Published in final edited form as:
Ann Stat. 2003 ; 31(2): 512–535. doi:10.1901/jaba.2003.31-512.

CURRENT STATUS AND RIGHT-CENSORED DATA STRUCTURES WHEN OBSERVING A MARKER AT THE CENSORING TIME[1]

MARK J. VAN DER LAAN and **NICHOLAS P. JEWELL**
University of California, Berkeley

Abstract

We study nonparametric estimation with two types of data structures. In the first data structure n i.i.d. copies of $(C, N(C))$ are observed, where N is a finite state counting process jumping at time-variables of interest and C a random monitoring time. In the second data structure n i.i.d. copies of $(C \wedge T, I(T \le C), N(C \wedge T))$ are observed, where N is a counting process with a final jump at time T (e.g., death). This data structure includes observing right-censored data on T and a marker variable at the censoring time.

In these data structures, easy to compute estimators, namely (weighted)-pool-adjacent-violator estimators for the marginal distributions of the unobservable time variables, and the Kaplan–Meier estimator for the time T till the final observable event, are available. These estimators ignore seemingly important information in the data. In this paper we prove that, at many continuous data generating distributions the ad hoc estimators yield asymptotically efficient estimators of $\sqrt{n}$-estimable parameters.

Key words and phrases

Asymptotically linear estimator; asymptotically efficient estimator; current status data; right; censored data; isotonic regression

1. Introduction

In this paper we study nonparametric estimation with two types of data structures. First, we discuss these two data structures in detail. Subsequently, we provide an overview of the rest of the paper.

1.1. Current status data on a finite counting process

Consider a finite state counting process $N(t)=\sum_{j=1}^{k} I(T_j \le t)$, $T_1 < \cdots < T_k$, where T_j is the time-variable at which a specified event occurs and where N jumps from value $j-1$ to j at time T_j. The number of jumps k is fixed and known. We allow that there is a positive probability that the counting process never reaches jump j_0 for any particular $j_0 \in \{1, \ldots, k\}$; since $T_1 < \cdots < T_k$, this implies that there is also a positive probability that N never reaches jump j for $j = j_0, \ldots, k$: that is, we

[1]Supported by a FIRST award (GM53722) from the National Institute of General Medical Sciences and the National Institutes of Health.

DIVISION OF BIOSTATISTICS, UNIVERSITY OF CALIFORNIA, BERKELEY, CALIFORNIA 94720, E-MAIL: laan@stat.berkeley.edu.

allow multivariate distributions of $(T_1, \ldots, T_k)$ with $P(T_j = \infty) > 0$ for $j = j_0, \ldots, k$. In this manner we allow applications in which the number of jumps of N is random on $\{1, \ldots, k\}$.

We consider the data structure $(C, N(C))$ for a single random monitoring time C. The only assumption is that C is independent of N: the cumulative distribution G of C, and the probability distribution F of N are unspecified. Note that the distribution of N, denoted by F, is not a cumulative distribution function, but a probability distribution that is identified by the multivariate cumulative distribution of $(T_1, \ldots, T_k)$.

Such data structures occur in cross-sectional studies where each subject is monitored once. For example, in some carcinogenicity experiments, one can only determine a discretized occult tumor size at time t in a randomly sampled mouse, as measured by $N(t)$, by sacrificing a mouse at time t. In this example, T_1 might represent time till onset of the tumor and $T_2, \ldots, T_k$ might correspond with times till increasing sizes of the tumor. Similarly, T_j might denote the age at which a child has mastered the j th skill among a set of k skills ordered in difficulty. We refer to Jewell and van der Laan (1995) for additional applications.

The distribution of $(C, N(C))$ depends on the distribution of $\vec{T} = (T_1, \ldots, T_k)$ only through the marginal distributions F_j of T_j, $j = 1, \ldots, k$ (see Section 2). In this problem, the NPMLE of the distribution of T_j requires an iterative algorithm. On the other hand, an ad hoc method for estimation of the distribution of T_j is directly available: reduce the observation $(C, N(C))$ to a standard current status observation $(C, \Delta_j = I(T_j \leq C))$ on T_j. Then one can estimate the distribution of T_j with the NPMLE based on the reduced current status observations, which we will refer to as the reduced data NPMLE (RNPMLE). This estimator provides regular and asymptotically linear estimators of pathwise differentiable functionals of F_j such as $\mu_j = \int(1 - F_j)(u)r(u)\,du$, for a given r, in the nonparametric model under certain conditions [Groeneboom and Wellner (1992)]. Previous work and examples of traditional current status data on a time variable T can be found in Diamond, McDonald and Shah (1986), Jewell and Shiboski (1990), Diamond and McDonald (1992), Keiding (1991) and Sun and Kalbfleisch (1993). In its nonparametric setting, the current status data structure is also known as case I interval censored data [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) give two examples that arise from studies of Human Immunodeficiency Virus (HIV) disease.

Note that the RNPMLE of F_j ignores the value of $N(C)$, beyond information on whether $N(C) \geq j$ or not. For example, if $N(t)$ is tumor size in a carcinogenicity experiment, then the simple current status estimator of the distribution of time, T_1, till onset of tumor would not distinguish between an observation $(C, N(C))$ with $N(C)$ large and an observation $(C, N(C))$ with $N(C)$ small but larger than 0, while the latter observation seems to suggest that onset occurred recently. Nonetheless, we establish that the RNPMLE yields efficient estimators of pathwise differentiable parameters at a large class of continuous data generating distributions of interest.

1.2. Current status data on a finite counting process when the final event is right censored

We also consider the data structure $(\tilde{T}_k \equiv C \wedge T_k, N(\tilde{T}_k))$ for a finite state counting process $N(t)=\sum_{j=1}^{k} I(T_j \leq t)$, where T_k represents the final event (say death) which is right censored by the monitoring time C, and k is known. Note that this observation includes observing the failure indicator $I(\tilde{T}_k = T_k)$. For example, consider a carcinogenicity experiment with mice in which T_1 is time till onset of colon tumor, T_2 time to liver metastasis and T_3 time to death from tumor, where we assume that colon tumors do not cause death except through liver failure secondary to metastasis. Here C is either a sacrificing time or time till death from any unrelated cause.

Consider another example concerning estimation of the survival function of the time $T = J - I$ between time I at seroconversion and time J at death of a hemophiliac patient infected with HIV. For this purpose we observe n i.i.d. subjects in a fixed time-interval of 10 years. If we assume that the time I at seroconversion of the subject is observed (which is approximately true for hemophiliacs), then the subject's survival time T is right censored by $C \equiv 10 - I$, where T will play the role of T_k. We define T_j as the time till a given monotone "surrogate" process $Z(t)$ achieves a particular value among a set of $k - 1$ increasing values, $j = 1, \ldots, k - 1$, where we assume that death $T = T_k$ always and only occurs after the value $Z(T_{k-1})$ has been reached. Let $N(t)=\sum_{j=1}^{k} I(T_j \le t)$ be the counting process. Here $Z(t)$ measures the progression of the disease of the subject t years after seroconversion; for example, $Z(t)$ might be a measure of viral load of the subject t years after seroconversion, where it may be reasonable to assume that the viral load is a nondecreasing process in the absence of treatment.

Suppose that for every subject who did not die before the end of the study C one measures the "surrogate" $Z(C)$ at time C only. In other words, we observe failure times only for subjects who fail before end of follow up and for every subject who is alive at end of follow up we also have a marker indicating future prognosis. Note that the observed data on a subject is given by $(\tilde{T} = T \wedge C, Z(\tilde{T}))$. We only assume that C is independent of Z. A seemingly ad hoc estimator of $S(t) = P(T > t)$ is the Kaplan–Meier estimator which simply ignores the marker information. In this example, a natural question is whether one can improve on the Kaplan–Meier estimator using the information in the surrogate process Z. In this paper we prove that the Kaplan–Meier estimator is asymptotically efficient at many continuous data generating distributions for which F_j have compact support.

A special case of this data structure has been treated in the literature. Consider a carcinogenicity experiment with $N(t)=\sum_{j=1}^{2} I(T_j \le t)$, T_1 is time till onset of tumor and T_2 is time till death from tumor. Thus one observes $(\tilde{T}_2 \equiv C \wedge T_2, N(\tilde{T}_2))$. This data structure has been considered in Kodell, Shaw and Johnson (1982), Dinse and Lagakos (1982), Turnbull and Mitchell (1984), van der Laan, Jewell and Peterson (1997), and recently Groeneboom (1998). The NPMLE for this data structure requires an iterative algorithm: Turnbull and Mitchell (1984) implemented the NPMLE via the EM-algorithm (using an initial distribution with point masses at each data point so that the EM-algorithm indeed converges to the NPMLE), while Groeneboom (1998) implements the NPMLE by maximizing the actual likelihood with a modern optimization algorithm. In this problem, an ad hoc estimator of the distribution of T_2 is the Kaplan–Meier estimator based on the reduced data $(\tilde{T}_2, \Delta_2 = I(\tilde{T}_2 = T_2))$. In Dinse and Lagakos (1982), the Kaplan–Meier estimator of F_2 was proposed and it was suggested that the NPMLE might be more efficient than the Kaplan–Meier estimator. In van der Laan, Jewell and Peterson (1997) it is shown that the Kaplan–Meier is efficient under a weak condition on (F_1, F_2). Moreover, an isotonic regression estimator of F_1 was provided: note that estimation of F_1 is complicated by the fact that for some subjects one only observes T_2 and thus that $T_1 < T_2$, where T_2 cannot be viewed as an independent monitoring time for T_1. We note here that, in van der Laan, Jewell and Peterson (1997), a simulation study was carried out which incorrectly implements the NPMLE, so that finite sample comparisons between the Kaplan–Meier estimator and the NPMLE remain open to study [specifically the derivation of the score equations in van der Laan, Jewell and Peterson (1997) for the NPMLE was not valid since the authors incorrectly assumed that the NPMLE $\hat{F}_1$ is strictly smaller than the NPMLE $\hat{F}_2$].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

1.3. Organization and overview of results

In Section 2 we prove, for the data structure of Section 1.1, that if the F_j's are continuous with Lebesgue density bounded away from zero on $[0, \tau_j]$ and zero elsewhere, and G is also continuous, then any estimator of a parameter $\mu=\Phi(F)\mathbb{R}$ that is regular and asymptotically linear at $P_{F,G}$ is also asymptotically efficient. The complexity of the NPMLE is discussed including that it is more efficient at many data generating distributions with singular pairs F_{j1}, F_{j2}; for example, F_1 discrete and F_2 continuous

In Section 3, we prove an analogous result for the nonparametric model with the data structure $(C \wedge T_k, N(C \wedge T_k))$. This shows that the Kaplan–Meier estimator of the distribution of T_k, based on the reduced data $(\tilde{T}_k, \Delta_k \equiv I(T_k \leq C))$, is asymptotically efficient at many continuous data generating distributions, extending the result in van der Laan, Jewell and Peterson (1997) for the case $k = 2$. Moreover, simple isotonic regression estimators for the distributions F_j, $j = 1, \ldots, k - 1$, are proposed that also yield asymptotically efficient estimators of smooth functionals by our general result.

2. Current status data on a counting process

2.1. Traditional current status data

Traditional current status data can be viewed as current status data on a simple counting process as follows. Let T be a univariate failure time of interest and define the process $\Delta(t) = I(T \leq t)$ as the counting process with one single jump at point T. Let $Y = (C, \Delta(C))$ represent current status data on Δ at a monitoring time C. We assume that C is independent of T [i.e., of $\Delta(\cdot)$]. The parameter of interest is the distribution F of T.

The properties of the NPMLE F_n of the distribution of T were established in Groeneboom and Wellner (1992). Here the NPMLE is defined as the maximum likelihood estimator over all discrete distributions with jumps at the monitoring times. Beyond proving a limit distribution result for F_n, these authors also established efficiency of smooth functionals of F_n with a closed form expression of the limit variance so that Wald-type confidence intervals are directly available. Huang and Wellner (1995) provide an alternative proof of asymptotic linearity of the NPMLE of smooth functionals of F under weak conditions.

We refer to Bickel, Klaassen, Ritov and Wellner (1993) for definitions of a *regular, asymptotically linear and efficient estimator* and *influence curve* of an estimator. The semiparametric-information bound at $P_{F,G}$ is defined as the infimum of parametric information bounds over a specified class of parametric submodels. We choose as parametric one-dimensional submodels

$$\left\{\varepsilon \to P_{F_{\varepsilon,h_1},G_{\varepsilon,h_2}} : \| h_j\|_\infty<\infty,\ j=1,2,\ \int h_1 dF=\int h_2 dG=0\right\},$$

where $dF_{\varepsilon,h_1}(\cdot) = (1 + \varepsilon,h_1(\cdot))\, dF(\cdot)$, $dG_{\varepsilon,h_2}(\cdot) = (1 + \varepsilon,h_2(\cdot))\, dG(\cdot)$ and ε is the unknown parameter with parameter space $[-\delta, \delta]$ for some small $\delta > 0$. The tangent space at $P_{F,G}$ is now defined as the closure in $L_0^2(P_{F,G})$ of the linear span of all the scores of these one-dimensional submodels, where, for a given measure μ, we define $L_0^2(\mu)=\{h: \int h^2 d\mu<\infty, \int h d\mu=0\}$ as the Hilbert space endowed with inner product $\langle h_1, h_2\rangle_\mu = \int h_1(y)h_2(y)\, d\mu(y)$. Thus the tangent space at $P_{F,G}$ is a sub-Hilbert space of $L_0^2(P_{F,G})$.

In this paper it is particularly important to realize that efficiency of an estimator is a local property in the sense that a regular estimator can be efficient at a particular $P_{F,G}$ and inefficient at another element of the model.

Lemma 2.1—*Consider the nonparametric model for* $Y = (C, \Delta(C))$, *where* $\Delta(\cdot) \equiv I(T \leq \cdot)$, T *has unspecified distribution F and C is independent of T with unspecified distribution G. We observe n i.i.d. observations of* $Y = (C, \Delta(C))$. *Consider the parameter* $\mu = \int(1 - F_n)(u)r(u)\,du$ *for a given function r. Consider the estimator* $\mu = \int(1 - F_n)(u)r(u)\,du$, *where* F_n *is the NPMLE of F. We have that* μ_n *is regular and asymptotically linear at any* (F, G) *for which F is continuous with density* $f_T > 0$ *on* $[0, M]$ *and zero elsewhere* $(M < \infty)$, $g(x) = dG/dx > 0$ *on* $[0, M]$, *and r is bounded on* $[0, M]$.

The influence curve of μ_n is given by

$$IC(Y|F, g, r)=\frac{r(C)}{g(C)}(F(C)(1-\Delta)-(1-F(C))\Delta). \qquad (1)$$

The variance of IC is given by

$$\mathrm{VAR}(IC)=\int\frac{r^2(c)}{g(c)}F(c)(1-F(c))dc.$$

This lemma is proved in Huang and Wellner (1995).

We can also prove the following tangent space result.

Lemma 2.2—*Consider the nonparametric model for* $Y = (C, \Delta(C))$, *where* $\Delta(\cdot) \equiv I(T \leq \cdot)$, T *has unspecified distribution F and C is independent of T with unspecified distribution G. We observe n i.i.d. observations of* $Y = (C, \Delta(C))$. *Suppose that:*

1. *F has a Lebesgue density f with* $f > 0$ *on* $[0, \tau_F)$ *and, if* $\tau_F < \infty(\tau_F = \infty$ *is allowed), then* $f = 0$ *on* $(\tau_F = \infty)$, *and*
2. G has a Lebesgue density g.

We allow $F(\{\infty\}) > 0$. *Then the tangent space at* $P_{F,G}$ *equals* $L_0^2(P_{F,G})$. *This implies that an estimator of a parameter* $\mu(F)$ *which is regular and asymptotically linear at* $P_{F,G}$ *is also asymptotically efficient if F, G satisfy* (1) *and* (2).

In Gill, van der Laan and Robins (1997) it is proved that if one only assumes that the conditional distribution of the observed data Y, given the full data T, satisfies “coarsening at random” (CAR), then the tangent space at $P_{F,G}$ is saturated, that is, equals $L_0^2(P_{F,G})$. The tangent space generated by $G(\cdot \mid T)$ under the sole assumption CAR equals $T_{\mathrm{CAR}}=\{v(Y) \in L_0^2(P_{F,G}):E(v(Y)|T)=0\}$. Therefore, the main idea of the proof below is to show that under the independent censoring model $G(\cdot \mid T) = G(\cdot)$, the tangent space of the marginal distribution G equals T_{CAR} at a $P_{F,G}$ satisfying (1) and (2) of Lemma 2.2. The proof below will be an ingredient of the proofs of our two main theorems.

Proof of Lemma 2.2—Let $A:L_0^2(F) \to L_0^2(P_{F,G})$, $A(h)(Y) = E_F(h(T)|Y)$ be the score operator for F and let $A^\top:L_0^2(P_{F,G}) \to L_0^2(F)$, $A(V)(T) = E_G(V(Y)|T)$ be its adjoint. The closure of the range of a Hilbert space operator equals the orthogonal complement of the null-space of its adjoint; that is, $\overline{R(A)}=N(A^\top)^\perp$, where $\overline{R(A)}$ is the closure of the range of the score operator and $N(A^\top)$ is the null space of $A^\top$. Thus $L_0^2(P_{F,G})=\overline{R(A)}+N(A^\top)$.

The data generating distribution is indexed by two locally variation-independent parameters F and G, so that the tangent space at $P_{F,G}$ can be obtained as a sum of two tangent spaces, namely the tangent space for F, which is given by $\overline{R(A)}$, and the tangent space for G. For every $h \in L_0^2(G)$ with finite supremum norm, we have that $\varepsilon \to (1 + \varepsilon h_2)\, dG$ is a one-dimensional submodel through G at $\varepsilon = 0$. Thus the tangent space corresponding with submodels $\varepsilon \to P_{F,G_\varepsilon}$ equals $L_0^2(G)$. Thus we have that the tangent space is given by $\overline{R(A)}+L_0^2(G)$. We conclude that it suffices to show that $N(A^\top)=L_0^2(G)$.

We have

$$A^\top(V)(T)=\int_0^T V(c,0)dG(c) + \int_0^\infty V(c,1)dG(c).$$

Thus $\int V(c, \Delta(c))\, dG(c) = 0$ F-a.e. implies that

$$\int_0^T \{V(c,0)-V(c,1)\}\, g(c)dc = \int_0^\infty V(c,1)dG(c) \quad \text{for } T \in [0,\tau_F). \tag{2}$$

Differentiation w.r.t T yields $V(C, 0) = V(C, 1)$ on $[0, \tau_F)$ G-a.e. If $\tau_F < \infty$ and $c > \tau_F$, then $c > T$ and thus $V(c, \Delta(c)) = V(c, 1)$. Thus $V(C, 0) = V(C, 1)$ G-a.e. which proves $N(A^\top)=L_0^2(G)$. □

It is of interest to note that one can represent $F_T(t)$ as a monotonic regression of Δ on C since $F(t) = E(\Delta(C) \mid C = t)$. This suggests that one can estimate F_T with the estimator $F_n(t)$ which minimizes $\sum_{i=1}^n (\Delta(C_i) - F_T(C_i))^2$ over all distribution functions F_T. $F_n(t)$ can be computed using the pool-adjacent-violator-algorithm [see Barlow, Bartholomew, Bremner and Brunk (1972)] which, in fact, yields the NPMLE.

2.2. Current status data on a counting process

Let the process of interest be a counting process $N(t)=\sum_{j=1}^k I(T_j \le t)$, $T_1 < \cdots < T_k$, where T_j is the time-variable at which an event occurs and where N jumps from value $j-1$ to j. Let C be a monitoring time and consider the data structure $Y = (C, N(C))$. We observe n i.i.d. copies of Y. We only assume that C is independent of N.

The distribution of $(C, N(C))$ depends on the distribution of $\vec{T}$ only through the marginal distributions F_j of T_j, $j = 1, \ldots, k$. To be precise, we have (denoting $S_i = 1 - F_i$), for $j \in \{0, \ldots, k\}$,

$$\begin{aligned} P_{F,G}(dc, N(C)=j) = &I(j=0)S_1(c)dG(c) + I(j=k)F_k(c)dG(c) \\ &+ I(j=1)\{S_2(c) - S_1(c)\}dG(c) \\ &+ \cdots + I(j=k-1)\{S_k(c) - S_{k-1}(c)\}dG(c). \end{aligned}$$

Thus the distribution of $Y = (C, N(C))$ only identifies the marginal distributions of T_j, $j = 1, \ldots, k$.

The NPMLE does not exist in closed form and can only be computed with an iterative algorithm. For a given j, we can reduce the observation $(C, N(C))$ to simple current status data $(C, \Delta_j = I(T_j \le C))$ on T_j, and estimate F_j with the RNPMLE. Under the conditions stated in Lemma 2.1, with $F = F_j$ and $G = G$, this estimator provides regular and asymptotically linear estimators of smooth functionals of the type $\mu_j = \int(1 - F_{T_j})(u)r(u)\,du$,, for a given r in the nonparametric model. The following theorem proves that, at a data generating distribution of Y satisfying a specified condition, any regular asymptotically linear estimator will provide asymptotically efficient estimators of smooth functionals of F_{T_j}. We decided to state a condition (3) which is easy to understand, but our proof shows that this can be weakened, for example, to allow the analogue of (3) for the case where all distributions $G, F_1, \ldots, F_k$ are discrete with a finite number of support points; that is, the support points of F_j are contained in the support points of F_j are contained in the support points of F_{j+1}, $j = 1, \ldots, k-1$, and G is discrete with support contained in the support of F_k.

Theorem 2.1—*Let $T_1 < T_2 < \cdots < T_k$ be time-variables corresponding to the chronological events of interest. Define the counting process with jumps of size* 1 *at these T_j's by*

$$N(t) = \sum_{j=1}^{k} I(T_j \le t).$$

Let $Y = (C, N(C))$. Consider the following semiparametric model for Y: Let $C \sim G$ be independent of $\vec{T} \sim F$, but leave G and F unspecified. Then, the distribution of Y only depends on the multivariate distribution F of $\vec{T} = (T_1, \ldots, T_k)$ through the marginal distributions $F_1, \ldots, F_k$ of $T_1, \ldots, T_k$.

Consider a data generating distribution $P_{F,G}$ in the model above, satisfying the following condition (3): *For certain $\tau_1 < \cdots < \tau_k < \infty$ let F_j have Lebesgue density f_j on $[0, \tau_j]$ with*

$$\begin{array}{lll} f_j > 0 & \text{on } [0, \tau_j] \text{ and } f_j = 0 & \text{on } (\tau_j, \infty),\ j = 1, \ldots, k, \\ F_j > F_{j+1} & \text{on } (0, \tau_j],\ j = 1, \ldots, k-1, & \\ G & \text{has Lebesgue density } g. & \end{array} \tag{3}$$

We allow that $p_j \equiv P(T_j = \infty) > 0$ for $j = j_0, \ldots, k$ and $j_0 \in \{1, \ldots, k\}$.

Then the tangent space at $P_{F,G}$ equals $L_0^2(P_{F,G})$ and is thus saturated.

This implies that any estimator of a real valued parameter of F that is a regular and asymptotically linear estimator at $P_{F,G}$ is also asymptotically efficient if $P_{F,G}$ satisfies (3). *In particular, given $j \in \{1, \ldots, k\}$, if $P_{F,G}$ satisfies* (3), *and F_j, G satisfy the conditions of* Lemma 2.1 *for the RNPMLE of μ_{F_j} based on* $(C, I(T_j = C))$ (*thus with $F = F_j$ and $G = G$*), *then the RNPMLE of μ_{F_j} is asymptotically efficient.*

2.2.1. Heuristic understanding of the difference between NPMLE and RNPMLE

—To understand the difference between the NPMLE and the RNPMLE, we consider the special case $k = 2$ in detail. In this case N can have three possible values:

$$N(C)=\begin{cases} 0, & \text{if } C<T_1, \\ 1, & \text{if } T_1<C<T_2 \\ 2, & \text{if } C>T_2. \end{cases},$$

Let us assume that C has a Lebesgue density g. The likelihood of $(C, N(C))$ is given by

$$p_{F_1,F_2,G}(c, N(c)=j)=S_1(c)^{I(j=0)}(S_2-S_1)(c)^{I(j=1)}F_2(c)^{I(j=2)}g(c).$$

We note that the density $p_{F_1,,F_2,G}$ can be reparametrized as

$$p_{R,F_2,G}(c, \delta)=R(c)^{I(j=0)}(1-R(c))^{I(j=1)}S_2(c)^{I(j\in\{0,1\})}F_2(c)^{I(j=2)}g(c),$$

where $R(t) \equiv S_1(t)/S_2(t)$. Thus, if we ignore the relation between F_2 and R, then the NPMLE of F_2 of the likelihood corresponding with $p_{R,F_2,G}$ would actually be equal to the reduced data NPMLE based on the reduced data $(C, I(T_2 \leq C))$. However, F_2 and R are related since S_2R has to be a survival function. Therefore, it is not possible to determine the NPMLE by separate maximization w.r.t. F_2 and R, which explains why the NPMLE and the RNPMLE of F_2 differ.

Theorem 2.1 shows that this relation between F_2 and R is not informative for estimation of smooth functionals of F_2 at a large class of data generating distributions, since the RNPMLE, which ignores this relation, is still asymptotically efficient for estimation of $\sqrt{n}$-estimable parameters. Our proof of Theorem 2.1 for $k = 2$ shows that the efficient score operator (for the definition of an efficient score operator, see the proof) of F_2 equals the efficient score operator for F_2 in the reduced data model based on (C, Δ_2). This implies that, at (F_1, F_2) satisfying (3), the efficient influence curve for any smooth functional of F_2 equals the influence curve of the RNPMLE as given in Lemma 2.1. Closer inspection of the proof for $k = 2$ also shows that, if (e.g.) F_2 is continuous while F_1 is discrete on $[0, \tau_1]$, or F_2 is discrete with support not containing the support of a discrete F_1, then the efficient score operator for F_2 is not the same as the efficient score operator for F_2 in the reduced data model, so that, in particular, the efficient influence curves (and information bounds) differ for the two models. Thus, at such (F_1, F_2), the RNPMLE of smooth functionals of F_2 is inefficient.

Here, we provide a likelihood-based explanation of this fact. Let R_n be the NPMLE of R. The NPMLE of F_2 maximizes the likelihood corresponding with p_{R_n,F_2} over all F_2 for which S_2R_n is a survival function, while the RNPMLE maximizes the likelihood over all distributions F_2. Suppose now that the model consists of discrete F_1's and continuous F_2's. This model, though smaller than the model with F_1, F_2 being unspecified, has the same semiparametric efficiency bound at a (F_1, F_2) in this smaller model as the efficiency bound in the original model. This follows from the fact that the class of one-dimensional submodels as needed to compute the tangent space can still be chosen the same. In this smaller model, an $R = S_1/S_2$ will be discrete at the support points of F_1, and the shape of R between the support points equals the shape of $1/S_2$. As a consequence, since R determines the shape of F_2 between the support points, knowing R in the smaller model helps enormously in estimating S_2. In particular, for a given R_n, maximizing the likelihood corresponding with $p_{R_n,,F_2}$ over F_2 with S_2R_n being a survival function, is very different from maximizing this likelihood over all possible distributions F_2. This shows that the RNPMLE in the smaller model is inefficient at such (F_1, F_2). Since the efficiency bound in the smaller model is the same as the efficiency bound in the original model, this also shows that the RNPMLE will also be inefficient at such (F_1, F_2).

Proof of Theorem 2.1—We need to prove that assumption (3) implies that the tangent space at $P_{F,G}$ equals $L_0^2(P_{F,G})$, and is thus saturated. The data generating distribution $P_{F,G}$ is indexed by F and G, where the dependence on F is only through the marginals F_j, $j = 1, \ldots, k$. Thus, the tangent space at $P_{F,G}$ can be obtained as a sum of two tangent spaces, namely the tangent space for F and the tangent space for G, where the latter equals $L_0^2(G)$. Let F, G be given and satisfy (3). We now claim that the tangent space for F is given by the closure of the sum of the k tangent spaces for F_j calculated as if the F_j 's are variation-independent parameters, $j = 1, \ldots, k$. We will show this now. Let $h_j \in L_0^2(F_j)$ have finite supremum norm, and let F_{j,ε,h_j} be the one-dimensional perturbation $\varepsilon \to \int_0^{\cdot}(1+\varepsilon h_j)dF_j$ through F_j at $\varepsilon = 0$, $j = 1, \ldots, k$. First, note that the support of F_{j,ε,h_j} equals the support of F_j, $j = 1, \ldots, k$. Since $F_j > F_{j+1}$ (strictly) on $(0, \tau_j]$ we have that, given an arbitrarily small $\delta_1 > 0$, there exists a neighborhood $\varepsilon \in (-\delta, \delta)$ with $F_{j,\varepsilon,h_j} \geq F_{j+1,\varepsilon,h_{j+1}}$ on $(\delta, \tau_j]$ for all $j = 1, \ldots, k-1$. Thus, $P_{F_{j,\varepsilon,hj}, j = 1,\ldots,k, G}$ satisfies the constraints $F_j \geq F_{j+1}$, $j = 1, \ldots, k-1$, of our model except on an arbitrarily small neighborhood of 0. Thus, by modifying h_j on an arbitrarily small neighborhood of 0, we can make $\varepsilon \to P_{F_{j,\varepsilon,hj}, j=1,\ldots,k,G}$ a true one-dimensional submodel. Since a tangent space for F is obtained as the *closure* in $L_0^2(F)$ of the linear span of scores of all possible one-dimensional submodels, it follows that the score of the unmodified $\varepsilon \to P_{F_{j,\varepsilon,hj}, j=1,\ldots,k,G}$ also belongs to the tangent space. This proves our claim.

Let $j \in \{1, \ldots, k\}$ be given. For a given $h_j \in L_0^2(F_j)$, we consider the one-dimensional submodel $F_{j,\varepsilon}$ given by $\varepsilon \to (1 + \varepsilon h_j(t))\, dF_j(t)$ which goes through F_j at $\varepsilon = 0$. For notational convenience, define the random variable $R = N(C) + 1 \in \{1, \ldots, k+1\}$, and let F_{-j} be the $(k-1)$-dimensional vector of c.d.f.'s excluding F_j. This one-dimensional submodel $F_{j,\varepsilon}$ implies a score for $P_{F_{j,\varepsilon}, F_{-j}, G}$ given by

$$
\begin{aligned}
A_1(h_1)&=I(R=1)\frac{\int_c^\infty h_1 dF_1}{S_1(c)}-I(R=2)\frac{\int_c^\infty h_1 dF_1}{(S_2-S_1)(c)} \quad \text{if } j=1,\\
A_j(h_j)&=I(R=j)\frac{\int_c^\infty h_j dF_j}{(S_j-S_{j-1})(c)}\\
&\quad -I(R=j+1)\frac{\int_c^\infty h_j dF_j}{(S_{j+1}-S_j)(c)} \quad \text{if } j\in\{2,\ldots,k-1\},\\
A_k(h_k)&=I(R=k)\frac{\int_c^\infty h_k dF_k}{(S_k-S_{k-1})(c)}-I(R=k+1)\frac{\int_c^\infty h_k dF_k}{F_k(c)} \quad \text{if } j=k.
\end{aligned}
$$

If we define $S_0 \equiv 0$ and $S_{k+1} \equiv 1$, then, for $j = 1, \ldots, k$,

$$
A_j(h_j)=I(R=j)\frac{\int_c^\infty h_j dF_j}{(S_j-S_{j-1})(c)} - I(R=j+1)\frac{\int_c^\infty h_j dF_j}{(S_{j+1}-S_j)(c)},
$$

where we use that $S_1 - S_0 = S_1$, and $S_{k+1} - S_k = F_k$. Here $A_j:L_0^2(F_j) \to L_0^2(P_{F,G})$ is called the score operator of, $F_j = 1, \ldots, k$. The tangent space for F_j is given by the closure of the range of A_j denoted by $\overline{R(A_j)}$. Define $A_F:L_0^2(F_1) \times \cdots \times L_0^2(F_k) \to L_0^2(P_{F,G})$ by $A_F(h_1, \ldots, h_k) = A_1(h_1) + \ldots + A_k(h_k)$. Then, the tangent space for F equals $\overline{R(A_F)}$ so that the tangent space at $P_{F,G}$ is given by $\overline{R(A_F)+L_0^2(G)}$. Thus, to prove the theorem, it suffices to show that $\overline{R(A_F)+L_0^2(G)}=L_0^2(P_{F,G})$ at any F, G satisfying (3).

The remaining task is to understand the range of A_F. We decompose A_F as a sum of efficient score operators A_j^*, where A_j^* is defined as A_j minus its projection, on the sum-space spanned by the ranges of the other score operators $A_1, \ldots, A_{j-1}, A_{j+1}, \ldots, A_k$, $j = 1, \ldots, k$. We will prove that the efficient score operator of F_j at a $P_{F,G}$ satisfying (3) equals $A_j^*(h_j)=E(h_j(T_j))|(C, \Delta_j=I(T_j \le C))$, which is the score operator for the reduced current status data structure (C, Δ_j), $j = 1, \ldots, k$. Since the information bounds for smooth functionals of F_j are, in both models, solely expressed in terms of the efficient score operator for F_j, the latter result proves that an efficient estimator of μ_j based on (C, Δ_j), $j = 1, \ldots, k$, like the RNPMLE, is also efficient in the model for the more informative data structure $(C, N(C))$ [e.g., Bickel, Klaassen, Ritov and Wellner (1993)]. This proves that the RNPMLE actually yields efficient estimators. Subsequently, we show that this special structure of the efficient score operators implies that the tangent space at a $P_{F,G}$ satisfying (3) is saturated, proving the more general statement of Theorem 2.1.

Derivation of the efficient score operators of F_j—Since $E(A_l(h_l)A_m(h_m)(Y))$ is equal to 0 if $|l - m| \ge 2$, it will follow that the efficient score operators mainly involve projections of the type $\prod(A_j|\overline{R(A_{j-1})})$ and $\prod(A_j|\overline{R(A_{j+1})})$. Therefore we first obtain closed form expressions, in general, for these projection operators.

If the projection $\prod(A_j(h_j)|\overline{R(A_{j-1})})$ is actually an element of $R(A_{j-1})$, then this projection is given by (compare with the formula $X(X'X)^{-}X'Y$ for the least squares estimator):

$$\prod(A_j(h_j)|\overline{R(A_{j-1})}) = A_{j-1}(A_{j-1}^{\top}A_{j-1})^{-}A_{j-1}^{\top}A_j(h_j), \quad (4)$$

where $A_{j-1}^{\top}:L_0^2(P_{F,G}) \to L_0^2(F_{j-1})$ is the adjoint of $A_{j-1}:L_0^2(F_{j-1}) \to L_0^2(P_{F,G})$, and $(A_{j-1}^{\top}A_{j-1})$ stands for the generalized inverse of $A_{j-1}^{\top}A_{j-1}:L_0^2(F_{j-1}) \to L_0^2(F_{j-1})$. Similarly,

$$\prod(A_j(h_j)|\overline{R(A_{j+1})}) = A_{j+1}(A_{j+1}^{\top}A_{j+1})^{-}A_{j+1}^{\top}A_j(h_j). \quad (5)$$

The adjoint $A_l^{\top}$ is defined by

$$\langle A_l(h_l), \eta\rangle_{P_F} = \langle h_l, A_l^{\top}(\eta)\rangle_{F_l} \quad \text{for all } h_l \in L_0^2(F_l) \text{ and } \eta \in L_0^2(P_{F,G}).$$

It is easily shown that for $l \in \{1, \ldots, k\}$,

$$A_l^{\top}(V)(T_l) = \int_0^{T_l} \{V(c,l) - V(c,l+1)\}\, dG(c).$$

We have that

$$A_l^{\top}A_l(h_l)(T_l) = \int_0^{T_l} \varphi_l(c) \int_c^{\infty} h_l dF_l dG(c),$$

where

$$\varphi_1 = \frac{S_2}{S_1(S_2 - S_1)},$$
$$\varphi_l = \frac{S_{l+1} - S_{l-1}}{(S_{l+1} - S_l)(S_l - S_{l-1})}, \ l = 2, \ldots, k-1,$$
$$\varphi_k = \frac{F_{k-1}}{(S_k - S_{k-1})F_k},$$

or, in fact, with our convention of $S_0 = 0$ and $S_{k+1} = 1$,

$$\varphi_l = \frac{S_{l+1} - S_{l-1}}{(S_{l+1} - S_l)(S_l - S_{l-1})}, l=1, \ldots, k.$$

Here $\varphi_l(t) \equiv 0$ if $S_l(t) = 0$.

If $p_l = P(T_l = \infty) > 0$, then we can write

$$A_l^\top A_l(h_l)(T_l) = -\int_0^{\min(T_l,\tau_l)} \varphi_l(c) \int_0^c h_l dF_l dG(c) + I(T_l=\infty) h_l(\infty) p_l \int_{\tau_l}^\infty \varphi_l(c) dG(c).$$

Thus, given a K with $K \ll G$, a solution (if it exists) of $A_l^\top A_l(h_l)=K$ has to satisfy: for G-a.e., $c \in [0, \tau_l]$,

$$\int_0^c h_l dF_l = -\frac{dK}{dG} \frac{1}{\varphi_l(c)}, l=1, \ldots, k, \tag{6}$$

and, if $p_l = P(T_l = \infty) > 0$, then the equation $A_l^\top A_l(h_l)(\infty)=K(\infty)$ yields

$$h_l(\infty) = \frac{1}{p_l \int_0^\infty \varphi_l(c) dG(c)} \left\{ K(\infty) - \int_0^{\tau_l} \varphi_l(c) \int_c^{\tau_l} h_l dF_l dG(c) \right\}. \tag{7}$$

Thus, even when $p_l > 0$, (6) is the principal equation to solve (and will imply our conditions) since its solution h_l on $[0, \tau_l]$ yields the complete solution $h_l(T_l) = h_l(T_l) I_{[0,\tau_l]}(T_l) + I(T_l = \infty)$ $h_l(\infty)$. This two-step method for solving for h_l in $A_l^\top A_l(h_l)=K$ first solves for $h_l I_{[0,\tau_l]}$ and then uses that, if $p_l > 0$, $h_l(\infty)$ is a function of $h_l I_{[0,\tau_l]}$.

We have, for $l \in \{1, \ldots, k-1\}$,

$$\begin{aligned} A_l^\top A_{l+1}(h_{l+1}) &= \int_0^{T_l} [A_{l+1}(h_{l+1}) I(R=l) - A_{l+1}(h_{l+1}) I(R=l+1)] dG(c) \\ &= -\int_0^{T_l} A_{l+1}(h_{l+1}) I(R=l+1) dG(c) \\ &= -\int_0^{T_l} \frac{1}{S_{l+1}-S_l} \int_c^\infty h_{l+1} dF_{l+1} dG(c). \end{aligned}$$

We note that this element is indeed absolutely continuous w.r.t. G. Similarly, it follows that, for $l \in \{1, \ldots, k-1\}$,

$$A_{l+1}^\top A_l(h_l) = -\int_0^{T_{l+1}} \frac{1}{S_{l+1} - S_l} \int_c^\infty h_l dF_l dG(c).$$

Thus, $h_{j-1,j} \equiv (A_{j-1}^\top A_{j-1})^- A_{j-1}^\top A_j(h_j)$ is the h satisfying

$$-\int_0^c hdF_{j-1} = -\frac{1}{S_j - S_{j-1}}\int_c^\infty h_j dF_j \frac{1}{\varphi_{j-1}} \quad \text{for } G-\text{a.e.}, c \in [0, \tau_{j-1}] \tag{8}$$

for G-a.e., $c \in [0, \tau_{j-1}]$ and, if $p_{j-1} > 0$, then $h(\infty)$ is a simple function of $hI_{[0,\tau_{j-1}]}$ as given above. Similarly, $h_{j+1,j} \equiv (A_{j+1}^\top A_{j+1})^- A_{j+1}^\top A_j(h_j)$ is the h satisfying

$$-\int_0^c hdF_{j+1} = -\frac{1}{S_{j+1} - S_j}\int_c^\infty h_j dF_j \frac{1}{\varphi_{j+1}} \quad \text{for } G-\text{a.e.}, c \in [0, \tau_{j+1}] \tag{9}$$

for G-a.e., $c \in [0, \tau_{j+1}]$ and, if $p_{j+1} > 0$, then $h(\infty)$ is a simple function of $hI_{[0,\tau_{j+1}]}$. If we can take a derivative of the right-hand sides in (8) and (9) w.r.t. F_{j-1} and F_{j+1}, then, in terms of h, equations (8) and (9) have a solution. This is possible if $F_j \ll F_l$ (i.e., F_j is absolute continuous w.r.t. F_l) on $[0, \tau_l]$, $l \in \{j-1, j+1\}$, which holds under assumption (3) since we assumed that all F_j have positive Lebesgue density on $[0, \tau_j]$. The efficient score operator A_j^* also involves projections requiring existence of solutions $h_{l-1,l}$, $h_{l+1,l}$ for l different from j. Therefore, the assumed condition (3) includes (via an easy to understand condition) the necessary and sufficient conditions for the existence of $h_{l-1,l}$, $h_{l+1,l}$ for all possible l, as needed below.

This gives the following closed form expressions for the projections (4) and (5) by simply replacing $\int_c^\infty hdF_l$ in $A_l(h)$ by the expressions above. We have, for $j = 1, \ldots, k-1$,

$$\begin{aligned}\prod(A_j(h_j)|\overline{R(A_{j+1})}) &= A_{j+1}(h_{j+1,j}) \\ &= -\frac{\int_c^\infty h_j dF_j}{(S_{j+1}-S_j)^2\varphi_{j+1}} I(R=j+1) \\ &\quad + \frac{\int_c^\infty h_j dF_j}{(S_{j+2}-S_{j+1})(S_{j+1}-S_j)\varphi_{j+1}} I(R=j+2)\end{aligned} \tag{10}$$

and, for $j = 2, \ldots, k$,

$$\begin{aligned}\prod(A_j(h_j)|\overline{R(A_{j-1})}) &= A_{j-1}(h_{j-1,j}) \\ &= -\frac{\int_c^\infty h_j dF_j}{(S_j-S_{j-1})(S_{j-1}-S_{j-2})\varphi_{j-1}} I(R=j-1) \\ &\quad + \frac{\int_c^\infty h_j dF_j}{(S_j-S_{j-1})^2\varphi_{j-1}} I(R=j).\end{aligned} \tag{11}$$

For simplicity we derive the efficient score operators for the case $k = 3$. (The proof generalizes to the general case.) First, define

$$A_j^l = A_j - \prod(A_j|\overline{R(A_l)}).$$

The efficient score operators $A_j^*: L_0^2(F_j) \to L_0^2(P_F)$ are given by

$$A_3^*=A_3-\prod(A_3|\overline{R(A_1+A_2)})=A_3-\prod\left(A_3|\overline{R(A_2^1)}\right),$$
$$A_2^*=A_2-\prod(A_2|\overline{R(A_1+A_3)})=A_2-\prod(A_2|\overline{R(A_1)})-\prod(A_2\overline{R(A_3)}),$$
$$A_1^*=A_1-\prod(A_1|\overline{R(A_2+A_3)})=A_1-\prod\left(A_1|\overline{R(A_2^3)}\right).$$

Calculation of A_2^*. Applying (10) and (11) with j = 2 gives us

$$\prod(A_2(h_2)|\overline{R(A_1)})=-\frac{\int_c^\infty h_2dF_2}{(S_2-S_1)(S_1-S_0)\varphi_1}I(R=1)+\frac{\int_c^\infty h_2dF_2}{(S_2-S_1)^2\varphi_1}I(R=2)$$

and

$$\prod(A_2(h_2)|\overline{R(A_3)})=-\frac{\int_c^\infty h_2dF_2}{(S_3-S_2)^2\varphi_3}I(R=3)+\frac{\int_c^\infty h_2dF_2}{(S_4-S_3)(S_3-S_2)\varphi_3}I(R=4).$$

Thus,

$$A_2^*(h_2)=\frac{\int_c^\infty h_2dF_2}{(S_2-S_1)\varphi_1}I(R=I)+\left\{\frac{1}{S_2-S_1}-\frac{1}{(S_2-S_1)^2\varphi_1}\right\}\int_c^\infty h_2dF_2I(R=2)+\left\{\frac{1}{(S_3-S_2)^2\varphi_3}-\frac{1}{S_3-S_2}\right\}\int_c^\infty h_2dF_2I(R=3)-\frac{\int_c^\infty h_2dF_2}{(S_4-S_3)(S_3-S_2)\varphi_3}I(R=4).$$

Now, notice that

$$(S_2-S_1)S_1\varphi_1=S_2,$$
$$(S_4-S_3)(S_3-S_2)\varphi_3=S_4-S_2=F_2,$$
$$\frac{1}{(S_3-S_2)^2\varphi_3}-\frac{1}{S_3-S_2}=-\frac{1}{F_2},$$
$$\frac{1}{S_2-S_1}-\frac{1}{(S_2-S_1)^2\varphi_1}=\frac{1}{S_2}.$$

Thus (using $\int_0^\infty h_2dF_2=0$),

$$A_2^*(h_2) = \frac{\int_c^\infty h_2 dF_2}{S_2(c)} I(R \in \{1,2\}) + \frac{\int_0^c h_2 dF_2}{F_2(c)} I(R \in \{3,4\}).$$

Calculation of A_1^*. Formula (10) with $j = 1$ gives us

$$\prod(A_2(h_2)|\overline{R(A_3)}) = -\frac{\int_c^\infty h_2 dF_2}{(S_3 - S_2)^2 \varphi_3(c)} I(R=3) + \frac{\int_c^\infty h_2 dF_2}{(S_4 - S_3)(S_3 - S_2)\varphi_3(c)} I(R=4).$$

Thus,

$$\begin{aligned} A_2^3(h_2) = & A_2(h_2) - \prod(A_2(h_2)|\overline{R(A_3)}) \\ = & \frac{\int_c^\infty h_2 dF_2}{(S_2 - S_1)(c)} I(R=2) \\ & + \left\{\frac{1}{(S_3 - S_2)^2 \varphi_3(c)} - \frac{1}{(S_3 - S_2)(c)}\right\} \int_c^\infty h_2 dF_2 I(R=3) \\ & - \frac{\int_c^\infty h_2 dF_2}{(S_4 - S_3)(S_3 - S_2)\varphi_3(c)} I(R=4). \end{aligned}$$

We now note that

$$(S_4 - S_3)(S_3 - S_2)\varphi_3 = F_2, \quad \frac{1}{(S_3 - S_2)^2 \varphi_3(c)} - \frac{1}{(S_3 - S_2)(c)} = -\frac{1}{F_2}.$$

Thus,

$$A_2^3(h_2) = \frac{\int_c^\infty h_2 dF_2}{(S_2 - S_1)(c)} I(R=2) - \frac{\int_c^\infty h_2 dF_2}{F_2(c)} I(R \in \{3,4\}).$$

It is easily verified that the adjoint $A_2^{3\top} : L_0^2(P_F) \to L_0^2(F_2)$ is given by

$$A_2^{3\top}(V) = \int_0^{T_2} \left\{ V(c,2) - \frac{(S_3 - S_2)(c)}{F_2(c)} V(c,3) - \frac{F_3(c)}{F_2(c)} V(c,4) \right\} dG(c).$$

Subsequently, we can now verify that

$$A_2^{3\top} A_2^3(h_2) = \int_0^{T_2} \varphi_2^3(c) \int_c^\infty h_2 dF_2 dG(c),$$

where

$$\varphi_2^3 \frac{F_1}{F_2(S_2 - S_1)}.$$

We need to find $h_{23,1} \equiv (A_2^{3\top} A_2^3)^-(K)$ with

$$K = A_2^{3\top} A_1(h_1) = -\int_0^{T_2} \frac{\int_c^\infty h_1 dF_1}{(S_2 - S_1)(c)} dG(c).$$

This solution has to satisfy on $[0, \tau_2]$:

$$-\int_0^c h_{23,1} dF_2 = \frac{dK}{dG}(c) \frac{1}{\varphi_2^3(c)} = -\frac{F_2}{F_1}(c) \int_c^\infty h_1 dF_1$$

and, as shown previously, $h_{23,1}(\infty)$ is a simple function of $h_{23,1} I_{[0,\tau_2]}$. We note that $h_{23,1}$ exists under the assumption $F_j \equiv F_k$ (i.e., $F_j \ll F_k$ and $F_k \ll F_j$) on $[0, \tau_j]$, $j = 1, \ldots, k - 1$, which follows from (3). We conclude that

$$\prod(A_1(h_1) | \overline{R(A_2^3)}) = A_2^3(h_{23,1}) = -\frac{F_2}{F_1(S_2 - S_1)} \int_c^\infty h_1 dF_1 I(R=2) + \frac{\int_c^\infty h_1 dF_1}{F_1} I(R \in \{3, 4\}).$$

Using $F_2/(F_1(S_2 - S_1)) - 1/(S_2 - S_1) = -1/F_1$ and $\int_c^\infty h_1 dF_1 = -\int_0^c h_1 dF_1$ yields

$$A_1^*(h_1) = A_1(h_1) - \prod(A_1(h_1) | \overline{R(A_2^3)}) = \frac{\int_c^\infty h_1 dF_1}{S_1(c)} I(R=1) + \frac{\int_0^c h_1 dF_1}{F_1(c)} I(R \in \{2, 3, 4\}).$$

Calculation of A_3^*. This calculation is very similar to the one above for A_1^* and is omitted. We have

$$A_3^*(h_3)=\frac{\int_0^c h_3 dF_3}{F_3(c)} I(R=4)+\frac{\int_c^\infty h_3 dF_3}{S_3(c)} I(R\in\{1,2,3\}).$$

Proving that the tangent space is saturated—Given the expressions for the efficient score operators derived above, we now prove that the tangent space at a $P_{F,G}$ satisfying (3) is saturated. Under our assumption (3), the tangent space equals $L_0^2(G)$ (scores generated by G) plus the closure of the range of $A^*:L_0^2(F_1)\times\cdots\times L_0^2(F_k)\to L_0^2(P_F)$ defined by

$$(h_1,\ldots,h_k)\to A_1(h_1)+\cdots+A_k^*(h_k),$$

where the marginal efficient score operators are given by $A_j^*(h_j)=E(h_j(T_j))|(C, \Delta_j=I(T_j\le C)),\ j=1,\ldots,k$. The closure of the range of a Hilbert space operator equals the orthogonal complement of the null-space of its adjoint, that is, $\overline{R(A^*)}=N(A^{*\top})^{\perp}$. Thus we need to show that $N(A^{*\top})=L_0^2(G)$. The adjoint $A^{*\top}:L_0^2(P_F)\to L_0^2(F_1)\times\cdots\times L_0^2(F_k)$ is given by

$$A^{*\top}(V)=(A_1^{*\top}(V),\ldots,A_k^{*\top}(V)),$$

where it is easily verified that the adjoint $A_j^{*\top}:L_0^2(P_F)\to L_0^2(F_j)$ of A_j^* is given by

$$A_j^{*\top}(V)=E(E(V(C,R)|C,\Delta_j)|T_j).$$

Consider the operator $B_j^\top:L_0^2(C,\Delta_j)\to L_0^2(F_j)$ given by $B_j^\top(\eta)=E(\eta(C,\Delta_j)|T_j)$, where $L_0^2(C,\Delta_j)$ is the space of functions of (C, Δ_j) with finite variance and zero mean (both taken w.r.t. $P_{F,G}$). Using precisely the same proof as the proof of Lemma 2.2, it follows that, if F_j has a Lebesgue density $f_j > 0$ on $[0, \tau_j]$, then the null-space $N(B_j^\top)=L_0^2(G)$, that is, it consists of functions independent of Δ_j. Thus, under (3), $A_j^{*\top}(V)=0$ implies that $E(V(C, R) \mid C, \Delta_j) = E(V(C, R) \mid C) \equiv \varphi(C)$, $j = 1, \ldots, k$.

Setting $\Delta_1 = 0$ yields $\varphi(C) = E(V(C, R) \mid C, \Delta_1 = 0) = V(C, 1)$. Now, we note that

$$P(R=m|\Delta_j=1, C=c)=I(m \geq j+1)\frac{P(R=m|c)}{F_j(c)}, j=1, \ldots, k,$$

where $P(R = m \mid c) = (S_m - S_{m-1})(c)$. Thus, $E(V(C, R) \mid C, \Delta_j = 1)$ is given by

$$\sum_{m \geq j+1} V(c, m)\frac{(S_m - S_{m-1})(c)}{F_j(c)}=\varphi(c), j=1, \ldots, k.$$

For $j = k$, this equality gives $V(c, k + 1) = \varphi(c)$. For $j = k - 1$, this equality gives then

$$V(c, k)\frac{(S_k - S_{k-1})(c)}{F_{k-1}(c)} = \left(1 - \frac{F_k(c)}{F_{k-1}(c)}\right)\varphi(c) = \frac{(S_k - S_{k-1})(c)}{F_{k-1}(c)}\varphi(c)$$

so that $V(c, k) = \varphi(c)$. In this manner, we subsequently find $\varphi(c) = V(c, k + 1) = V(c, k) = \ldots = V(c, 2)$. This shows that $V(C, R)$ does not depend on R. This completes the proof.

3. Current status data on a counting process when final event is right censored

The following theorem proves efficiency of any regular asymptotically linear estimator at a specified rich sub-model.

Theorem 3.1

Let $N(t)$ be a counting process $N(t)=\sum_{j=1}^{k} I(T_j \leq t)$ for random variables $T_1 < \ldots < T_k$. Let C be a random censoring time. For every subject we observe the following data structure:

$$Y=(\tilde{T}=T_k \wedge C, \Delta=I(T_k \leq C), N(\tilde{T})).$$

We assume that C is independent of $(T_1, \ldots, T_k)$. The distribution of Y only depends on the multivariate distribution F of $(T_1, \ldots, T_k)$ through the marginal distributions $F_1, \ldots, F_k$ of $(T_1, \ldots, T_k)$.

Consider a data generating distribution $P_{F,G}$ in the model above satisfying the following condition (12): *For certain $\tau_1 < \ldots < \tau_k < \infty$, let F_j have Lebesgue density f_j on $[0, \tau_j]$ with*

$$\begin{array}{ll} f_j>0 & on\ [0,\tau_j]\ and\ f_j=0 \quad on\ (\tau_j,\infty),\ j=1,\ldots,k, \\ F_j>F_{j+1} & on\ (0,\tau_j],\ j=1,\ldots,k-1, \\ G & has\ Lebesgue\ density\ g. \end{array} \tag{12}$$

We allow that $p_j \equiv P(T_j = \infty) > 0$ *for* $j = j_0, \ldots, k$ and $j_0 \in \{1, \ldots, k\}$.

Then, the tangent space at $P_{F,G}$ *equals* $L_0^2(P_{F,G})$ *and is thus saturated. This implies that an estimator of a real valued parameter of the distribution F which is regular and asymptotically linear at* $P_{F,G}$ *is also asymptotically efficient if* $P_{F,G}$ *satisfies* (12). *In particular, if* $\bar{G}(t) > 0$ *and F, G satisfy* (12), *then the Kaplan–Meier estimator* $S_{k,KM}(t)$ *of* $S_k(t) = P(T_k > t)$, *based on the i.i.d. data* $(\tilde{T}, \Delta)$, *is asymptotically efficient.*

3.1. Regular and asymptotically linear estimators—The important implication of Theorem 3.1 is that, if we can construct an estimator of $\sqrt{n}$-estimable parameters of F_j which is regular, then this estimator will be asymptotically efficient at any F satisfying (12), $j = 1, \ldots, k$. In this subsection, we provide relatively simple regular and asymptotically linear estimators.

First, consider estimation of $S_k(t) = P(T_k > t)$. It is well known that $S_{k,KM}(t)$ is a regular asymptotically linear estimator of $S_k(t)$ whenever $\bar{G}(t) > 0$. Second, consider estimation of $S_j(t) = P(T_j > t)$, $j = 1, \ldots, k-1$. Let $\Delta_j \equiv I(T_j \le C)$. Under independent censoring (we can weaken this to noninformative censoring of T_k), we have

$$E(1-\Delta_j | C=c, T_k>c) = \frac{S_j(c)}{S_k(c)} \equiv R_j(c). \tag{13}$$

So

$$\begin{aligned} S_j(c) &= S_k(c) E(1-\Delta_j | C=c, T_k>c) \\ &= E(S_k(c)(1-\Delta_j) | C=c, T_k>c). \end{aligned} \tag{14}$$

In other words, estimating S_j can be viewed as estimating a monotonic regression of $S_k(C)(1-\Delta_j)$ on the observed C's. This suggests replacing S_k by the efficient Kaplan–Meier estimator $S_{k,KM}$ and minimizing

$$\frac{1}{n}\sum_{i=1}^{n} w_i \left\{ S_{k,KM}(C_i)(1-\Delta_{ji}) - S_j(C_i) \right\}^2 I(C_i \le T_{ki}) \tag{15}$$

over the vector $(S_j(C_i): i = 1, \ldots, n)$, under the constraint that S_j is monotone, where w_i, $i = 1, \ldots, n$, is a given set of weights possibly assigning more mass to observations with smaller variance. The solution $S_{j,n}$ of this problem can be obtained with the pool-adjacent-violator-algorithm (PAVA) [see, e.g., Barlow, Bartholomew, Bremner and Brunk (1972)].

A simple calculation shows that

$$\begin{aligned}&\mathrm{VAR}\{S_k(C)(1-\Delta_j)|C=c,T_k>c\}\\=&S_k(c)^2\mathrm{VAR}\left\{1-\Delta_j|C=c,T_k>c\right\}=S_k^2(c)R_j(c)\left\{1-R_j(c)\right\}.\end{aligned} \quad (16)$$

Since R_j is not identified from the data at a better rate than S_j, a good set of weights is $w_i=1/S^2_{k,KM}(C_i), i=1,\ldots,n$ [see van der Laan, Jewell and Peterson (1997)].

It is beyond the scope of this paper to prove that smooth functionals of $S_{j,n}$ are regular and asymptotically linear. Since it is straightforward to prove such a theorem for a standard histogram regression estimator of the regression of $S_k(C)(1-\Delta_j)$ on the observed C's, one expects that the more sophisticated isotonic regression estimate $S_{j,n}$ (which only differs because it selects its bins adaptively) is regular and asymptotically linear under the same conditions. We note that the choice of weights w_i, $i = 1, \ldots, n$, has no effect on the limit distribution of smooth functionals of $S_{j,n}$.

3.2. Proof of Theorem 3.1—In the first part of the proof we establish that, if condition (12) holds, then the efficient score operator of F_k equals the efficient score operator of F_k in the reduced data model for $(\tilde{T}_k, \Delta_k)$, hereby establishing a proof of the efficiency of the Kaplan–Meier estimator $S_{KM}(t)$. Subsequently, exploiting this special form of the efficient score operator of F_k, we prove saturation of the tangent space and thus Theorem 3.1.

Consider the data structure $(\tilde{T}_k = T_k \wedge C, N(\tilde{T}_k))$, where $N(t)=\sum_{j=1}^{k} I(T_j\leq t)$ and $T_1 < T_2 < \ldots < T_k$ are ordered random variables. Let $R = N(T_k) + 1$. The density of the data is given by

$$P(d\tilde{T}_k,R=j)=\prod_{m=1}^{k}(S_m-S_{m-1})(\tilde{T}_k)^{R=m}dF_k(\tilde{T}_k)^{R=k+1}dG(t)^{R\neq k+1}\overline{G}(t)^{R=k},$$

where $S_0 \equiv 0$ and $S_{k+1} \equiv 1$. We refer to the beginning of the proof of Theorem 2.1 to show that the tangent space at a $P_{F,G}$ satisfying condition (12) is the closure of the sum of the tangent spaces generated by F_j, $j = 1, \ldots, k$ and the tangent space of G, treating F_j as locally variation-independent. We have that the score operators: $A_j:L_0^2(F_j)\to L_0^2(P_{F,G})$ for F_j, $j= 1, \ldots, k-1$, are given by

$$A_j(h_j)=\frac{\int_c^\infty h_j dF_j}{(S_j-S_{j-1})(c)}I(R=j)-\frac{\int_c^\infty h_j dF_j}{(S_{j+1}-S_j)(c)}I(R=j+1)$$

and

$$A_k(h_k)=h_k(\tilde{T}_k)I(R=k+1)+\frac{\int_c^\infty h_k dF_k}{(S_k-S_{k-1})(c)}I(R=k).$$

Derivation of efficient score operator of F_k: We first determine the efficient score operator for F_k. For notational convenience, we consider the case $k = 3$. We have

$$A_3^*(h_3)=A_3(h_3) - \prod\left(A_3(h_3)|\overline{R(A_2^1)}\right)$$

where

$$A_2^1=A_2 - \prod(A_2|\overline{R(A_1)}).$$

Applying formula (11) gives

$$\prod(A_2(h_2)|\overline{R(A_1)})= - \frac{\int_c^\infty h_2 dF_2}{S_2(c)} I(R=1) + \frac{\int_c^\infty h_2 dF_2}{(S_2 - S_1)^2 \varphi_1(c)} I(R=2),$$

where we need to assume that $F_2 \ll F_1$ on $[0, \tau_1]$. Thus, an easy calculation shows that

$$A_2^1(h_2)=\frac{\int_c^\infty h_2 dF_2}{S_2(c)} I(R \in \{1,2\}) - \frac{\int_c^\infty h_2 dF_2}{(S_3 - S_2)(c)} I(R=3).$$

Another straightforward calculation shows that the adjoint $A_2^{1\top}:L_0^2(P_{F,G}) \to L_0^2(F_2)$ of $A_2^{1\top}:L_0^2(F_2) \to L_0^2(P_{F,G})$ is given by

$$A_2^{1\top}(V)=\int_0^{T_2}\left\{V(c,1)\frac{S_1}{S_2}(c)+\frac{(S_2 - S_1)}{S_2}(c)V(c,2) - V(c,3)\right\} dG(c).$$

A straightforward calculation now shows that

$$A_2^{1\top}A_2^1(h_2) = \int_0^{T_2}\int_0^\infty h_2 dF_2 \frac{S_3}{S_2(S_3 - S_2)}(c) dG(c).$$

We also have

$$A_2^{1\top}A_3(h_3) = -\int_0^{T_2} \frac{\int_c^\infty h_3 dF_3}{(S_3 - S_2)(c)} dG(c).$$

This shows that $h_{21,3} \equiv (A_2^{1\top}A_2^1)^{-}A_2^{1\top}A_3(h_3)$ satisfies, on $[0, \tau_2]$,

$$-\int_0^c h_{21,3} dF_2 = -\frac{S_2}{S_3}(c)\int_c^\infty h_3 dF_3,$$

and, if $p_2 = P(T_2 = \infty) > 0$, then $h_{21,3}(\infty)$ is a simple function of $h_{21,3}I_{[0,\tau_2]}$ as shown above (7). Here we need to assume that this equation can be solved in $h_{21,3}$. This is true if $F_3 \ll F_2$ on $[0, \tau_2]$. Then

$$\prod(A_3(h_3)|\overline{R(A_2^1)}) = A_2^1(h_{21,3}) = -\frac{\int_c^\infty h_3 dF_3}{S_3(c)} I(R \in \{1,2\}) + \frac{S_2(c)}{S_3(S_3-S_2)(c)}\int_c^\infty h_3 dF_3 I(R=3).$$

This proves that

$$A_3^*(h_3) = h_3(\tilde{T}_3)I(R=4) + \left\{\frac{1}{S_3-S_2} - \frac{S_2}{(S_3-S_2)S_3}\right\}(c)I(R=3) + \frac{\int_c^\infty h_3 dF_3}{S_3(c)} I(R \in \{1,2\}) = h_3(\tilde{T}_3)I(R=4) + \frac{\int_c^\infty h_3 dF_3}{S_3(c)} I(R \in \{1,2,3\}).$$

Thus, we have proved that, if $F_k \equiv F_j$ on $[0, \tau_j]$, $j = 1, \ldots, k-1$, then the efficient score $A_k^*(h_k) = E(h_k(T_k)|\tilde{T}_k, \Delta_k)$. The latter condition holds, in particular, if (12) holds. This proves the statement of Theorem 3.1 regarding efficiency of the Kaplan–Meier estimator S_{KM}.

Saturated tangent space result: Note that, for a random variable Y, we define $L_0^2(Y) = \{h(Y): Eh^2(Y) < \infty, Eh(Y) = 0\}$. For simplicity, we prove saturation for $k = 3$. Let $A: L_0^2(F_1) \times L_0^2(F_2) \to L_0^2(P_{F,G})$ be defined by $A(h_1, h_2) = A_1(h_1) + A_2(h_2)$. Then, the tangent space of F is given by $\overline{R(A_1)+R(A_2)+R(A_3)} = \overline{R(A_1)+R(A_2)} \oplus \overline{R(A_3^*)}$. Thus, the tangent space at $P_{F,G}$ is given by $\overline{R(A)} \oplus \overline{R(A_3^*)} \oplus \overline{R(B)}$, where $B: L_0^2(G) \to L_0^2(\tilde{T}_3, \Delta_3)$ is the score operator for the censoring mechanism G, given by $B(h) = E(h(C) \mid \tilde{T}_3, \Delta_3)$. By factorization of the likelihood into F and G parts, we have that $R(B)$ is orthogonal to F-scores. It is well known that $\overline{R(A_3^*) \oplus R(B)} = L_0^2(\tilde{T}_3, \Delta_3)$. The latter result simply states that the tangent space for the nonparametric right-censored data model for $(\tilde{T}_3, \Delta_3)$, only assuming that C is independent of T, is saturated [e.g., Bickel, Klaassen, Ritov and Wellner (1993)]. Thus, we need to prove that

$\overline{R(A)} \oplus L_0^2(\tilde{T}_3, \Delta_3) = L_0^2(P_{F,G})$ which is equivalent to proving $N(A^\top) = L_0^2(\tilde{T}_3, \Delta_3)$, where $A^\top : L_0^2 P_{(F,G)} \to L_0^2(F_1) \times L_0^2(F_2)$ is the adjoint of A and $N(A^\top)$ denotes its null space.

First, we decompose $A_1 + \ldots + A_{k-1}$ into a sum of orthogonal operators (efficient score operators in the model with F_k known). Let $A_1' = A_1 - \prod(A_1 | \overline{R(A_2)})$ and $A_2' = A_2 - \prod(A_2 | \overline{R(A_1)})$. By (4), it follows that

$$A_1'(h_1) = \frac{\int_c^\infty h_1 dF_1}{S_1(c)} I(R=1) - \frac{\int_c^\infty h_1 dF_1}{(S_3 - S_1)} I(R \in \{2,3\}),$$
$$A_2'(h_2) = \frac{\int_c^\infty h_2 dF_2}{S_2(c)} I(R \in \{1,2\}) + \frac{\int_0^c h_2 dF_2}{(S_3 - S_2)(c)} I(R=3),$$

where we need the equivalence assumptions $F_j \equiv F_{j+1}$ on $[0, \tau_j]$ for $j = 1, \ldots, k$, again. A more compact manner of representing these operators $A_j' : L_0^2(F_j) \to H(C, R) \equiv \{V(C,R) I(R<4) \in L_0^2(P_{F,G}) : V\}$ is

$$A_j'(h_j) = E(h_j(T_j) | C, \Delta_j, T_3 > C) I(T_3 > C), \; j = 1, 2. \tag{17}$$

Consider the operator $A' : L_0^2(F_1) \times L_0^2(F_2) \to H(C,R)$ defined by $A'(h_1, h_2) = A_1'(h_1) + A_2'(h_2)$. Proving $N(A^\top) = L_0^2(\tilde{T}_3, \Delta_3)$ is equivalent to proving $N(A'^\top) = L_0^2(\tilde{T}_3, \Delta_3)$, where $A'^\top$ is the adjoint of A'.

From the representation (17), the adjoint $A_j'^\top : H(C,R) \to L_0^2(F_j)$ is given by

$$A_j'^\top(V) = E(E(V(C,R) I(T_3 > C) | C, \Delta_j, T_3 > C) | T_j), \; j = 1, 2,$$

and thus, $N(A'^\top) = N(A_1'^\top) \cap N(A_2'^\top)$.

Consider now a solution $V I(T_3 > C) \in H(C, R)$ satisfying $A_j'^\top(V I(T_3 > C)) = 0, \; j = 1, 2$. In order to prove $V \in L_0^2(\tilde{T}_3, \Delta_3)$, it suffices to show $I(T_3 > C)V = I(T_3 > C)\varphi(C)$ for some φ. Using precisely the same proof as the proof of Lemma 2.2, it follows that, if F_j has a Lebesgue density $f_j > 0$ on $[0, \tau_j]$ and G has a Lebesgue density, then, for any function $I(T_3 > C)\eta(C, \Delta_j)$, $E(I(T_3 > C)\eta(C, \Delta_j) \mid T_j) = 0$ implies $\eta(C, 1) = \eta(C, 0)$. This proves that $E(V(C, R) I(T_3 > C) \mid C, \Delta_j, T_3 > C) = E(V(C, R) I(T_3 > C) \mid C, T_3 > C) \equiv I(T_3 > C)\varphi(C)$ does not depend on Δ_j, $j = 1, 2$.

Setting $\Delta_1 = 0$ yields $I(T_3 > C)\varphi(C) = E(V(C, R) I(T_3 > C) \mid C, \Delta_j, T_3 > C) = V(C, 1) I(T_3 > C)$. Now, we note that

$$P(R=m|\Delta_j=1, C=c, T_3>c)$$
$$=I(m \geq j+1, m<4)\frac{(S_m-S_{m-1})}{(S_3-S_j)(c)}, j=1, 2.$$

Thus, $E(V(C, R)I(T_3 > C) \mid C, \Delta_j = 1, T_3 > C)$ is given by

$$I(T_3>C)\sum_{m\geq j+1, m<4} V(C, m)\frac{(S_m - S_{m-1})(C)}{(S_3 - S_j)(C)} = I(T_3>C)\varphi(C), j=1, 2.$$

For $j = 2$, this equality gives $I(T_3 > C)V(C, 3) = I(T_3 > C)\varphi(C)$. For $j = 1$, this equality gives

$$I(T_3>C)\left\{V(C, 2)\frac{(S_2 - S_1)(c)}{(S_3 - S_1)(C)}+\varphi(C)\frac{(S_3 - S_2)(C)}{(S_3 - S_1)(C)}\right\} = I(T_3>C)\Phi(C),$$

so that $I(T_3 > C)V(C, 2) = I(T_3 > C)\varphi(C)$. We have shown $I(T_3 > C) \times V(C, 1) = I(T_3 > C) V(C, 2) = I(T_3 > C)V(C, 3)$ which proves that $V = I(T_3 < C)V_1(T_3) + I(T_3 > C)\varphi(C)$ for some functions V_1 and f, and thus that $V \in L_0^2(\tilde{T}_3, \Delta_3)$. This completes the proof. □

Acknowledgments

The authors thank the referees and Associate Editor for their helpful comments.

References

Barlow, RE.; Bartholomew, DJ.; Bremner, JM.; Brunk, HD. Statistical Inference under Order Restrictions. Wiley; New York: 1972.

Bickel, PJ.; Klaassen, CAJ.; Ritov, Y.; Wellner, JA. Efficient and Adaptive Estimation in Semi-Parametric Models. Johns Hopkins Univ. Press; 1993.

Diamond, ID.; McDonald, JW. The analysis of current status data. In: Trussell, J.; Hankinson, R.; Tilton, J., editors. Demographic Applications of Event History Analysis. Oxford Univ. Press; 1992. p. 231-252.

Diamond ID, McDonald JW, Shah IH. Proportional hazards models for current status data: Application to the study of differentials in age at weaning in Pakistan. Demography 1986;23:607–620. [PubMed: 3803657]

Dinse GE, Lagakos SW. Nonparametric estimation of lifetime and disease onset distributions from incomplete observations. Biometrics 1982;38:921–932. [PubMed: 7168795]

Gill, RD.; van der Laan, MJ.; Robins, JM. Proc First Seattle Symposium in Biostatistics. Lecture Notes in Statist. Vol. 123. Springer; New York: 1997. Coarsening at random: Characterizations, conjectures and counterexamples; p. 255-294.

Groeneboom, PJ. Special topics course 593C: Nonparametric estimation for inverse problems: algorithms and asymptotics. 1998. Technical Report 344, Dept. Statistics, Univ. Washington. (For related software see www.stat.washington.edu/jaw/RESEARCH/SOFTWARE/software.list.html.)

Groeneboom, P.; Wellner, JA. Information Bounds and Nonparametric Maximum Likelihood Estimation. Birkhäuser; Basel: 1992.

Huang J, Wellner JA. Asymptotic normality of the NPMLE of linear functionals for interval censored data, case I. Statist Neerlandica 1995;49:153–163.

Jewell NP, Malani HM, Vittinghoff E. Nonparametric estimation for a form of doubly censored data with application to two problems in AIDS. J Amer Statist Assoc 1994;89:7–18.

Jewell NP, Shiboski SC. Statistical analysis of HIV infectivity based on partner studies. Biometrics 1990;46:1133–1150. [PubMed: 2085629]

Jewell NP, van der Laan MJ. Generalizations of current status data with applications. Lifetime Data Analysis 1995;1:101–109. [PubMed: 9385086]

Jongbloed, G. Three statistical inverse problems. Delft Univ. Technology; 1995. Ph.D. dissertation

Keiding N. Age-specific incidence and prevalence: A statistical perspective (with discussion). J Roy Statist Soc Ser A 1991;154:371–412.

Kodell RL, Shaw GW, Johnson AM. Nonparametric joint estimators for disease resistance and survival functions in survival/sacrifice experiments. Biometrics 1982;38:43–58. [PubMed: 7082762]

Sun J, Kalbfleisch JD. The analysis of current status data on point processes. J Amer Statist Assoc 1993;88:1449–1454.

Turnbull BW, Mitchell TJ. Nonparametric estimation of the distribution of time to onset for specific diseases in survival/sacrifice experiments. Biometrics 1984;40:41–50. [PubMed: 6733233]

van der Laan MJ, Jewell NP, Peterson DR. Efficient estimation of the lifetime and disease onset distribution. Biometrika 1997;84:539–554.