'_0\}^T\mathbf{IC}'_0]$. This general technique for estimating the variance-covariance matrix of a vector of estimates can also be applied to the locally efficient estimators discussed next.

4. LOCALLY EFFICIENT ONE-STEP ESTIMATOR

In this section, we construct a locally efficient one-step estimator of μ by adding to the estimator μ^0_n in (5) an estimate of the empirical mean of the estimated efficient influence function in the semiparametric model characterized by the restrictions (1), (2), and (3). Our first task is to provide a representation of the efficient influence function at a given data-generating distribution (F_X, G), which will then be estimated by substitution of estimators of the unknown components of F_X and G. This representation has two pieces. The first is given by the influence function of μ^0_n when using the known G, which is given by (8).

The second piece is the projection $\mathrm{IC}^*_{\mathrm{nu}}$ in $L^2_0(P_{F_X,G})$ of IC_0 onto the nuisance tangent space of G only, assuming CAR [i.e., (1)], which we will denote by T_{CAR}, where

$$T_{\mathrm{CAR}}(G) = \left\{\int H(u, \overline{X}(u))\, dM(u) : H\right\}.$$

By CAR, $\mathrm{IC}^*_{\mathrm{nu}}$ is a function of Y. Robins and Rotnitzky (1992) showed that $\mathrm{IC}^*_{\mathrm{nu}}$ is given by

$$\mathrm{IC}^*_{\mathrm{nu}}(Y|F_X, G) = \mathrm{IC}^*_{\mathrm{nu}}(Y|Q, G) = -\int Q(u)\frac{dM(u)}{G(u|X)}, \quad (10)$$

where $Q(u) = E(B|\overline{X}(u), \widetilde{T} \geq u)$ is the conditional expectation of B, given $\overline{X}(u) = (X(s) : 0 \leq s < u)$ and $\widetilde{T} \geq u$. [This formula follows from the fact that, for any $H(X, C), \Pi(H(X, C)|T_{\mathrm{CAR}}) = \int\{E(H(X, C)|\overline{X}(u), C = u) - E(H(X, C)|\overline{X}(u), C \geq u)\}\, dM(u)$.] Note that $\mathrm{IC}^*_{\mathrm{nu}}(\cdot|F_X, G)$ only depends on F_X through $Q(u)$. Thus, if $B = I(\mathbf{T} > \mathbf{t})$, then

$$\mathrm{IC}^*_{\mathrm{nu}}(Y|Q, G) = -\int S(\mathbf{t}|\overline{X}(u), \widetilde{T} \geq u)\frac{dM(u)}{G(u|X)}. \quad (11)$$

In addition, for $u > t^* \equiv \max(t_1, \ldots, t_k)$, $Q(u) = S(\mathbf{t}|\overline{X}(u), \widetilde{T} > u)$ is a known function of $\overline{X}(u)$ and equal to either 1 or 0.

From results in the appendix of Robins and Rotnitzky (1992) (see also Hubbard, van der Laan, and Robins 1999), it follows that the efficient influence curve IC^* at (F_X, G) for estimation of μ is given by

$$\mathrm{IC}^*(Y|Q, G, \mu) = \mathrm{IC}_0(Y|G, \mu) - \mathrm{IC}^*_{\mathrm{nu}}(Y|Q, G). \quad (12)$$

A one-step estimator is obtained by estimating IC^* as follows:

$$\mathrm{IC}^*(Y|Q_n, G_n, \mu^0_n) = \mathrm{IC}_0(Y|G_n, \mu^0_n) - \mathrm{IC}^*_{\mathrm{nu}}(Y|Q_n, G_n), \quad (13)$$

where μ^0_n is the IPCW estimator defined in (5), G_n is the Cox estimator of G based on (3), and Q_n is the estimator of Q described in the next section.

The locally efficient one-step estimator is

$$\mu^1_n = \mu^0_n + \frac{1}{n}\sum_{i=1}^{n}\mathrm{IC}^*(Y_i|Q_n, G_n, \mu^0_n). \quad (14)$$

For example, setting $B = I(\mathbf{T} > \mathbf{t})$ in (14) yields the one-step estimator $S^1_n(\mathbf{t})$ of $S(\mathbf{t})$.

Note that $\sum_i \mathrm{IC}_0(Y_i|G_n, \mu^0_n) = 0$ and therefore

$$\mu^1_n = \frac{1}{n}\sum_{i=1}^{n}\frac{B_i\Delta_i}{G_n(V_i|X_i)} - \mathrm{IC}^*_{\mathrm{nu}}(Y_i|Q_n, G_n).$$

We chose the representation (14) in order to reflect that μ^1_n is just the classical one-step estimator as defined in Bickel, Klaassen, Ritov, and Wellner (1993, p. 395); that is, by its definition, μ^1_n is the first step in the Newton–Raphson algorithm for solving the optimal estimating equation (corresponding to the efficient influence curve)

$$0 = \frac{1}{n}\sum_{i=1}^{n}\mathrm{IC}^*(Y_i|Q_n, G_n, \mu) \quad (15)$$

for μ, with nuisance parameters Q and G, and where we chose μ^0_n as the initial estimator. This follows from the fact that the derivative of the estimating equation (15) with respect to μ equals -1. In fact, in our special case the estimating equation is linear in μ so that μ^1_n is also the exact solution of (15).

Double-Protection Property. We will view $\mathrm{IC}^*(Y|Q, G, \mu)$ as an estimating function in μ with nuisance parameters G and $Q = Q(F_X)$. This estimating function satisfies the following unbiasedness property:

$$E_{F_X,G}\mathrm{IC}^*(Y|Q_1, G_1, \mu) = 0 \quad \text{if } G_1 = G \text{ or } Q_1 = Q(F_X). \quad (16)$$

This double-protection property can be proved directly but it holds, in general, in CAR-censored data models (Robins et al. 1999; van der Laan and Zhuo 2001), and it is the basis of the "doubly robust" asymptotic properties of our one-step estimator. That is, $\mathrm{IC}^*(Y|Q_1, G_1, \mu)$ has mean 0 if either (but not necessarily both) Q_1 or G_1 is equal to the truth.

Efficiency Considerations. To understand the potential increase in efficiency provided by μ^1_n, consider the extreme case in which the observed past is a perfect predictor of B in the sense that $Q(u) = E(B|\overline{X}(u), T > u) = B$ with probability 1 for all u. Then μ^1_n using the true Q reduces to the empirical distribution of the complete data, $\mu^1_n = (1/n)\sum_i B_i$, and thus has succeeded in recovering all information lost due to censoring. A heuristic explanation of how μ^1_n manages to recover this information is as follows. Consider two subjects censored at time c, $c < V$, so that B is not observed. Suppose, based on evidence obtained from the uncensored subjects concerning the unknown prediction function $Q(u)$, subject 1 has an observed past medical history that predicts a small value of B, while subject 2's past predicts a large value of B. An efficient estimator such as μ^1_n will use the estimated $Q(u)$ to predict each subject's B and thus will use the two subjects quite differently.

Based on the simulations in Section 6 and the simulations in Hubbard et al. (1999), μ_n^1 can significantly improve on the IPCW estimator that ignores the covariates and just uses the Kaplan–Meier estimate for G, even when censoring is uninformative.

4.1 Construction of Confidence Intervals

We first consider the case in which we are willing to assume that the model (3) for the censoring mechanism is correct. In Theorem A.1, we show that μ_n^1 of (14) is asymptotically linear with an influence curve $\mathrm{IC}^*(Y|Q_1, G, \mu) - \Pi(\mathrm{IC}^*(\cdot|Q_1, G, \mu)|T_G)$, where $\mathrm{IC}^*(Y|Q_1, G, \mu)$ is the limit as $n \to \infty$ of $\widehat{\mathrm{IC}}(Y) = \mathrm{IC}^*(Y|Q_n, G_n, \mu_n^0)$ and $\Pi(\cdot|T_G)$ denotes the projection onto the tangent space of G under the assumed Cox proportional-hazards model. Thus, the variance of this influence curve is smaller than or equal to the variance of $\mathrm{IC}_1^*(Y) \equiv \mathrm{IC}^*(Y|Q_1, G, \mu)$. Therefore, a conservative estimate of the asymptotic variance of μ_n^1 is given by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\{\widehat{\mathrm{IC}}(Y_i)\}^2.$$

This can be used to construct a conservative 95% confidence interval for μ,

$$\mu_n^1 \pm 1.96\frac{\hat{\sigma}}{\sqrt{n}}. \tag{17}$$

This confidence interval has asymptotic level 95% if $Q_n(u)$ is consistent for $Q(u)$ [i.e., $Q_1(u) = Q(u)$] since $\Pi(\mathrm{IC}^*(\cdot|Q, G, \mu)|T_G) = 0$. In addition, one gets the confidence interval for "free" after computing μ_n^1. By calculating the projection onto T_G, as in Section 3, Robins (1996) and van der Laan and Robins (2001) provided nonconservative asymptotic interval estimators. However, these interval estimators are unlikely to be worth the effort since, unless the model for $Q(u)$ is very poorly specified, the conservative interval will have actual coverage rate that is only slightly greater than the nominal (see the simulation study of Hubbard et al. 1999).

Theorem A.1 and its proof demonstrate that μ_n^1 remains asymptotically linear if either the censoring mechanism G or the model for $Q(u)$ is estimated consistently. However, when the model for G is misspecified and that for $Q(u)$ is correct, the interval estimator (17) is no longer guaranteed to be conservative. Therefore, to avoid the technical difficulty of calculating the asymptotic variance of μ_n^1, we recommend estimating the asymptotic variance with the nonparametric bootstrap (Gill 1989). In this case, the bootstrap works (i.e., the estimators are regular and asymptotically normal) if either the censoring distribution or $Q(u)$ is consistently estimated.

4.2 Guaranteed Improvement of the One-Step Estimator Relative to the IPCW Estimator

If the model for $Q(u)$ is badly misspecified, then the one-step estimator μ_n^1 can be less efficient than the initial estimator μ_n^0, even asymptotically. For the interested reader, we provide here a locally efficient one-step estimator μ_n^* which asymptotically always improves on the initial IPCW estimator μ_n^0 (Robins and Rotnitzky 1992; Robins 1993; Hubbard et al. 1999; van der Laan and Robins 2001). In this case, one must assume that the model for G is correct and thus the estimator will no longer benefit from the double-protection property.

Under regularity conditions, μ_n^0 is consistent and asymptotically linear with influence curve $\mathrm{IC}_0' = \mathrm{IC}_0 - \Pi(\mathrm{IC}_0|T_G)$ as in (6), where T_G is the tangent space of the Cox regression model (3) assumed for G and the closed-form representation of the projection $\mathrm{IC}_{\mathrm{nu}} \equiv \Pi(\mathrm{IC}_0|T_G)$ onto T_G is given by (7). Let $\mathrm{IC}_{\mathrm{CAR}}(Y) \equiv \Pi(\mathrm{IC}_0'(Y)|T_{\mathrm{CAR}})$ be the projection of IC_0' onto the tangent space of G when only assuming CAR. Since $\mathrm{IC}_{\mathrm{nu}} \in T_{\mathrm{CAR}}$ and $\Pi(\mathrm{IC}_0|T_{\mathrm{CAR}}) = \mathrm{IC}_{\mathrm{nu}}^*$ provided above (11), this projection is given by

$$\mathrm{IC}_{\mathrm{CAR}}(Y) = \mathrm{IC}_{\mathrm{nu}}^*(Y) - \mathrm{IC}_{\mathrm{nu}}(Y).$$

Let $\widehat{\mathrm{IC}}_0'(Y|\mu)$ and $\widehat{\mathrm{IC}}_{\mathrm{CAR}}$ be estimates of IC_0' and $\mathrm{IC}_{\mathrm{CAR}}$, where we assume that $\widehat{\mathrm{IC}}_0'$ consistently estimates IC_0' while $\widehat{\mathrm{IC}}_{\mathrm{nu}}^*$ and thus $\widehat{\mathrm{IC}}_{\mathrm{CAR}}$ are allowed to be inconsistent because of misspecification of the model for $Q(u)$. Define $c_n = P_n\{\widehat{\mathrm{IC}}_0'\widehat{\mathrm{IC}}_{\mathrm{CAR}}\}/P_n\widehat{\mathrm{IC}}_{\mathrm{CAR}}^2$, where, given a function $f(Y)$, $P_n f \equiv (1/n)\sum_i f(Y_i)$ denotes the empirical mean of $f(Y)$. Note that $c_n\widehat{\mathrm{IC}}_{\mathrm{CAR}}$ estimates consistently the projection of $\widehat{\mathrm{IC}}_0'$ onto the one-dimensional space $\langle\widehat{\mathrm{IC}}_{\mathrm{CAR}}\rangle$ spanned by $\widehat{\mathrm{IC}}_{\mathrm{CAR}}$. Now define the one-step estimator:

$$\mu_n^* = \mu_n^0 + \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathrm{IC}}_0'(Y_i|\mu_n^0) - c_n\widehat{\mathrm{IC}}_{\mathrm{CAR}}(Y_i). \tag{18}$$

Application of our general asymptotics theorem shows that, under specified regularity conditions, this estimator will be asymptotically linear with influence curve $\mathrm{IC}_0' - \Pi(\mathrm{IC}_0'|\langle\mathrm{IC}_{\mathrm{CAR},1}\rangle)$, where $\mathrm{IC}_{\mathrm{CAR},1}$ represents the asymptotic limit of $\widehat{\mathrm{IC}}_{\mathrm{CAR}}$. Thus, the variance of this influence curve is smaller than the variance of IC_0', which demonstrates that μ_n^* is asymptotically more efficient than μ_n^0 even when $\mathrm{IC}_{\mathrm{CAR},1}$ is not close to $\mathrm{IC}_{\mathrm{CAR}}$. Note also that when the model for $Q(u)$ is consistently estimated, $c_n \to 1$ and thus μ_n^* is locally efficient. This leads to a model selection criterion for the estimation of $Q(u)$, specifically the distance of c_n from 1. We demonstrate in our data analysis (Section 7) that such a criterion can be useful in choosing an estimate of $Q(u)$ among competing models.

5. ESTIMATION OF THE FULL-DATA NUISANCE PARAMETER

In this section, we provide methods for estimation of $Q(u)$. In Section 5.1, we use inverse weighting (by G) to estimate Q. This means that, if the model for G is misspecified, then G_n and Q_n will be inconsistent so that our asymptotics fully rely on a correctly specified model for G. Thus, by using the method detailed in Section 5.1, the estimator no longer has the double-protection property. However, for time-dependent covariates, this method yields practical estimators. In Section 5.2, we provide likelihood-based methods assuming full-data models for F_X which preserve the double-protection property (16).

5.1 A Regression Method for Estimation of the Conditional Expectation

To evaluate (14), we need an estimate of $Q(u)$ at and only at those times u at which a subject has been censored (assuming a Cox model for censoring). We will estimate this conditional expectation $Q(u) = E(B|\overline{X}(u), \widetilde{T} \geq u)$ by using a general regression approach with covariates extracted from $\overline{X}(u)$, based on the subsample with $\widetilde{T} \geq u$. Note that, under (1),

$$E(B|\overline{X}(u), \widetilde{T} \geq u) = E(O_G|\overline{X}(u), \widetilde{T} \geq u),$$

where

$$O_G \equiv \frac{B\Delta G(u|X)}{G(V|X)}. \tag{19}$$

Given an estimate G_n of G, $O_{G_n}(u)$ is an observed random variable. The idea of representing the conditional probability $Q(u)$ as a regression of an observable random variable $O_{G_n}(u)$ on observed covariates is due to Robins (1993) and Robins and Rotnitzky (1992). Consequently, for every u equal to an observed C_i, we perform a parametric or nonparametric regression estimation of $O_{G_n}(u)$ on one or more relevant for B user-supplied summary measures $W_1(u), \ldots, W_m(u)$ of $\overline{X}(u)$, restricted to subjects with $\widetilde{T} \geq u$. For example, one could use the SPLUS function GAM (Hastie and Tibshirani 1990) to fit, separately at each censoring time u, a generalized additive logistic regression of O_{G_n} on $W_1(u), \ldots, W_m(u)$,

$$E(O_{G_n}|W_1(u), \ldots, W_m(u)) = \frac{\exp(f_1(W_1(u)) + \cdots + f_m(W_m(u)))}{1 + \exp(f_1(W_1(u)) + \cdots + f_m(W_m(u)))}, \tag{20}$$

where the f_i are unknown functions and GAM allows the user to specify the number of degrees of freedom used to fit $f_i, i = 1, \ldots, m$. We propose to smooth the V-specific regression estimates w.r.t. u.

When (3) holds, the one-step estimator will be CAN and will be highly efficient and perform well in moderate samples even when (i) $\mathbf{T}$ has many components k and (ii) the GAM regression models corresponding to different censoring times u are incompatible with any joint distribution of X.

For the purpose of selecting summary measure, $W_j(u), j = 1, \ldots, m$, it is useful to note that $(\overline{X}(u), \widetilde{T} \geq u)$ is equivalent to $\Delta(u) = (\Delta_1(u), \ldots, \Delta_k(u), \Delta_1(u)T_1, \ldots, \Delta_k(u)T_k, \overline{L}(u))$, where $\Delta_j(u) \equiv I(T_j \leq u), j = 1, \ldots, m$.

5.2 Estimation of the Conditional Expectation by Assuming a (Semi)parametric Full-Data Model

We now describe an alternative approach that preserves double robustness by (i) specifying a (semi)parametric model for the joint law of $(\mathbf{T}, \overline{L}(T))$, (ii) estimating the model parameters by maximum likelihood, and (iii) using the MLE of $Q(u)$ in μ_n^1.

To implement this technique with time-dependent covariates, it is convenient to represent the joint density of $(\mathbf{T}, \overline{L}(T))$ as a product integral over time of conditional densities of the current events, given the past events, where each of these conditional densities can be further split into conditional densities of survival time events and covariate events. The observed data likelihood will now consist of this product integral up until the minimum of censoring and T. Each of these conditional densities can be modeled with standard models such as multiplicative intensity models for the failure time events. Our one-step estimator will be CAN if either the model (3) or the model for $(\mathbf{T}, \overline{L}(T))$ is correctly specified. The only difficulty with this approach is that, if L is time dependent, then the complexity of the model can become cumbersome. Although burdensome, we believe the effort may be worthwhile for the following reason. As a practical matter, $u_{n.}^1$ will always be somewhat biased, because both the censoring model (3) and the model for the joint law of $(\mathbf{T}, \overline{L}(T))$ will inevitably be somewhat misspecified. However, we might expect that, if both models are reasonably close (say as measured by the minimum Kullback–Liebler distance to the truth), the bias in $u_{n.}^1$ will be less than that of $u_{n.}^0$. In future work, we plan to compare by simulation the biases of $u_{n.}^1$ and $u_{n.}^0$ under misspecification

Next, consider the case where L is time independent. For the sake of presentation, let $k = 2$ and $B = I(\mathbf{T} > \mathbf{t})$. Then $Q(u) = E(B|\overline{X}(u), \widetilde{T} > u) = S(t_1, t_2|\overline{X}(u), \widetilde{T} > u)$, where

$$S(\mathbf{t}|\overline{X}(u), \widetilde{T} > u) = \begin{cases} I(T_1 > t_1)\dfrac{P(T_2 > t_2 \vee u|T_1, L)}{P(T_2 > u|T_1, L)} & \text{if } \Delta(u) = (1, 0), \\ I(T_2 > t_2)\dfrac{P(T_1 > t_1 \vee u|T_2, L)}{P(T_1 > u|T_2, L)} & \text{if } \Delta(u) = (0, 1), \\ \dfrac{P(T_1 > t_1 \vee u, T_2 > t_2 \vee u|L)}{P(T_1 > u, T_2 > u|L)} & \text{if } \Delta(u) = (0, 0). \end{cases}$$

Thus, estimation of $Q(u)$ only requires an estimate of the bivariate conditional distribution of (T_1, T_2), given L. One could model this conditional survival function with the bivariate survival model based on the Copula family (Genest and MacKay 1986; Genest and Rivest 1993; Joe 1993). Frailty models (Clayton 1978; Clayton and Cuzick 1985; Hougaard 1986, 1987; Clayton 1991; Klein 1992; Costigan and Klein 1993) assume that the two components of $\mathbf{T}$ are conditionally independent, given a subset of covariates W contained in L and an unobserved frailty Z, where the frailty distribution is known to have mean 1 and variance σ. The bivariate distributions generated by frailty models are a subclass of the (Archimedean) Copula family (Oakes 1989). A particular Copula family, corresponding to a gamma frailty, is given by

$$S(t_1, t_2|W) = (S_1(t_1|W)^{-\sigma} + S_2(t_2|W)^{-\sigma} - 1)^{-1/\sigma}, \tag{21}$$

where S_1, S_2 denote the conditional marginal survival functions. Under this assumption, the conditional distribution of $\mathbf{T}$, given W, is parametrized by σ and the univariate conditional distributions of T_j, given the frailty Z and $W, j = 1, 2$. It is commonly assumed that these distributions follow the proportional-hazards models

$$\lambda_{T_j}(t|W, Z) = Z\lambda_{0,j}(t)\exp(\beta_j^T W), \qquad j = 1, 2.$$

This model can be fit with the SPLUS function *coxph* using a stratum variable indicating for each line in the data file which of the two failure time components it represents and the gamma frailty option.

For the burn victim data (Table 1), we used all the time-independent baseline covariates as regressors and allowed the coefficients of the Cox model to differ according to the time variable (excision versus infection). Only two baseline covariates were found to be significantly related to either of the failure times: treatment and gender. The SPLUS syntax for the final model was

$$coxph(Surv(t, d) \sim strata(time) + L1 + L2 + frailty(id)), \quad (22)$$

where t and d are the failure and censoring indicators, respectively. *time* indicates whether the observation refers to excision ($time = 1$) or infection ($time = 2$), *id* indicates the subject, and $L1 = L_1(0) = 0$ (bathing) or 1 (body cleansing) and $L2 = L_2(0) = 0$ (male) or 1 (female), that is, $W = (L_1(0), L_2(0))$. The estimation procedure returns an estimate of the baseline survival for each failure time, the coefficients associated with the covariates, and the variance, σ, of the gamma frailty model. These, in turn, can be used to estimate (21) and thus $Q(u)$.

With smaller sample sizes, one might simply choose a parametric survival regression model and assume independence of the time variables, conditional on the covariates. For example, in our analysis of the burn victim data, we also report results based on assuming that the conditional log hazard of both T_1 and T_2 is a linear function of the same two covariates discussed previously and fit two exponential models using the *survReg* procedure in SPLUS:

$$survReg(Surv(t1, d1) \sim L1 + L2, dist = \text{"exponential"}),$$
$$survReg(Surv(t2, d2) \sim L1 + L2, dist = \text{"exponential"}).$$

Now, $Q(u)$ will be a reasonably simple function of u, $\bar{X}(u)$, and the coefficients returned by these regression procedures.

Finally, when no covariate process $\bar{L}(\tilde{T})$ has been recorded for data analysis, the following nonparametric method is available. By CAR, we have

$$\begin{aligned} P(T_1 > t_1 | T_2) &= P(T_1 > t_1 | T_2, C > T_2) \\ &= P(T_1 > t_1 | \tilde{T}_2, \Delta_2 = 1). \end{aligned}$$

So we can nonparametrically estimate this conditional distribution at $T_2 = u_2$ with the Kaplan–Meier estimator based on the observations with an observed T_2 close to u_2. Similarly, we can nonparametrically estimate $P(T_2 < t_2 | T_1)$. However, due to the curse of dimensionality, this method should not be used when the sample size is moderate and the dimension k of $\mathbf{T}$ is larger than 2 or so.

6. SIMULATION RESULTS

A simulation study was performed to examine the relative performance of the NPMLE, the IPCW (μ_n^0), and the locally efficient one-step estimators (μ_n^1). In all simulations, the failure times are bivariate, censoring is independent of the failure times, and the Cox regression model with all available covariates included as regressors is used to estimate the censoring distribution for both μ_n^0 and μ_n^1. We used two methods to estimate $Q(u)$ (used in the one-step estimator): (1) the known values derived analytically from the data-generating distributions (subsequently referred to as μ_n^1) and (2) linear regression of O_G against time-independent covariates, W, as described in Section 4.1 (subsequently referred to as $\mu_{n,\text{REG}}^1$). We want to compare the performance of the one-step estimators under different models for $Q(u)$. Using the true function $Q(u)$ in μ_n^1 is essentially equivalent to a low-dimensional correctly specified parametric model. Thus, in our simulations, μ_n^1 will be fully efficient. Under the distributions chosen for our simulations, however, the true regression $Q(u)$ of O_G on the aforementioned covariates is nonlinear and thus $\mu_{n,\text{REG}}^1$ will be somewhat inefficient owing to misspecification of the model for $Q(u)$. The ratio of the mean-squared errors (MSE's), based on the 1,000 trials of each simulation, is used to compare the efficiency of the competing estimators. In addition, we report the percentage of iterations in which the conservative 95% confidence interval [based on (17)] includes the true value for both the μ_n^1 and the $\mu_{n,\text{REG}}^1$ approaches. All simulations have a sample size of 500.

6.1 Unordered T_1, T_2

Simulation 1. The parameter μ is $S_{T_2}(t_2) = P(T_2 > t_2)$. Both C and T_1 are $U(5, 10)$ random variables, whereas $T_2 = T_1 + e$, where e is logistically distributed with mean 0 and standard deviation .4 [the logistic distribution was chosen because it allows one to determine $Q(u)$ analytically]. The only covariate available is $W = T_1$, which obviously contains important information about the location of T_2.

The results given in Table 2 show significant improvement of the IPCW and one-step estimators relative to the Kaplan–Meier estimator (in this case, the NPMLE). As expected, the estimator μ_n^1 performs best. Misspecification of the model for $Q(u)$ results in an estimator $\mu_{n,\text{REG}}^1$ that does not always improve upon the IPCW estimator, but also does not have performance worse than μ_n^0. Finally, in all the simulations, our confidence intervals for both μ_n^1 and $\mu_{n,\text{REG}}^1$ perform well. Note, as predicted by theory, this is true even when the model for $Q(u)$ is misspecified.

Table 2. Simulation 1

t	S_2	SE^1	SE^2	SE^3	SE^4	RSE^1	RSE^2	RSE^3	CI^1	CI^2
5.7	.86	2.5	2.5	2.5	2.5	1.0	1.0	1.0	95	95
6.0	.81	3.3	3.2	3.2	3.2	1.0	1.0	1.0	95	95
6.2	.76	4.3	4.0	3.9	4.0	1.1	1.1	1.1	94	95
6.4	.71	4.8	4.4	4.3	4.4	1.1	1.1	1.1	95	95
6.7	.67	5.7	5.0	4.9	5.0	1.1	1.2	1.1	94	95
6.9	.62	6.6	5.5	5.3	5.6	1.2	1.2	1.2	94	94
7.1	.57	6.9	5.7	5.4	5.7	1.2	1.3	1.2	95	95
7.4	.52	7.4	6.1	5.7	6.0	1.2	1.3	1.2	96	96
7.6	.48	8.1	6.4	6.0	6.3	1.3	1.3	1.3	95	94
7.9	.43	8.9	6.7	6.2	6.6	1.3	1.4	1.3	94	94
8.1	.38	9.3	6.9	6.1	6.6	1.3	1.5	1.4	95	95
8.3	.33	9.1	6.7	5.8	6.4	1.4	1.6	1.4	95	94
8.6	.29	9.6	7.0	6.0	6.6	1.4	1.6	1.4	95	94

NOTE: SE^1 is MSE (×100) Kaplan–Meier estimator, SE^2 is MSE (×100) μ_n^0, SE^3 is MSE (×100) μ_n^1, SE^4 is MSE (×100) $\mu_{n,\text{REG}}^1$, $RSE^1 = MSE^1/MSE^2$, $RSE^2 = MSE^1/MSE^3$, $RSE^3 = MSE^1/MSE^4$, and CI^1 and CI^2 are the percentage of iterations that the μ_n^1 and the $\mu_{n,\text{REG}}^1$ confidence intervals contain the true $S_2(t)$, respectively.

6.2 Ordered T_1, T_2

Simulation 2. In these simulations, $T_1 \sim U(5, 7)$, $C \sim U(5, 13.5)$, and $T_2 = T_1 + L(0) + e$, where $L(0) \sim U(0, 5)$ and $e \sim U(0, .5)$. Both T_1 and $L(0)$ serve as the covariates $W = (T_1, L(0))$ for μ_n^0, μ_n^1, and $\mu_{n,\mathrm{REG}}^1$. The parameter μ of interest is $P(T_2 - T_1 > t)$. In addition to μ_n^0 and μ_n^1, we also calculated both the NPMLE based on discretization of the data defined in Section 2, ignoring the covariate $L(0)$, and the estimator of Lin et al. (1999) described in Section 3. In this setting, the NPMLE (1) estimates the marginal distribution of T_1 with the Kaplan–Meier estimator, (2) groups the uncensored observations $(\widetilde{T}_1, \Delta_1 = 1)$ into p equal size groups with $\widetilde{T}_1 \in (a_j, a_{j+1}]$, $j = 1, \ldots, p$, and (3) estimates the conditional distribution of T_2, given $T_1 = t$ with $t \in (a_j, a_{j+1}]$, with the Kaplan–Meier estimator based on the observations with an uncensored $\widetilde{T}_1 \in (a_j, a_{j+1}]$. The smaller the number of groups, p, the greater the smoothing. In this case, we found that the optimal number of groups is 8, which was used in the simulations.

The results (Table 3) show an equivalent performance for both the estimator of Lin et al. and the NPMLE, even though theory implies that the NPMLE is asymptotically more efficient. However, μ_n^0, μ_n^1, and $\mu_{n,\mathrm{REG}}^1$ have substantially increased performance by utilizing the information contributed by the covariate $L(0)$.

Simulation 3. In this simulation, we estimated the joint distribution, $P(T_2 > t_2, T_1 > t_1)$, using the data-generating distributions and covariates as in simulation 2. In addition to μ_n^0 and μ_n^1, the NPMLE described in simulation 2 is computed.

The results (Table 4) suggest that the NPMLE and μ_n^0 perform equivalently. As in the preceding simulations, the one-step estimator, μ_n^1, has high relative efficiency, whereas $\mu_{n,\mathrm{REG}}^1$ gains little over the IPCW estimator. In Figure 1, we show an example of the true $Q(u)$ (for a fixed u and T_1) and the best linear fit versus W; a simple linear model does not fit well and this is the reason that $\mu_{n,\mathrm{REG}}^1$ is not as efficient as μ_n^1. Thus, to maximize the efficiency of the one-step estimator, one should explore different possible models for $Q(u)$. However, even in the case that the model is poorly chosen, there appears little risk in trying to improve upon the IPCW estimator and, as discussed in Section 4.2, this risk can be further reduced by the alternative one-step estimator μ_n^*.

Table 3. Simulation 2

t	S	SE^1	SE^2	SE^3	SE^4	SE^5	RSE^1	RSE^2	RSE^3	RSE^4	CI
.7	.90	2.3	2.3	2.2	2.1	2.3	1.0	1.0	1.1	1.0	94, 94
1.1	.80	4.1	4.1	3.7	3.3	3.8	1.0	1.1	1.2	1.1	96, 95
1.5	.70	5.6	5.6	4.8	4.4	4.7	1.0	1.2	1.3	1.2	95, 95
1.9	.60	6.7	6.7	5.5	5.0	5.5	1.0	1.2	1.3	1.2	94, 94
2.3	.50	7.2	7.2	5.9	5.4	5.9	1.0	1.2	1.3	1.2	95, 95
2.6	.40	7.2	7.2	5.7	5.1	5.7	1.0	1.3	1.4	1.3	95, 95
3.0	.30	6.9	6.8	5.5	4.5	5.5	1.0	1.2	1.5	1.3	95, 95
3.4	.20	5.8	5.8	5.0	4.0	5.0	1.0	1.2	1.5	1.3	94, 94
3.8	.10	3.7	3.8	3.5	2.5	3.5	1.0	1.1	1.5	1.2	95, 95

NOTE: SE^1 is MSE (×100) estimator of Lin et al. (1999), SE^2 is MSE (×100) NPMLE, SE^3 is MSE (×100) μ_n^0, SE^4 is MSE (×100) μ_n^1, SE^5 is MSE (×100) $\mu_{n,\mathrm{REG}}^1$, $RSE^1 = MSE^1/MSE^2$, $RSE^2 = MSE^1/MSE^3$, $RSE^3 = MSE^1/MSE^4$, $RSE^4 = MSE^1/MSE^5$, and CI is the percentage of iterations that the μ_n^1 and the $\mu_{n,\mathrm{REG}}^1$ confidence intervals contain the true $S(t) \equiv P(T_2 - T_1 > t)$, respectively.

Table 4. Simulation 3

t_1	t_2	S	SE^1	SE^2	SE^3	SE^4	RSE^1	RSE^2	RSE^3	CI^1	CI^2
6.6	5.5	.70	4.7	4.6	4.3	4.7	1.0	1.1	1.0	95	94
7.8	5.5	.52	6.1	6.3	5.3	6.1	1.0	1.1	1.0	95	95
7.1	6.0	.45	5.4	5.5	5.1	5.5	1.0	1.1	1.0	94	94
8.2	6.0	.31	5.5	5.6	4.8	5.4	1.0	1.1	1.0	94	95
8.9	5.5	.30	5.6	6.2	4.8	5.4	.9	1.2	1.0	95	95
7.6	6.5	.21	4.0	3.8	3.4	4.0	1.0	1.2	1.0	95	94
9.4	6.0	.17	4.4	4.5	3.2	4.0	1.0	1.4	1.1	94	94
8.8	6.5	.14	3.4	3.1	2.5	3.2	1.1	1.3	1.0	95	96
9.9	6.5	.07	2.5	2.2	1.4	2.2	1.2	1.8	1.1	95	94

NOTE: SE^1 is MSE (×100) μ_n^0, SE^2 is MSE (×100) NPMLE, SE^3 is MSE (×100) μ_n^1, SE^4 is MSE (×100) $\mu_{n,\mathrm{REG}}^1$, $RSE^1 = MSE^1/MSE^2$, $RSE^2 = MSE^1/MSE^3$, $RSE^3 = MSE^1/MSE^4$, and CI^1 and CI^2 are the percentage of iterations that the μ_n^1 and the $\mu_{n,\mathrm{REG}}^1$ confidence intervals contain the true $S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$, respectively.

7. DATA ANALYSIS

In this section, we apply our methodology to the burn victim data on time to wound excision (T_1) and time to *Staphylococcus aureus* wound infection (T_2) among 154 burn victims. The times are not necessarily ordered.

In preliminary exponential and Cox regressions, only two of the ten available covariates were found to be significantly associated with T_1 or T_2: cleansing treatment ($1 =$ body cleansing, $0 =$ routine bathing) and gender ($0 =$ male, $1 =$ female). Henceforth, we shall ignore data on other baseline covariates. In Table 5, we provide results for five different estimators. The first two are the IPCW estimators, $\mu_{n,\mathrm{KM}}^0$, and $\mu_{n,\mathrm{COX}}^0$; the next two are the one-step estimators, $\mu_{n,\mathrm{EXP}}^1$ and $\mu_{n,\mathrm{COX}}^1$; the final is Dabrowska's estimator, $\mu_{n,\mathrm{DAB}}$. For $\mu_{n,\mathrm{KM}}^0$, the censoring distribution was estimated, ignoring all covariates; that is, α in model (3) was set to 0 a priori, which could only be legitimate if censoring and failure were known to be independent. For $\mu_{n,\mathrm{COX}}^0$ and the two one-step estimators, the censoring distribution was estimated by partial likelihood, assuming the Cox model (3) with covariates gender and cleansing treatment. For both $\mu_{n,\mathrm{COX}}^1$ and $\mu_{n,\mathrm{EXP}}^1$, $Q(u)$ was estimated conditionally on the time-independent covariates gender and cleansing treatment by the conditionally independent Cox and exponential regression models, respectively. Originally, we fit the frailty model (21), but the estimate of the variance of the

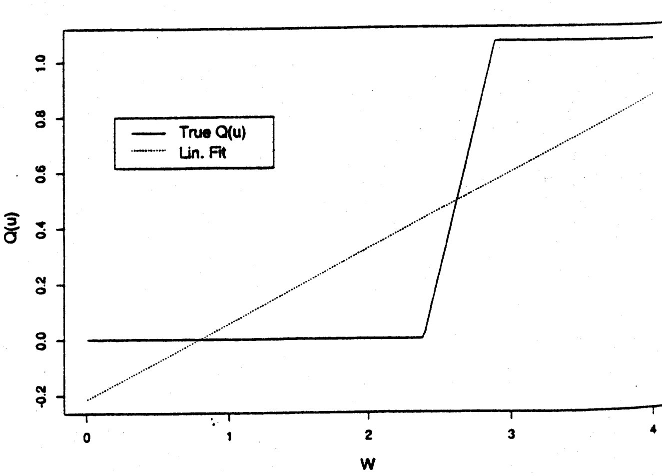

Figure 1. Plot of the True $Q(u)$ and the Best Linear Fit Versus W.

Table 5. Results of Estimation of $P(T_1 > t_1, T_2 > t_2)$

t_1, t_2	$\mu^0_{n,KM}$	$\mu^0_{n,COX}$	$\mu^1_{n,EXP}$	$\mu^1_{n,COX}$	$\mu_{n,DAB}$
8.3, 15	.52 (.05/.07)	.52 (.05/.07)	.51 (.05/.05)	.49 (.04/.04)	.46 (.04)
16.6, 15	.22 (.04/.05)	.22 (.04/.05)	.22 (.04/.04)	.28 (.04/.04)	.25 (.04)
25, 15	.10 (.04/.06)	.11 (.04/.06)	.11 (.04/.04)	.14 (.04/.03)	.10 (.04)
8.3, 30	.35 (.06/.07)	.36 (.07/.07)	.36 (.06/.07)	.42 (.04/.04)	.37 (.05)
16.6, 30	.20 (.05/.07)	.22 (.06/.08)	.22 (.05/.06)	.24 (.04/.04)	.23 (.04)
25, 30	.07 (.04/.04)	.08 (.04/.04)	.08 (.04/.04)	.12 (.04/.03)	.09 (.04)
8.3, 45	.22 (.08/.08)	.22 (.09/.09)	.22 (.09/.10)	.40 (.04/.05)	.37 (.14)
16.6, 45	.13 (.07/.08)	.14 (.08/.09)	.14 (.08/.08)	.24 (.04/.04)	.23 (.09)

NOTE: The estimators include (1) IPCW estimators using both the Kaplan–Meier ($\mu^0_{n,\mathrm{KM}}$) and the Cox regression ($\mu^0_{n,\mathrm{COX}}$) estimators of censoring, (2) the one-step estimators using exponential ($\mu^1_{n,\mathrm{EXP}}$) and Cox regression ($\mu^1_{n,\mathrm{COX}}$) for estimating $Q(u)$, and (3) Dabrowska's estimator ($\mu_{n,\mathrm{DAB}}$). The standard errors are given in parentheses (bootstrap/analytic); Dabrowska's estimator has only bootstrap standard errors.

latent frailty was found to be nearly 0 (implying conditional independence of T_1 and T_2), so our final semiparametric model used for $Q(u)$ was based on simple Cox regression models fit separately by the two time variables. With the exception of Dabrowska's estimator for which we only calculated the bootstrap estimator of the standard error, all estimators had standard errors calculated using two methods: nonparametric bootstrapping and an explicit representation of their influence curves. For the IPCW estimators, these standard errors were calculated using the explicit representation of the influence curve based on a Cox regression model of the censoring distribution discussed in Section 3. For the locally efficient estimators, the naive standard errors were calculated using the method outlined in Section 4.1.

The results are listed in Table 5. First, the IPCW estimators are very similar, implying that utilizing the covariate information using the IPCW estimator, at least as done here, does not improve the efficiency of the estimator. Likewise, $\mu^1_{n,\mathrm{EXP}}$ also has similar standard errors and so does not appear to benefit significantly from utilizing the covariate information. On the other hand, the one-step estimator, $\mu^1_{n,\mathrm{COX}}$, appears to result in a significant improvement over the IPCW estimators and the one-step using exponential regression. In addition, it appears to have significantly higher efficiency than $\mu_{n,\mathrm{DAB}}$ at the larger quantiles of T_2.

For the burn data, exponential regression is a poor model for $Q(u)$ and thus $\mu^1_{n,\mathrm{COX}}$ appears to be a better estimator than $\mu^1_{n,\mathrm{EXP}}$. For instance, the exponential model assumes a constant hazard in time, whereas the Cox regression models used to estimate $\mu^1_{n,\mathrm{COX}}$ suggest that, for both T_1 and T_2, the baseline hazards increase significantly with time. There are many ways of evaluating the relative fit of competing models for $Q(u)$, but one possible model fit statistic can be derived from the estimator discussed in Section 4.2. For the burn data, this estimator given by (18) is not a compelling alternative for either $\mu^1_{n,\mathrm{COX}}$ or $\mu^1_{n,\mathrm{EXP}}$, because both appear to be either as good as or better than the IPCW estimators. However, one of the byproducts of (18) is the estimated projection constant c_n and this constant lends itself to selecting a good fit for $Q(u)$. As discussed in Section 4.2, c_n close to 1 suggests that the estimated influence curve is close to the efficient influence curve and thus that the estimated $Q(u)$ is close to the true $Q(u)$. In Table 6, we report the c_n's for all the quantiles using both the exponential and the Cox model estimates of $Q(u)$. As expected, the c_n's are almost always closer to 1 for the Cox model relative to the exponential model, providing more evidence that the Cox model is a better choice for estimating $Q(u)$.

Table 6. c_n's for Using Both Exponential and Cox Regressions to Estimate $Q(u)$ in the One-Step Estimators

t_1	t_2	EXP	COX
8.33	15.00	1.9	.9
16.67	15.00	2.4	.7
25.00	15.00	.9	.8
8.33	30.00	−.7	.8
16.67	30.00	.9	.8
25.00	30.00	.1	.5
8.33	45.00	.4	.9
16.67	45.00	2.8	.9

Dependent censoring was not a factor for these burn data. However, dependent censoring can be a major problem in other substantive settings. For example, Robins and Finkelstein (2000) demonstrated the importance of dependent censoring attributable to strong time-dependent prognostic factors in their analysis of the effect of bactrim versus aerosolized pentamidine on the survival of AIDS patients in ACTG trial 021. They showed that adding strong prognostic factors to a Cox model for censoring not only helped correct the bias due to dependent censoring but also greatly improved efficiency. Presumably, a reanalysis of the same data but now with a bivariate survival outcome (say time to pneumocystis pneumonia PCP and time to death) would demonstrate an even greater advantage of our one-step estimator.

APPENDIX: ASYMPTOTICS OF THE ONE–STEP ESTIMATOR

An estimator μ_n of μ is asymptotically linear at $P_{F_X,G}$ with influence curve $\mathrm{IC}(Y|F_X, G, \mu)$ if $\mu_n - \mu = n^{-1}\sum_{i=1}^n \mathrm{IC}(Y_i|F_X, G, \mu) + o_P(n^{-1/2})$. 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, IC^*. The efficient influence curve is also called the canonical gradient and it is given by $\mathrm{IC}^*(Y|Q, G, \mu)$ as defined in Section 3.

In this appendix, we prove two asymptotic theorems. Theorem A.1 assumes consistent estimation of the censoring mechanism. Theorem A.2 assumes *either* consistent estimation of the censoring mechanism G *or* consistent estimation of $Q(u) = E(B|\overline{X}(u), \widetilde{T} \geq u)$: This does not require choosing which of the two quantities are consistently estimated. Obviously, the last theorem provides the most general nonparametric consistency and asymptotic normality result, but the price one has to pay is that one cannot use the explicit conservative confidence interval (17). A confidence interval either requires calculating another influence curve or using the nonparametric bootstrap. Because of this and the fact that it will typically be easier to estimate the univariate censoring mechanism than it is to estimate $Q(u)$, we feel that Theorem A.1 deserves a separate treatment.

A.1 Asymptotics Assuming Consistent Estimation of the Censoring Mechanism

Theorem A.1 shows that, if the regression $Q(u) = E(B|\overline{X}(u), \widetilde{T} \geq u)$ 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 the model for $Q(u) = E(B|\overline{X}(u), \widetilde{T} \geq u)$ is misspecified. This is due to the fact that $\mathrm{IC}^*_{\mathrm{nu}} = \int H(u, \overline{X}(u))\, dM(u)$ for a particular H and that, for any function H, $\int H(u, \overline{X}(u)) dM(u)$ has mean 0, given X, because $E(dM(u)|X) = 0$. It is important to emphasize that, for any function $H(u, \overline{X}(u))$, the stochastic integral

$$\int H(u, \overline{X}(u))\, dM(u) = H(C, \overline{X}(C))(1-\Delta)I(C \leq t) - \int_0^{\widetilde{T}\wedge t} H(u, \overline{X}(u))\Lambda_C(du|X)$$

is a function of the observed data Y because $\lambda_C(u|X)$ depends on X only through $\overline{X}(u)$.

When the model for $Q(u) = E(B|\overline{X}(u), \widetilde{T} \geq u)$ is misspecified, the influence curve of μ_n^1 depends on the model for the nuisance parameter $G(c|X)$. Characterization of this dependence requires that we introduce the notion of a tangent space. Denote by $L_0^2(P_{F_X,G})$ the Hilbert space of functions of Y with finite variance and mean 0 endowed with the covariance inner product $\langle v_1, v_2\rangle_{P_{F_X,G}} \equiv \sqrt{\int v_1 v_2\, dP_{F_X,G}}$. 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: In the main part of this article, we denoted this space by T_G.

Theorem A.1 can be used as a template to prove the local efficiency result for a one-step estimator μ_n^1. Condition (ii) in the theorem is an empirical process condition. For empirical process theory, we refer to van der Vaart and Wellner (1996). This condition is technical and depends on the submodel chosen models for the unknown parameters. After having assumed (i) $G(V|X) > \delta > 0$, this condition can typically be considered as a regularity condition. Condition (iii) basically says that G_n needs to be consistent and $Q_n(u)$ needs to converge to some function $Q_1(u)$ not necessarily equal to $Q(u)$. Given condition (i) so that the denominators are bounded away from 0, condition (iv) requires that terms of the type

$$E_X\left\{\int (G_n - G)(u|X)(Q_n - Q_1)(u)\lambda_C(u|X)\right\} \qquad \text{(A.1)}$$

are $o_P(1/\sqrt{n})$. Since smooth functionals of non/semiparametric or parametric maximum likelihood estimators for a given model are efficient under regularity conditions, condition (v) will hold under regularity conditions if G_n is a (non/semi)parametric maximum likelihood estimator of G under a given model. Condition (v) is not a condition on the choice of model for G; it just states that, whatever correct model one chooses for G, one should use an estimation procedure which is efficient for that model. Recall the notation $Pf = \int f(x)\, dP(x)$.

Theorem A.1. Let $\mu = EB$ be given and consider the one-step estimator

$$\mu_n^1 = \mu_n^0 + \frac{1}{n}\sum_{i=1}^n \mathrm{IC}^*(Y_i|Q_n, G_n, \mu_n^0).$$

We assume

(i) $G(V|X) > \delta$, F_X-a.e. for some $\delta > 0$.
(ii) $\mathrm{IC}^*(\cdot|Q_n, G_n, \mu_n^0)$ falls in a $P_{F_X,G}$-Donsker class with probability tending to 1.
(iii) For some Q^1, we have

$$\|\mathrm{IC}^*(\cdot|Q_n, G_n, \mu_n^0) - \mathrm{IC}^*(\cdot|Q^1, G, \mu)\|_{P_{F_X,G}} \to 0$$

in probability.
(iv)

$$P_{F_X,G_n-G}\{\mathrm{IC}(\cdot|Q_n, G_n, \mu) - \mathrm{IC}(\cdot|Q^1, G_n, \mu)\} = o_P(1/\sqrt{n}).$$

Define, for a G_1,

$$\Phi(G_1) = P_{F_X,G}\{\mathrm{IC}^*(\cdot|Q^1, G_1, \mu)\}.$$

(v) $\Phi(G_n)$ 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

$$\mathrm{IC} \equiv \Pi(\mathrm{IC}^*(\cdot|Q^1, G, \mu)|T_2^{\perp}(P_{F_X,G})).$$

In particular, if $\mathrm{IC}^*_{\mathrm{nu}}(\cdot|Q^1, G) = \mathrm{IC}^*_{\mathrm{nu}}(\cdot|Q, G)$ (i.e., $Q^1 = Q$), then μ_n^1 is asymptotically efficient.

It is interesting to consider what the distribution of μ_n^1 would be when $G(\cdot|x)$ is known and its known value is used in the one-step estimator. In that case, T_2 is empty. Thus, by Theorem A.1, the influence curve of μ_n^1 is given by $\mathrm{IC}^*(\cdot|Q^1, G, \mu)$, which has variance greater than or equal to that of the influence curve IC based on $G(\cdot|X)$ estimated by G_n. Because μ_n^1 reduces to μ_n^0 if one sets $Q_n(u) = 0$, Theorem A.1 also provides us with the asymptotics of the IPCW estimator μ_n^0. It teaches us that the IPCW estimator μ_n^0 using the partial likelihood estimator G_n of G assuming the Cox model (3) is asymptotically linear with influence curve equal to $\mathrm{IC}_0(Y|G, \mu)$ minus its projection on the tangent space of G as provided in Section 3. On the other hand, the IPCW estimator μ_n^0 of Lin et al. (1999) using the Kaplan–Meier estimator G_{KM} of G assuming the independent censoring model has influence curve equal to $\mathrm{IC}_0(Y|G, \mu)$ minus its projection on the smaller tangent space of G for the submodel $\lambda_C(c|X) = \lambda(c)$ of the Cox proportional-hazards model. This proves that our claim in Section 3 stating that the IPCW estimator μ_n^0 using the partial likelihood estimator of G is always at least as efficient as the estimator of Lin et al. (1999). Lemma A.1 provides a general understanding of the fact that efficient estimation of a known orthogonal nuisance parameter often leads to improvements in efficiency.

Proof of Theorem A.1. We have

$$\mu^1 = \mu_n^0 + (P_n - P_{F_X,G})\left\{\mathrm{IC}^*(\cdot|Q_n, G_n, \mu_n^0)\right\} + P_{F_X,G}\left\{\mathrm{IC}^*(\cdot\|Q_n, G_n, \mu_n^0)\right\}.$$

For empirical process theory, we refer to van der Vaart and Wellner (1996). Conditions (ii) and (iii) in the theorem imply that the empirical process term on the right-hand side is asymptotically equivalent to $(P_n - P_{F_X,G})\{\mathrm{IC}^*(\cdot|Q^1, G, \mu)\}$ plus an $o_P(1/\sqrt{n})$ term. The last term can be written as

$$P_{F_X,G}\left\{\mathrm{IC}^*(\cdot|Q_n, G_n, \mu_n^0) - \mathrm{IC}^*(\cdot|Q_n, G, \mu_n^0\right\} + P_{F_X,G}\left\{\mathrm{IC}^*(\cdot|F_{X,n}, G, \mu_n^0)\right\}.$$

Denote the first term by A. Because $P_{F_X,G}\mathrm{IC}_0(\cdot|G, \mu_n^0) = \mu - \mu_n^0$ and $P_{F_X,G}\mathrm{IC}^*_{\mathrm{nu}}(\cdot|Q_n, G) = 0$, for any Q^1,

$$P_{F_X,G}\mathrm{IC}^*(\cdot|Q^1, G, \mu_n^0) = \mu - \mu_n^0. \tag{A.2}$$

Recall the definition

$$\Phi(G_1) = P_{F_X,G}\{\mathrm{IC}^*(\cdot|Q^1, G_1, \mu)\}.$$

If one assumes that

$$P_{F_X,G}\left\{\mathrm{IC}^*(\cdot|Q_n, G_n, \mu_n^0) - \mathrm{IC}^*(\cdot|Q_n, G, \mu_n^0)\right\} = \Phi(G_n) - \Phi(G) + o_P(1/\sqrt{n}),$$

then it follows that the first term A equals

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

We now prove the equivalency of this assumption and (iv). For notational convenience, let

$$\mathrm{IC}_n(G) \equiv \mathrm{IC}(\cdot|Q_n, G, \mu),$$
$$\mathrm{IC}(G) \equiv \mathrm{IC}(\cdot|Q^1, G, \mu).$$

Note that $P_{F_X,G}\mathrm{IC}_n(G) = 0$. Similarly, $P_{F_X,G_n}\mathrm{IC}_n(G_n) = 0$. Thus,

$$P_{F_X,G}\mathrm{IC}_n(G_n) - \mathrm{IC}_n(G) \approx P_{F_X,G}\mathrm{IC}_n(G_n) \approx P_{F_X,G-G_n}\mathrm{IC}_n(G_n).$$

This proves that

$$P_{F_X,G}\{\mathrm{IC}_n(G_n) - \mathrm{IC}_n(G)\} - P_{F_X,G}\{\mathrm{IC}(G_n) - \mathrm{IC}(G)\} = P_{F_X,G_n-G}\mathrm{IC}_n(G_n) - \mathrm{IC}(G_n),$$

which proves the claim.

Thus, one can conclude that, by assumption (iv), the first term A equals

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

Thus, μ_n^1 is asymptotically linear with influence curve $\mathrm{IC}^*(\cdot|Q^1, G, \mu) + \mathrm{IC}_{\mathrm{nu},2}(\cdot|Q_1, G)$, where $\mathrm{IC}_{\mathrm{nu},2}(\cdot|F_X, G)$ is the influence curve of $\Phi(G_n)$. Now, the same argument as given in the proof of Lemma A.1 proves that this is given by

$$\Pi(\mathrm{IC}^*(\cdot|Q^1, G, \mu)|T_2^{\perp}).$$

Finally, the efficiency statement for the case that $\mathrm{IC}^*_{\mathrm{nu}}(\cdot|Q^1, G) = \mathrm{IC}^*_{\mathrm{nu}}(\cdot|Q, G)$ follows from the fact that $\mathrm{IC}_0(\cdot|G, \mu) - \mathrm{IC}^*_{\mathrm{nu}}(\cdot|Q, G)$ equals the efficient influence curve which has no component in $T_1^{\perp} \supset T_2$. □

The following lemma shows how optimal estimation of an orthogonal nuisance parameter leads to an asymptotic improvement of the estimator.

Lemma A.1. Let $Y \sim P_{F_X,G}$, G satisfying the coarsening at random condition (1). Denote the tangent space for the parameter F_X by $T_1(P_{F_X,G})$. Consider the parameter μ which is a real-valued functional of F_X. Let $\mu_n(G)$ be a regular asymptotically linear estimator of μ with influence curve $\mathrm{IC}_0(\cdot|F_X, G)$ which uses the true $G(\cdot|x)$. Assume now that, for an estimator G_n,

$$\mu_n(G_n) - \mu = \mu_n(G) - \mu + \Phi(G_n) - \Phi(G) + o_P(1/\sqrt{n}) \tag{A.3}$$

for some functional Φ of G_n. Assume that $\Phi(G_n)$ is an asymptotically efficient estimator of $\Phi(G)$ for a given model $\{G_\eta : \eta \in \Gamma\}$ with tangent space $T_2(P_{F_X,G})$. Then $\mu_n(G_n)$ is asymptotically linear with influence curve

$$\mathrm{IC}_1(\cdot|F_X, G) = \Pi(\mathrm{IC}_0(\cdot|F_X, G)|T_2(P_{F_X,G})^{\perp}).$$

Proof. $L_0^2(P_{F_X,G})$ is decomposed orthogonally in $T_1(P_{F_X,G}) + T_2(P_{F_X,G}) + T_{\perp}(P_{F_X,G})$, where $T_{\perp}(P_{F_X,G})$ is the orthogonal complement of $T_1 + T_2$. The assumptions in the lemma imply that $\mu_n(G_n)$ is asymptotically linear with influence curve $\mathrm{IC} = \mathrm{IC}_0 + \mathrm{IC}_{\mathrm{nu}}$, where $\mathrm{IC}_{\mathrm{nu}}$ is an influence curve corresponding to an estimator of the nuisance parameter $\Phi(G)$ estimated under the model with nuisance tangent space T_2. Let $\mathrm{IC}_0 = a_0 + b_0 + c_0$ and $\mathrm{IC}_{\mathrm{nu}} = a_{\mathrm{nu}} + b_{\mathrm{nu}} + c_{\mathrm{nu}}$ according to the orthogonal decomposition of $L_0^2(P_{F_X,G})$ given previously. From now on, the proof uses the following two general facts about influence curves of regular asymptotically linear estimators: An influence curve is orthogonal to the nuisance tangent space and the efficient influence curve lies in the tangent space. Since $\mathrm{IC}_{\mathrm{nu}}$ is an influence curve of $\Phi(G)$ in the model where nothing is assumed on F_X, it is orthogonal to T_1; that is, $a_{\mathrm{nu}} = 0$. Since $\Phi(G_n)$ is efficient, $\mathrm{IC}_{\mathrm{nu}}$ lies in the tangent space T_2 and hence $c_{\mathrm{nu}} = 0$ as well. In addition, $\mathrm{IC}_0 + \mathrm{IC}_{\mathrm{nu}}$ is an influence curve for an estimator of μ and hence it is orthogonal to T_2: so $b_0 + b_{\mathrm{nu}} = 0$. Consequently,

$$\mathrm{IC}_1 = \mathrm{IC}_0 + \mathrm{IC}_{\mathrm{nu}} = a_0 + c_0 = \Pi(\mathrm{IC}_0|T_2^{\perp}).$$

This completes the proof. □

A.2 Asymptotics Assuming That Either the Censoring Mechanism or the Full-Data Distribution Is Estimated Consistently

If one is only willing to assume that either the censoring mechanism or the full-data distribution is modeled correctly (i.e., one is not willing to point out which one of the two is modeled correctly), then one applies the following asymptotic theorem.

Theorem A.2. Let $\mu = EB$ be given and consider the one-step estimator

$$\mu_n^1 = \mu_n^0 + \frac{1}{n}\sum_{i=1}^{n} \mathrm{IC}^*(Y_i|Q_n, G_n, \mu_n^0).$$

Assume

(i) $G(V|X) > \delta$, F_X-a.e. for some $\delta > 0$.
(ii) $\mathrm{IC}^*(\cdot|Q_n, G_n, \mu_n^0)$ falls in a $P_{F_X,G}$-Donsker class with probability tending to 1.
(iii) For some (Q^1, G_1) with either $Q^1 = Q$ or $G_1 = G$,

$$\|\mathrm{IC}^*(\cdot|Q_n, G_n, \mu_n^0) - \mathrm{IC}^*(\cdot|Q^1, G_1, \mu)\|_{P_{F_X,G}} \to 0$$

in probability.
(iv) Define, at (Q^1, G_1),

$$\Phi_1(Q) = P_{F_X,G}\{\mathrm{IC}^*(\cdot|Q, G_1, \mu)\},$$
$$\Phi_2(G) = P_{F_X,G}\{\mathrm{IC}^*(\cdot|Q^1, G, \mu)\}.$$

Define

$$\Phi(Q, G^*) = P_{F_X,G}\{\mathrm{IC}^*(\cdot|Q, G^*, \mu)\}.$$

Assume that

$$\Phi(Q_n, G_n) - \Phi(Q^1, G_1) = \{\Phi_1(Q_n) - \Phi_1(Q^1)\} + \{\Phi_2(G_n) - \Phi_2(G_1)\} + o_P(1/\sqrt{n})$$

and that $\Phi_1(Q_n)$ and $\Phi_2(G_n)$ are regular asymptotically linear with influence curve $\mathrm{IC}_{\mathrm{nu},1}(\cdot|Q^1, G_1)$ and $\mathrm{IC}_{\mathrm{nu},2}(\cdot|Q^1, G_1)$, respectively.

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

$$IC \equiv IC^*(\cdot|Q^1, G_1, \mu) + IC_{nu}(\cdot|Q^1, G_1),$$

where

$$IC_{nu} = IC_{nu,1} + IC_{nu,2}$$
$$= \begin{cases} IC_{nu,1} & \text{if } Q^1 = Q(F_X), \\ IC_{nu,2} & \text{if } G_1 = G, \\ 0 & \text{if } (Q^1, G_1) = (Q(F_X), G)). \end{cases}$$

Thus, if $IC^*(\cdot|Q^1, G_1, \mu) = IC^*(\cdot|Q(F_X), G, \mu)$, then μ_n^1 is asymptotically efficient.

Proof. Note that

$$\mu^1 = \mu_n^0 + (P_n - P_{F_X,G})\{IC^*(\cdot|Q_n, G_n, \mu_n^0)\} + P_{F_X,G}\{IC^*(\cdot||Q_n, G_n, \mu_n^0)\}.$$

For empirical process theory, we defer to van der Vaart and Wellner (1996). Conditions (ii) and (iii) in the theorem imply that the empirical process term on the right-hand side is asymptotically equivalent to $(P_n - P_{F_X,G})\{IC^*(\cdot|Q^1, G_1, \mu)\}$ plus an $o_P(1/\sqrt{n})$ term. The last term can be written as

$$P_{F_X,G}\left\{IC^*(\cdot|Q_n, G_n, \mu_n^0) - IC^*(\cdot|Q^1, G_1, \mu_n^0)\right\} + P_{F_X,G}\left\{IC^*(\cdot|Q^1, G_1, \mu_n^0)\right\}.$$

Recall identity (A.2). The general identity (see van der Laan 1994)

$$P_{F_X,G}IC^*(\cdot|Q, G_1, \mu) = 0 \quad \text{for any } G_1 \text{ satisfying CAR.} \tag{A.4}$$

can be explicitly verified in our case. This implies that $P_{F_X,G}IC^*(\cdot|Q, G_1, \mu_n^0) = \mu - \mu_n^0$ for any G_1 satisfying CAR. Thus, if either $Q^1 = Q$ or $G_1 = G$, then

$$P_{F_X,G}IC^*(\cdot|Q^1, G_1, \mu_n^0) = \mu - \mu_n^0.$$

We conclude that μ_n^1 is asymptotically linear with influence curve $IC^*(\cdot|Q^1, G_1, \mu) + IC_{nu}(\cdot|Q^1, G_1)$, where $IC_{nu}(\cdot|Q^1, G_1)$ is the influence curve of $\Phi(Q_n, G_n)$, where one should note that the μ in the definition of $\Phi(Q_n, G_n)$ cancels out; that is, $\Phi(Q, G)$ does not depend on μ. Under condition (iv), $\Phi(Q_n, G_n)$ is asymptotically linear with influence curve $IC_{nu,1} + IC_{nu,2}$. Now, note that

1. If $G_1 = G$, then $\Phi_1(Q_n) - \Phi_1(Q) = 0$ and thus $IC_{nu,1} = 0$.
2. If $Q^1 = Q_X$, then $\Phi_2(G_n) - \Phi_2(G) = 0$ and thus $IC_{nu,2} = 0$. Thus, if both $G_1 = G$ and $Q^1 = Q_X$, then $IC_{nu} = 0$ and the estimator is efficient. □

[Received February 2000. Revised March 2001.]

REFERENCES

Andrews, D. F., and Herzberg, A. M. (1985), *Data: A Collection of Problems From Many Fields for Student and Research Worker*, New York: Springer Verlag.

Bang, H., and Tsiatis, A. A. (2000), "Estimating Medical Costs With Censored Data," *Biometrika*, 87, 329–343.

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.

Burke, M. D. (1988), "Estimation of a Bivariate Survival Function Under Random Censorship," *Biometrika*, 75, 379–382.

Clayton, D. G. (1978), "A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of a Familial Tendency in Chronic Disease Incidence," *Biometrika*, 65, 141–151.

——— (1991), "A Monte Carlo Method for Bayesian Inference in Frailty Models," *Biometrics*, 47, 467–485.

Clayton, D. G., and Cuzick, J. (1985), "Multivariate Generalizations of the Proportional Hazards Model" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 148, 82–117.

Costigan, T., and Klein, J. (1993), "Multivariate Survival Analysis Based on Frailty Models," in *Advances in Reliability*, ed. A. Basu, New York: North Holland, pp. 43–58.

Dabrowska, D. M. (1988), "Kaplan Meier Estimate on the Plane," *The Annals of Statistics*, 16, 1475–1489.

——— (1989), "Kaplan Meier Estimate on the Plane: Weak Convergence, LIL, and the Bootstrap," *Journal of Multivariate Analysis*, 29, 308–325.

Efron, B. (1967), "The Two Sample Problem With Censored Data," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California Press, pp. 831–853.

Genest, C., and MacKay, R. J. (1986), "The Joy of Copulas: Bivariate Distributions With Given Marginals," *The American Statistician*, 40, 280–283.

Genest, C., and Rivest, L. P. (1993), "Statistical Inference Procedures for Bivariate Archimedean Copulas," *Journal of the American Statistical Association*, 88, 1034–1043.

Gill, R. D. (1989), "Non- and Semi-Parametric Maximum Likelihood Estimators and the Von Mises Method," (Part 1) *Scand. J. Statist.* 16, 97–128.

——— (1992), "Multivariate Survival Analysis," *Theory of Probability and Its Applications*, 37, 18–31, 284–301.

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*, eds. D. Y. Lin and T. R. Fleming, Springer Lecture Notes in Statistics, pp. 255–294.

Gill, R. D., van der Laan, M. J., and Wellner, J. A. (1995), "Inefficient Estimators of the Bivariate Survival Function for Three Models," *Annales de l'Institut Henri Poincaré*, 31, 545–597.

Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models*, London: Chapman & Hall.

Heitjan, D. F., and Rubin, D. B. (1991), "Ignorability and Coarse Data," *The Annals of Statistics*, 19, 2244–2253.

Hougaard, P. (1986), "A Class of Multivariate Failure Time Distributions," *Biometrika*, 73, 671–678.

——— (1987), "Modelling Multivariate Survival," *Scandinavian Journal of Statistics*, 14, 291–304.

Hubbard, A. E., van der Laan, M. J., and Robins, J. M. (1999), "Nonparametric Locally Efficient Estimation of the Treatment Specific Survival Distribution With Right Censored Data and Covariates in Observational Studies," in *Statistical Models in Epidemiology: The Environment and Clinical Trials*, eds. E. Halloran and D. Berry, New York: Springer-Verlag, pp. 135–178.

Ichida, J. M., Wassell, J. T., and Keller, M. D. (1993), "Evaluation of Protocol Change in Burn-Care Management Using the Cox Proportional Hazards Model With Time-Dependent Covariates," *Statistics in Medicine*, 12, 301–310.

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.

Joe, H. (1993), "Parametric Families of Multivariate Distributions With Given Marginals," *Journal of Multivariate Analysis*, 46, 262–282.

Klein, J. P. (1992), "Semiparametric Estimation of Random Effects Using the Cox Model Based on the EM Algorithm," *Biometrics*, 48, 795–806.

Klein, J. P., and Moeschberger, M. L. (1997), *Survival Analysis: Techniques for Censored and Truncated Data*, New York: Springer-Verlag.

Lin, D. Y., Sun, W., and Ying, Z. (1999), "Nonparametric Estimation of the Gap Time Distribution for Serial Events With Censored Data," *Biometrika*, 86, 59–70.

Oakes, D. (1989), "Bivariate Survival Models Induced by Frailties," *Journal of the American Statistical Association*, 84, 487–493.

Prentice, R. L., and Cai, J. (1992a), "Covariance and Survivor Function Estimation Using Censored Multivariate Failure Time Data," *Biometrika*, 79, 495–512.

——— (1992b), "Marginal and Conditional Models for the Analysis of Multivariate Failure Time Data," in *Survival Analysis State of the Art*, eds. J. P. Klein and P. K. Goel, Dordrecht: Kluwer.

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 Finkelstein, D. M. (2000), "Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial With Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests, *Biometrics*, 56, 779–788.

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," in *AIDS Epidemiology: Methodological Issues*, Boston: Birkhauser.

——— (2002), Comments on "Inference for Semiparametric Models," by P. J. Bickel and J. Kwon, unpublished manuscript submitted to *Statistica Sinica*.

Robins, J. M., Rotnitzky, A., and Scharfstein, D. O. (1999), "Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models," in *Statistical Models in Epidemiology: The Environment and Clinical Trials*, eds. E. Halloran and D. Berry, New York: Springer-Verlag, pp. 1–94.

Robins, J. M., Rotnitzky, A., and van der Laan, M. J. (1999), Discussion of "On Profile Likelihood," by S. A. Murphy and A. W. van der Vaart, *Journal of the American Statistical Association*, 95, 477–482.

Scharfstein, D. O., and Robins, J. M. (2002), "Estimation of the Failure Time Distribution in the Presence of Informative Censoring," unpublished manuscript, revised for publication in *Biometrika*.

Strawderman, R. L. (2000), "Estimating the Mean of an Increasing Stochastic Process at Censored Stopping Time," *Journal of the American Statistical Association*, 95, 1192–1208.

Tsai, W. Y., Leurgans, S., and Crowley, J. (1986), "Nonparametric Estimation of a Bivariate Survival Function in the Presence of Censoring," *The Annals of Statistics*, 14, 1351–1365.

Turnbull, B. W. (1976), "The Empirical Distribution With Arbitrarily Grouped Censored and Truncated Data," *Journal of the Royal Statistical Society*, Ser. B, 38, 290–295.

van der Laan, M. J. (1994), "Modified EM-Equations-Estimator of the Bivariate Survival Function," *Mathematical Methods of Statistics*, 3, 213–243.

——— (1996a), "Efficient Estimator of the Bivariate Survival Function and Repairing NPMLE," *The Annals of Statistics*, 24, 596–627.

——— (1996b), *Efficient and Inefficient Estimation in Semiparametric Models*, CWI Tract 114, Amsterdam: Centre for Mathematics and Computer Science.

van der Laan, M. J., and Robins, J. M. (1998), "Locally Efficient Estimation With Current Status Data and Time-Dependent Covariates," *Journal of the American Statistical Association*, 93, 693–701.

——— (2001), *Unified Methods for Censored Longitudinal Data*, New York: Springer-Verlag.

van der Laan, M. J., and Zhuo, Y. (2001), Comments on the millennium "Inference for Semiparametric Models: Some Questions and an Answer," by P. J. Bickel and J. Kwon in the *Millennium Series of Statistica Series*, 910–917.

van der Laan, M. J., Robins, J. M., and Gill, R. D. (2000), "Locally Efficient Estimation in Censored Data Models," Technical Report 85, University of California, Berkeley, Division of Biostatistics.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer-Verlag.

Wang, W., and Wells, M. T. (1998), "Nonparametric Estimation of Successive Duration Times Under Dependent Censoring," *Biometrika*, 85, 561–572.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989), Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions, *Journal of the American Statistical Association*, 84, 1065–1073.