Lifetime Data Analysis, 1, 417–434 (1995)

An Analytic Method for Randomized Trials with Informative Censoring: Part II

JAMES M. ROBINS
Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Ave., Boston, Massachusetts 02115

Received October 20, 1994; accepted August 8, 1995

Abstract. Consider a randomized trial in which time to the occurrence of a particular disease, say pneumocystic pneumonia in an AIDS trial or breast cancer in a mammographic screening trial, is the failure time of primary interest. Suppose that time to disease is subject to informative censoring by the minimum of time to death, loss to and end of follow-up. In such a trial, the potential censoring time is observed for all study subjects, including failures. In the presence of informative censoring, it is not possible to consistently estimate the effect of treatment on time to disease without imposing additional non-identifiable assumptions. Robins (1995) specified two non-identifiable assumptions that allow one to test for and estimate an effect of treatment on time to disease in the presence of informative censoring. The goal of this paper is to provide a class of consistent and reasonably efficient semiparametric tests and estimators for the treatment effect under these assumptions. The tests in our class, like standard weighted-log-rank tests, are asymptotically distribution-free α-level tests under the null hypothesis of no causal effect of treatment on time to disease whenever the censoring and failure distributions are conditionally independent given treatment arm. However, our tests remain asymptotically distribution-free α-level tests in the presence of informative censoring provided either of our assumptions are true. In contrast, a weighted log-rank test will be an α-level test in the presence of informative censoring only if (1) one of our two non-identifiable assumptions hold, and (2) the distribution of time to censoring is the same in the two treatment arms. We also study the estimation, in the presence of informative censoring, of the effect of treatment on the evolution over time of the mean of repeated measures outcome such as CD4 count.

Keywords: Survival analysis; Causal inference; Dependent censoring; Semiparametric methods

1. Introduction

Consider a two-armed randomized trial in which time to the occurrence of a particular disease, say pneumocystic pneumonia in an AIDS trial or breast cancer in a mammographic screening trial, is the outcome of primary interest. Suppose that time to disease is subject to informative censoring by the minimum of time to death, loss to and end of follow-up. Our goal is both to test the null hypothesis of no effect of treatment on time to disease, and to estimate the magnitude of the treatment effect.

Lin, Robins, and Wei (1995) generalize an approach of Robins (1989, p. 158), Robins and Tsiatis (1991), and Robins and Rotnitzky (1992, Appendix 4). In Lin et al.'s approach, an asymptotically normal and unbiased estimator of the treatment effect on time to disease was constructed under the (non-identifiable) assumption that, on a logarithmic scale, time to disease and time to informative censoring satisfied a bivariate location shift model with a completely arbitrary baseline distribution. Their key idea was to use the fact that the censoring time is known, even for subjects observed to develop disease.

Robins (1995) specified non-identifiable assumptions that are weaker than those made by

Lin et al. and yet allow one to test for and estimate an effect of treatment on time to disease in the presence of informative censoring. In this paper, we provide consistent and reasonably efficient semiparametric estimators for the treatment effect under these assumptions. Since, in contrast to Lin et al. (1995), our goal is not the joint estimation of the effect of treatment on time to informative censoring (e.g., death) and disease, we choose to estimate time to informative censoring (completely) non-parametrically. Hence, our inference concerning the effect of treatment on time to disease is not invalidated by possible misspecification of a model (such as Lin et al.'s location shift model) for the effect of treatment on time to informative censoring.

The paper is organized as follows. In Section 2, we describe the available data. In Section 3, we consider a class of tests for the treatment effect under these assumptions. In Section 4, we show that there exists a test and estimator in our class that attain the semiparametric bound for our model when censoring is in fact independent (but the independence of censoring is not imposed by the model).

The price we pay for the enhanced robustness of our test in the presence of informative censoring is a loss of power if censoring is, in truth, independent. Specifically, under independent censoring, the power of the optimal test in our class will be less than or equal to that of the optimal weighted log rank test against Pitman alternatives, with strict inequality unless the distribution of censoring times is the same in the two treatment arms.

We stress that, if censoring is informative and both of our two non-identifiable identifying assumptions are false, tests in our class will fail to provide asymptotically α-level tests of the null hypothesis of no treatment effect on time to disease. In fact, as discussed by Robins (1995), our test, in contrast to standard log-rank tests, need not be asymptotically α-level under the null hypothesis when censoring and failure are dependent but the censoring mechanism is non-informative. In Section 4, we argue that the "M-estimators" considered in this paper should be more efficient than the "rank" estimators similar to those of Lin et al. (1995).

In Section 5, we extend our methods and study the estimation, in the presence of informative censoring, of the effect of treatment on the evolution over time of the mean of repeated measures outcome such as CD4 count. Glidden and Wei (1995) adopt a different, presumably less efficient approach to the estimation of the effect of treatment on a repeated measures outcome in the presence of informative censoring. Section 6 considers variance estimation. Until Section 7, we regard all components of censoring as possibly informative. In Section 7, we consider the additional assumption that some component of the censoring mechanism (e.g., time to loss to follow-up) is *a priori* independent of both time to disease and the other informative components of censoring (e.g., time to death).

2. The Data

Let Z denote a random vector of regressors with discrete components, Y denote the logarithm of time to (informative) censoring, and X^0 denote the logarithm of time to disease. We assume we observe n independent and identically distributed realizations of

$$Z, X \equiv X^0 \wedge Y, \sigma \equiv I\left(X^0 < Y\right), Y = Y \tag{2.1}$$

That is, we always observe time to censoring Y, but we only observe time to disease X^0 if less than time to censoring. Without loss of generality we assume the 0 vector is in the support of Z. We restrict consideration to discrete Z so that the law of Y given Z can be estimated non-parametrically without additional smoothing. In a randomized study, Z will often simply be the dichotomous treatment arm indicator.

3. A Model and A Class of Estimators

3.1. *The Model*

Define

$$m(u, z) = F_{Y|Z}(u \mid z) = \Pr[Y < u \mid Z = z]. \tag{3.1}$$

Let $v(t, u, z)$ be the unique function satisfying

$$S_{X^0}(t \mid Z = z, M = u) = S_{X^0}[v(t, u, z) \mid Z = 0, M = u] \tag{3.2}$$

where $M \equiv m(Y, Z)$. In the bivariate shift model of Lin et al. (1995), $v(t, u, z) = t - \theta_0 z$, where θ_0 is the shift parameter.

Remark. The functions $m(u, z)$ and $v(t, u, z)$ are related to the functions $m(u)$ and $v(t, u)$ of Robins (1995) as follows. If $Z \in \{0, 1\}$ is a dichotomous treatment arm indicator of a randomized trial as in Robins (1995), then the function $m(u)$ of Robins (1995) equals $F_{Y|Z}^{-1}[m(u, 1) \mid 0]$, and the function $v(t, u)$ of Robins (1995) equals $v(t, F_{Y|Z}^{-1}(u \mid Z = 0), 1)$. It is more convenient to work with the functions $v(t, u, z)$ and $m(u, z)$ than $v(t, u)$ and $m(u)$. However, since $F_{Y|Z}^{-1}(\cdot \mid 0)$ is a fixed function, it follows that the causal identification results of Robins (1995) remain relevant. Adapting the notation of Robins (1995), if Robins' (1995) rank preservation assumption (2.9) holds, which we write as $F_{Y(1)}[Y(1)] = F_{Y(0)}[Y(0)]\ w.p.1$, then $v(t, u, 1) = S^{-1}_{X^0(0)|F_{Y(0)}(Y(0))=u}\{S_{X^0(1)|F_{Y(0)}(Y(0))=u}(t)\}$ measures the causal effect of treatment on disease X^0 among subjects with $F_{Y(0)}(Y(0)) = u$. Even without rank preservation, if $v(t, u, 1) = v^*(t)$ for all (t, u) then $v^*(t) = S^{-1}_{X^0(0)}\{S_{X^0(1)}(t)\}$ is the marginal causal effect of treatment on disease X^0. However, $v(t, u, 1)$ is only identified on $A(m, v) = \{(t, u); t < \min[m^{-1}(u, 1), v^{-1}(u, u, 1)]\}$. Further, even if $v(t, u, 1) = v^*(t)$ on $A(m, v)$, this does not imply $v(t, u, 1) = v^*(t)$ on all (t, u) and thus does not imply the identifiable function $v^*(t)$ on $A(m, v)$ has a causal interpretation.

Note definitions (3.1) and (3.2) imply

$$\begin{gathered} Z \coprod (\epsilon, M) \\ \text{where } \epsilon \equiv v\left(X^0, M, Z\right), M \equiv m(Y, Z) \end{gathered} \tag{3.3}$$

We shall assume we have a correctly specified model

$$v(t, u, Z) = \overset{\circ}{v}(t, u, Z, \theta_0) \tag{3.4}$$

where $\overset{\circ}{v}(t, u, Z, \theta)$ is a known function continuously differentiable in t, u, and θ satisfying $\overset{\circ}{v}(t, u, 0, \theta) = \overset{\circ}{v}(t, u, z, 0) = t$. θ_0 is an unknown parameter vector.

Example. The bivariate shift model of Lin et al. (1995) implies that $\overset{\circ}{v}(t, u, z, \theta_0) = t - \theta_0 z$ does not depend on u.

Our goal is to estimate θ_0 in the semiparametric model A defined by the data (2.1) and restriction (3.4). [Note restriction (3.3) follows from the definitions of ϵ and M and thus is not a restriction on the joint distribution of the data.]

Remark. Whether the identifiable parameter θ_0 has a causal interpretation depends on additional non-identifiable assumptions beyond (3.4). For example, if Z is randomized and either of the non-identifiable assumptions (2.8) or (2.9) of Robins (1995) is true, the sharp null hypothesis (2.1) of Robins (1995) of no causal effect of treatment Z on disease X^0 implies $\theta_0 = 0$ and an α-level test of $\theta_0 = 0$ is an α-level test of that null hypothesis.

3.2. *Estimation*

We next describe our estimators and their large sample properties. To do so, we shall require some additional notation. Let $\widehat{M} \equiv \widehat{m}(Y, Z) \equiv \widehat{F}_{Y|Z}(Y \mid Z)$ where

$$\widehat{F}_{Y|Z}(u \mid z) \equiv \sum_i I(Y_i < u) I(Z_i = z) / \sum_i I(Z_i = z)$$

is the empirical distribution of Y given Z and $i = 1, \ldots, n$ indexes the n study subjects. Let

$$\widehat{\epsilon}(\theta) = \overset{\circ}{v}\left(X^\circ, \widehat{M}, Z, \theta\right), \quad \widehat{\mu}(\theta) = \overset{\circ}{v}\left(Y, \widehat{M}, Z, \theta\right), \text{ and}$$

$$\widehat{\nu}(\theta) = \overset{\circ}{v}\left(X, \widehat{M}, Z, \theta\right) = \min\{\widehat{\epsilon}(\theta), \widehat{\mu}(\theta)\}.$$

When Z is a dichotomous (0, 1) variable, our estimating function has a simple form. For dichotomous Z, let $\widehat{\mu}^*(\theta) = \overset{\circ}{v}\left(\widehat{A}, \widehat{M}, 1 - Z, \theta\right)$ where $\widehat{A} \equiv \widehat{m}^{-1}\left(\widehat{M}, 1 - Z\right)$ and, by definition, $\widehat{m}^{-1}(u, z) \equiv t$ if $\widehat{m}(t, z) = u$. Define the observables

$$\widehat{X}^*(\theta) = \min\left(\widehat{\epsilon}(\theta), \widehat{\mu}^*(\theta), \widehat{\mu}(\theta)\right), \widehat{\Delta}(\theta) = I\left[\widehat{\epsilon}(\theta) = \widehat{X}^*(\theta)\right].$$

$\widehat{\Delta}(\theta)$ is an artificial failure indicator since $\widehat{\Delta}(\theta)$ may be 0 and yet $\sigma = 1$.

Our estimating function will depend on two functions of dim θ chosen by the investigator. Specifically, let

$$\widehat{R}(\theta) = r\left\{\widehat{\Delta}(\theta), \widehat{X}^*(\theta), \hat{M}\right\} \equiv$$
$$\widehat{\Delta}(\theta) r_1\left(\widehat{\epsilon}(\theta), \widehat{M}\right) + \left\{1 - \widehat{\Delta}(\theta)\right\} r_2\left(\hat{X}^*(\theta), \hat{M}\right)$$

where $r_1(\cdot, \cdot)$ and $r_2(\cdot, \cdot)$ are arbitrary vector valued functions of dim θ chosen by the investigator. Let $\widehat{D}(\theta) \equiv \widehat{D}(\theta, r) \equiv \widehat{R}(\theta)\left\{Z - \widetilde{E}(Z)\right\}$ where $\widetilde{E}(H) \equiv n^{-1} \sum_i H_i$ so

$\widetilde{E}[\cdot]$ is the empirical mean operator. Note $\widehat{\Delta}(\theta)$ and thus $\widehat{D}(\theta)$ are discontinuous in θ. We now adopt the following

Notational Convention. If $\widehat{G}(\theta) \equiv g\left(\theta, \widehat{m}(\cdot,\cdot), \widehat{m}^{-1}(\cdot,\cdot), \widetilde{E}[\cdot]\right)$, then

$$G(\theta) \equiv g\left(\theta, m(\cdot,\cdot), m^{-1}(\cdot,\cdot), E[\cdot]\right),$$

$\widehat{G} \equiv \widehat{G}(\theta_0)$, $G \equiv G(\theta_0)$.

Note, by (3.3), $E(D(\theta_0)) = 0$ since $\min(\mu, \mu^*) = \min_{z\in\{0,1\}} v\left\{m^{-1}(M,z), M, z\right\}$ and thus $\min(\epsilon, \mu, \mu^*)$ and R depend only on ϵ, M and not on Z. Here D and R are defined in terms of $\widehat{D}(\theta)$ and $\widehat{R}(\theta)$ by the above convention. In the Appendix, we prove

THEOREM 2.1 *In model A characterized by restriction (3.4) and data (2.1), there exists* $\hat{\theta}$ *minimizing* $\tilde{E}[\widehat{D}(\theta)]'\tilde{E}[\widehat{D}(\theta)]$ *in the neighborhood of* θ_0 *[so that* $\hat{\theta}$ *solves* $n^{\frac{1}{2}}\tilde{E}[\widehat{D}(\theta)] = o_p(1)$*], such that* $n^{\frac{1}{2}}(\hat{\theta} - \theta_0)$ *and* $n^{-\frac{1}{2}}\sum_i D_i(\theta_0)$ *are asymptotically normal with mean 0 and variances* $\tau^{-1}\Sigma\,\tau^{-1}$ *and* Σ *respectively, where*

$$\tau = \partial E[D(\theta)]/\partial\theta'\,|_{\theta=\theta_0} \text{ and } \Sigma = var(\alpha).$$

In Section 6, the random variable α is characterized for certain choices of $r_1(\cdot,\cdot)$ and $r_2(\cdot,\cdot)$, and a consistent estimator $\widehat{\Sigma}(\theta_0)$ of Σ is described. τ can be consistently estimated by the numerical derivative with respect to θ of $\widehat{D}(\theta)$ evaluated at $\widehat{\theta}$ with step size $O(n^{-\frac{1}{2}})$.

It follows that when θ_0 is one-dimensional, alpha-level tests of the null hypothesis $\theta_0 = 0$ can be based on comparing $\widehat{D}(0)/\{\widehat{\Sigma}(0)\}^{\frac{1}{2}}$ to a standard normal distribution.

3.3. *Vector Valued Z*

We now let Z be an arbitrary vector with discrete components. Let $h(u,z,x)$ and $\gamma(u,z,x)$ be fixed vector-valued functions of dim θ chosen by the investigator. Let

$$\begin{aligned}\widehat{D}(\theta) = \widehat{D}(\theta,h,\gamma) \equiv & \textstyle\int_{-\infty}^{\infty} dN_{\hat{\epsilon}(\theta)}(u)\left\{h\left(u,z,\widehat{M}\right) - \widehat{\mathcal{L}}\left(u,\theta,h,\widehat{M}\right)\right\} \\ & - \textstyle\int_{-\infty}^{\infty} du\, I\{\widehat{v}(\theta) > u\}\left\{\gamma\left(u,z,\widehat{M}\right) - \widehat{\mathcal{L}}\left(u,\theta,\gamma,\widehat{M}\right)\right\}\end{aligned}$$

where $N_{\widehat{\epsilon}(\theta)}(u) = I[\widehat{v}(\theta) \le u, \sigma = 1]$, $\widehat{\mathcal{L}}(u,\theta,h,x) \equiv$

$$\tilde{E}\left[I\left\{\mathring{v}\left[\widehat{m}^{-1}(x,Z),x,Z,\theta\right] \ge u\right\} h(u,Z,x)\right] / \widetilde{E}\left\{I\left[\mathring{v}\left(\widehat{m}^{-1}(x,Z),x,Z,\theta\right) \ge u\right]\right\}.$$

Remark. $\widehat{D}(\theta,h,\gamma)$ is the natural generalization of the estimating function $\widehat{D}(\theta,r)$ of the last subsection to non-dichotomous Z. Specifically, when Z is dichotomous for given functions h and γ, $\hat{D}(\theta,h,\gamma)$ equals $\hat{D}(\theta,r)$ with

$$\begin{aligned} r_2(u,x) &\equiv -\int_{-\infty}^{u}\{\gamma(t,1,x) - \gamma(t,0,x)\}\,dt, \\ r_1(u,x) &\equiv h(u,1,x) - h(u,0,x) + r_2(u,x). \end{aligned} \quad (3.5)$$

Again Eq. (3.3) implies $E[D(\theta_0)] = 0$ since $Z \coprod (\epsilon, M)$ implies

$$\begin{aligned} E[h(u,Z,M) \mid v \geq u, M = x] &= E[h(u,Z,M) \mid v \geq u, \epsilon = u, M = x] \\ &= E[h(u,Z,M) \mid \mu \geq u, M = x] \\ &= E\left[h(u,Z,x) \mid v\left[m^{-1}(x,Z), x, Z\right] \geq u, M = x\right] \\ &= \mathcal{L}(u,h,x). \end{aligned}$$

Thus $0 = E[dN_\epsilon(u)\{h(u,Z,M) - \mathcal{L}(u,h,M)\}] =$

$$E[I(v > u)\{\gamma(u,Z,M) - \mathcal{L}(u,\gamma,M)\}]$$

implying $E(D(\theta_0)) = 0$. Using this fact, we prove in Appendix 1, that Theorem 2.1 is true for a vector-valued Z.

Remark. $D(\theta_0)$ will in general not be a "martingale." However, if $\gamma(u,z,M) = h(u,z,M)\lambda_\varepsilon(u \mid M)$ where $\lambda_\varepsilon(u \mid M)$ is the hazard of ε at u given M, then $D(\theta_0)$ would be a martingale.

4. Efficiency Considerations

4.1. Semiparametric Efficiency

The likelihood for a single subject (with respect to some dominating measure) under the model A characterized by restriction (3.4) and data (2.1) is

$$\begin{aligned} &\{\partial m(Y,Z;\omega_1)/\partial Y\}\left\{\partial\epsilon(\theta)/\partial X^0\right\}^\sigma f(Z;\omega_3) f(Y \mid Z;\omega_1) \\ &\qquad \{f[\epsilon(\theta) \mid m(Y,Z;\omega_1);\omega_2]\}^\sigma \\ &\left\{\int_{\overset{\circ}{v}(Y,m(Y,Z;\omega_1),Z,\theta)}^{\infty} f[u \mid m(Y,Z;\omega_1);\omega_2]\,du\right\}^{1-\sigma} \end{aligned} \tag{4.1}$$

with $m(Y,Z;\omega_1) \equiv F(Y \mid Z;\omega_1)$ where $\omega = (\omega_1, \omega_2, \omega_3)$ is a vector of nuisance parameters with ω_1 and ω_2 infinite dimensional, with true value $\omega_0 = (\omega_{10}, \omega_{20}, \omega_{30})$ indexing the law of $Y \mid Z$, the law of $\epsilon \mid M$ with $M \equiv m(Y,Z)$, and the law of Z. $\{\partial m(Y,Z;\omega_1)/\partial Y\}$ is the Jacobian for the transformation of Y to $m(Y,Z;\omega_1)$ and $\partial\varepsilon(\theta)/\partial X^0$ is the Jacobian for the transformation of X^0 to $\varepsilon(\theta)$. The lower limit of integration in the integral in (4.1) reflects the fact that for a subject with X^0 unobserved, we only know that $\varepsilon(\theta)$ exceeds $\overset{\circ}{v}(Y, m(Y,Z;\omega_1), Z, \theta)$. Let $S = s(\epsilon, M, Z) = \partial \ell n\{[\partial\epsilon(\theta_0)/\partial X] f[\epsilon(\theta_0) \mid M;\omega_{20}]\}/\partial\theta$ be the score for θ if X^0 is always observed. Let $h_{op}(u,Z,M) = s(u,M,Z) - E[S \mid \epsilon > u, Z, M]$, and

$$\gamma_{op}(u,Z,M) \equiv \gamma^{h_{op}}(u,Z,M)$$

where $\gamma^h(u,Z,M) \equiv h(u,Z,M)\lambda_\epsilon(u \mid M)$ and $\lambda_\epsilon(u \mid \cdot)$ is the hazard of ϵ at u given $\cdot$. Then $E\left[D\left(\theta_0, h_{op}, \gamma_{op}\right) \mid Z\right] = 0$ and the martingale $D_{op}(\theta_0) \equiv D\left(\theta_0, h_{op}, \gamma_{op}\right)$ is the efficient score for θ_0 in our model A when $\omega_1 = \omega_{10}$ is also known *a priori* (Robins,1993, Lemma A4.1). Lemma A4.1 of Robins (1993) applies because when ω_{10} is known, $M = F_{Y|Z}(Y \mid Z)$ is known and (Y, Z) or, equivalently, (Z, M), are ancillary for θ in (4.1) (Robins and Rotnitzky, 1992, Appendix 4).

Example. Suppose $\epsilon(\theta_0) = X^0 - \theta_0 Z$, as in Lin et al. (1995). Then $h_{op}(u, z, M) = z\partial \ell n \lambda_\epsilon (u \mid M)/\partial u$.

Let $\tilde{\theta}(h_{op}, \gamma_{op})$ solve $n^{\frac{1}{2}}\tilde{E}\left[D\left(\theta, h_{op}, \gamma_{op}\right)\right] = o_p(1)$. $\tilde{\theta}(h_{op}, \gamma_{op})$ is semiparametric efficient in model A when ω_{10} is known, since it has asymptotic variance equal to the inverse of the variance of the efficient score $\{Var[D(h_{op}, \gamma_{op})]\}^{-1}$. Now suppose that, in truth, but not imposed by the model A, censoring is independent in the sense that $X^0 \coprod Y \mid Z$ which implies $\epsilon \coprod M$. Then $h_{op}(u, z, M)$, $\gamma_{op}(u, z, M)$, and $\overset{\circ}{v}(t, m, z, \theta_0)$ do not depend on M. Then, Corollary (6.2) below implies that $\widehat{\theta}(h_{op}, \gamma_{op})$ of Theorem (2.1) has the same asymptotic variance as $\tilde{\theta}(h_{op}, \gamma_{op})$ and so must be efficient in model A when censoring is independent whether or not ω_{10} is completely unknown (as we assume) or completely known.

If $\epsilon \coprod M$ is false and w_{10} is unknown, we would obtain an efficient estimator θ_0 in model A by solving $n^{\frac{1}{2}}\tilde{E}[\tilde{D}(\theta, h_{op}, \gamma_{op})] = o_p(1)$ where $\tilde{D}(\theta)$ is $\hat{D}(\theta)$ with efficient estimators, under model A, $\widetilde{m}(\cdot,\cdot)$ of $F_{Y|Z}(\cdot \mid \cdot)$ and $\widetilde{m}^{-1}(\cdot,\cdot)$ substituted for $\widehat{m}(\cdot,\cdot)$ and $\widehat{m}^{-1}(\cdot,\cdot)$ in $\hat{D}(\theta)$. Now the empirical distribution function $\widehat{m}(y, z)$ of $F_{Y|Z}(y \mid z)$ is not efficient in model A, since it fails to fully exploit the restriction that $\epsilon(\theta_0) \coprod Z \mid M$. To construct an efficient estimator $\widetilde{m}(y, z)$ is beyond the scope of this paper. However, if the probability that $X_0 < Y$ is small, then $\hat{\theta}(h_{op}, \gamma_{op})$ will be nearly efficient, since there will be little information regarding $F_{Y|Z}(y \mid z)$ available from the restriction $\epsilon(\theta_0) \coprod Z \mid M$.

Since γ_{op} and h_{op} are unknown, $\widehat{\theta}(h_{op}, \gamma_{op})$ is not available for data analysis. However, we can adaptively estimate them as follows. Choose a flexible parametric model $\lambda(u \mid M; \eta)$ for $\lambda_\epsilon(u \mid M)$. Obtain $\widehat{\lambda}_\epsilon\left(u \mid \hat{M}\right) \equiv \lambda\left(u \mid \widehat{M}; \widehat{\eta}\right)$ by finding $\widehat{\eta}$ that maximizes

$$\prod_i \lambda\left[\widehat{\epsilon}_i\left(\hat{\theta}\right) \mid \hat{M}_i; \eta\right]^\sigma \exp\left[-\int_{-\infty}^{\widehat{v}_i(\theta)} \lambda\left[t \mid \widehat{M}_i; \eta\right] dt\right]$$

with $\widehat{\theta} = \widehat{\theta}(h, \gamma)$ a preliminary estimate of θ_0. We then estimate $\widehat{h}_{op}$ and $\widehat{\gamma}_{op}$ from the estimated law of (Y, ϵ) given Z determined by $\widehat{\eta}, \widehat{\theta}(h, \gamma)$ and $\hat{F}_{Y|Z}(y \mid z)$. [Note since $\epsilon \coprod Z \mid M$ implies $\epsilon \coprod \mu \mid M$, we have that the conditional cause-specific hazard of v at u given M corresponding to $\sigma = 1$ equals $\lambda_\epsilon(u \mid M)$. Hence, $\lambda\left(u \mid \hat{M}; \widehat{\eta}\right)$ will be consistent for $\lambda_\epsilon(u \mid M)$ if the model $\lambda(u \mid M; \eta)$ is correctly specified.]

Alternatively, we can estimate $\lambda_\epsilon(u \mid M)$ by a smoothed estimate of the cause-specific hazard of v at u given M (corresponding to $\sigma = 1$) using Dabrowska's (1987) nearest neighbor estimator, applied to the "right censored" failure time data, $\widehat{M}, \widehat{v}(\hat{\theta}), \sigma$. The resulting estimator $\widehat{\lambda}(u \mid \hat{M})$ is, under regularity conditions, guaranteed to be consistent for $\lambda_\epsilon(u \mid M)$. If $\widehat{\lambda}(u \mid \hat{M})$ is based on Dabrowska's estimator, $\widehat{h}_{op}$ and $\widehat{\gamma}_{op}$ are guaranteed to converge in probability to h_{op} and γ_{op} under regularity conditions. This in turn is sufficient to guarantee, by a standard argument, as in Robins, Mark, and Newey (1992), that $\widehat{\theta}(\widehat{h}_{op}, \widehat{\gamma}_{op})$ and $\widehat{\theta}(h_{op}, \gamma_{op})$ have the same limiting distribution. [More generally, if $\widehat{\gamma}^h \equiv h(u, z, m)\widehat{\lambda}_\epsilon(u \mid m) \xrightarrow{P} \gamma^h \equiv h(u, z, m)\lambda_\epsilon(u \mid m)$, then $\widehat{\theta}(h, \widehat{\gamma}^h)$ and $\widehat{\theta}(h, \gamma^h)$ will be asymptotically equivalent.] By the reciprocal relationship between the local power of tests and efficiency of estimators, it follows that, for θ_0 one-dimensional, the test based on

$$\psi\left(0, \hat{h}_{op}, \widehat{\gamma}_{op}\right) \equiv n^{-\frac{1}{2}} \sum_i \widehat{D}_i\left(0, \hat{h}_{op}, \widehat{\gamma}_{op}\right) / \left\{\hat{\Sigma}\left(0, \hat{h}_{op}, \widehat{\gamma}_{op}\right)\right\}^{\frac{1}{2}}$$

will be the locally most powerfully alpha-level *regular* test (see Robins and Rotnitzky, 1992, p. 317, for the definition of a regular test) of $\theta_0 = 0$ against alternatives given by (3.4) when censoring is in fact independent, but independence is not imposed. We now show that this test has, under independent censoring, local power less that of the optimal weighted log rank tests (which is based on imposing independent censoring) except when the distribution of Y does not depend on Z, in which case the local powers are equal.

Results in Robins (1993, Appendix 4) and Ritov and Wellner (1988) imply that if we additionally correctly impose the assumption of independent censoring in model A, $h_{op}(u, Z, M) \equiv h_{op}(u, Z)$ does not depend on M and the efficient score is

$$\int_{-\infty}^{\infty} dM_\epsilon^*(u) \left\{h_{op}(u, Z) - E\left[h_{op}(u, Z) \mid \mu > u\right]\right\}$$

with $dM_\epsilon^*(u) = dN_\epsilon(u) - I(v > u)\lambda_\epsilon(u)du$. The efficient score $D\left(\theta_0, h_{op}, \gamma_{op}\right)$ when the assumption that censoring is independent is true but not imposed *a priori* is $\int_{-\infty}^{\infty} dM_\epsilon^*(u)\left\{h_{op}(u, Z) - E\left[h_{op}(u, Z) \mid \mu > u, M\right]\right\}$. It follows from the uniqueness of the efficient score in a semiparametric model, that we always gain information by imposing independent censoring *a priori* unless $E\left[h_{op}(u, Z) \mid \mu > u, M\right] = E\left[h_{op}(u, Z) \mid \mu > u\right]$, a sufficient and essentially necessary condition for which is $Z \coprod M \mid \mu > u$. When the null hypothesis $\theta_0 = 0$ is true, a sufficient and essentially necessary condition for the above conditional independence is that $m(y, z)$ does not depend on z, i.e. censoring is independent of treatment arm. To see this, note $\theta_0 = 0$ implies $\mu = Y \equiv m^{-1}(M, Z)$. Hence, since $Z \coprod M$ by Eq. (3.3), $Z \coprod M \mid m^{-1}(M, Z) > u$ will be true essentially only when $m^{-1}(y, z)$ does not depend on z. On the other hand, when $\theta_0 \neq 0$ it will always be the case that information is gained by imposing independent censoring (when true). The aforementioned result concerning testing then follows from the reciprocal relationship between the power of tests and the efficiency of estimators.

Remark. When $\epsilon \coprod M$ is false, $\widehat{\theta}\left(h_{op}, \gamma_{op}\right)$ is not even the most efficient estimator in our class $\widehat{\theta}(h, \gamma)$. However, the optimal choices h_{eff}, γ_{eff} are extremely complex. Since the efficient estimator with known M, $\widetilde{\theta}\left(h_{op}, \gamma_{op}\right)$ must be more efficient than $\widehat{\theta}\left(h_{eff}, \gamma_{eff}\right)$ and since we have argued that if $\Pr\left(X^0 < Y\right)$ is small, $\widehat{\theta}\left(h_{op}, \gamma_{op}\right)$ is nearly as efficient as $\widetilde{\theta}\left(h_{op}, \gamma_{op}\right)$, little information will be lost by failing to estimate h_{eff}, γ_{eff} and using $\hat{h}_{op}, \gamma_{op}$ instead.

4.2. *An Alternative Estimation Procedure*

In this section, we describe a simpler but less efficient method of estimation. We do so because of its simplicity and because it is connected to the test statistics described in Robins (1995). Let $\widehat{u}^{\dagger}(\theta) = \min_z v\{\hat{m}^{-1}(\widehat{M}, z), \widehat{M}, z, \theta\}$ so for Z dichotomous, $\widehat{u}^{\dagger}(\theta) = \min\{\widehat{\mu}(\theta), \widehat{\mu}^*(\theta)\}$. As for Z dichotomous, let $\widehat{X}^*(\theta) = \min\{\widehat{\varepsilon}(\theta), \widehat{\mu}^{\dagger}(\theta)\}$ and $\widehat{\Delta}(\theta) = I[\widehat{\varepsilon}(\theta) = \widehat{X}^*(\theta)]$. Define the subject-specific contribution to a weighted log-rank statistic based on a function $g(u, Z)$ of dim θ chosen by the investigator to be $\widehat{S}(\theta) \equiv \widehat{S}(\theta, g) = \int_{-\infty}^{\infty} dN^{\dagger}_{\widehat{\varepsilon}(\theta)}(u)\{g(u, Z) - \overline{g}(u)\}$ where $N^{\dagger}_{\widehat{\varepsilon}(\theta)}(u) = I[\widehat{X}^*(\theta) \le u, \widehat{\Delta}(\theta) = 1], \overline{g}(u) = \widetilde{E}[g(u, Z) I\{\widehat{X}^*(\theta) > u\}]/\widetilde{E}\left[I\left\{\widehat{X}^*(\theta) > u\right\}\right]$. Let $\widehat{\theta}^{\dagger} \equiv \widehat{\theta}^{\dagger}(g)$ solve $n^{\frac{1}{2}}\widetilde{E}\left[\widehat{S}(\theta, g)\right] = o_p(1)$ and $\widetilde{\theta}^{\dagger} \equiv \widetilde{\theta}^{\dagger}(g)$ solve $n^{\frac{1}{2}}\widetilde{E}[S(\theta, g)] = o_p(1)$. Now, by the standard theory of weighted log-rank tests, $E[S(\theta_0)] = 0$ if

$$\lambda_{X^*,\Delta=1}(u \mid Z) = \lambda_{X^*,\Delta=1}(u) \tag{4.2}$$

where $\lambda_{X^*,\Delta=1}(u \mid Z)$ is the cause-specific hazard of X^* corresponding to $\Delta = 1$. But (4.2) is true since $Z \coprod (\varepsilon, M)$ implies $Z \coprod (\varepsilon, \mu^{\dagger})$ and thus $Z \coprod (X^*, \Delta)$. When $\theta = 0$, $n\widetilde{E}[S(\theta)]$ and $n\widetilde{E}\left[\hat{S}(\theta)\right]$ are the test statistic numerators Num^* and $N\hat{u}m^*$ of Robins (1995). $\hat{\theta}^{\dagger}$ is closely related to the rank estimators of Lin et al. (1995).

When $F_{Y|Z}(y \mid z)$ is known so (M, Z) is ancillary for θ_0, one would expect that if $M \not\coprod \varepsilon$ then $\widetilde{\theta}^{\dagger}(g_{opt})$ (with g_{opt} the optimal choice of g) would have larger asymptotic variance than $\widetilde{\theta}(h_{opt}, \gamma_{opt})$ since the estimating function $S(\theta, g_{opt})$ in contrast to $D(\theta, h_{opt}, \gamma_{opt})$ does not have mean zero conditional on the ancillary statistic M. (See Robins (1993, p. 281) for related discussion.). We would therefore expect that $\widehat{\theta}(h_{opt}, \gamma_{opt})$ would be more efficient than $\widehat{\theta}^{\dagger}(g_{opt})$ when $F_{Y|Z}(y \mid z)$ is unknown, at least when $pr(X^0 < Y)$ is small. Thus, in this paper, we concentrate on the estimator $\widehat{\theta}(h_{opt}, \gamma_{opt})$.

5. Repeated Measures Outcomes

Suppose now our interest is in estimating the effect of treatment on a repeated measures outcome such as CD4 lymphocyte count X_t recorded at T predetermined times from randomization. Without loss of generality, denote the T times by $t = 1, \ldots, T$. We observe

$$Z_i, \bar{X}_i(Y_i), Y_i, i = 1, \ldots, n \tag{5.1}$$

where, now, $\bar{X}(u) \equiv (X_1, \ldots, X_{int(u)})'$ and $int(u)$ is the largest integer less than or equal to both u and T. With Y, Z, and $M = m(Y, Z) = F_{Y|Z}(Y \mid Z)$ as defined previously, define $v(t, u, z) = E[X_t \mid M = u, Z = z] - E[X_t \mid M = u, Z = 0]$. We consider the model

$$v(t, u, z) = \overset{\circ}{v}(t, u, z, \theta_0)$$

where $\overset{\circ}{v}(t, u, z, \theta)$ is a known function such that $\overset{\circ}{v}(t, u, 0, \theta) = \overset{\circ}{v}(t, u, z, 0) = 0$, which is equivalent to the model

$$E\left[\epsilon\left(\theta_0\right) \mid M, Z\right]=E\left[\epsilon\left(\theta_0\right) \mid M\right] \tag{5.2}$$

where $\epsilon(\theta) \equiv\left(\epsilon_1(\theta), \ldots, \epsilon_T(\theta)\right)^{\prime}, \epsilon_t(\theta)=X_t-\stackrel{\circ}{v}(t, M, Z, \theta)$.

Identifiability of Causal Effects. Results for the identifiability of causal effects wholly analogous to those obtained in Robins (1995) can be derived. Specifically, let $X_t(z)$ be the possibly counterfactual value of CD4 count if treatment z were given. If Z was assigned by physical randomization and Robins' (1995) rank preservation assumption (2.8) holds, then

$$v(t, u, z)=E\left[X_t(z)-X_t(0) \mid F_{Y(0)}\{Y(0)\}=u\right]$$

and thus has the causal interpretation as the mean effect of treatment z compared to treatment 0 amongst subjects with $F_{Y(0)}(Y(0))=u$. If Z is randomized, and $v(t, u, z)=v^*(t, z)$ for all t, z, u then $v^*(t, z)=E\left[X_t(z)-X_t(0)\right]$ is the marginal mean effect of treatment z. Note $v(t, u, 0)=0$ and that $v(t, u, z)$ is only identified on $A(m, v, z)= \left\{(t, u, z) ; t<\min \left[m^{-1}(u, z), v^{-1}(u, u, z)\right]\right\}$. Further, if $v(t, u, z)=v^*(t, z)$ on $A(m, v, z)$ does not imply that $v(t, u, z)=v^*(t, z)$ for all t, u, z and thus does not imply either that $E\left[X_t(z)-X_t(0)\right]=v^*(t, z)$ or that the identifiable values of $v^*(t, z)$ on $A(m, v, z)$ have any causal interpretation.

Estimation. Let

$$\hat{\mu}(\theta, \gamma)=\left(\hat{\mu}_1\left(\theta, \gamma_1\right) \ldots, \hat{\mu}_T\left(\theta, \gamma_T\right)\right)^{\prime}, \hat{\mu}_t\left(\theta, \gamma_t\right)=I(Y>t)\left[\hat{\epsilon}_t(\theta)-\gamma_t\left(\hat{M}\right)\right],$$

$\widehat{\epsilon}_t(\theta)=X_t-\stackrel{\circ}{v}\left(t, \widehat{M}, Z, \theta\right)$, and $\gamma_t(\cdot)$ is any smooth real valued function. Define $\widehat{D}(\theta) \equiv \widehat{D}(\theta, h, \gamma) \equiv \sum_{t=1}^T \widehat{D}_t(\theta)$ where

$$\gamma=\left(\gamma_1, \ldots, \gamma_T\right)^{\prime}, \quad \widehat{D}_t(\theta)=\widehat{D}_t\left(\theta, h_t, \gamma_t\right)=\left\{h_t\left(Z, \widehat{M}\right)-\widehat{\mathcal{L}}\left(h_t, \widehat{M}\right)\right\} \widehat{u}_t\left(\theta, \gamma_t\right),$$

$$h \equiv h(z, u)=\left(h_1(z, u), \ldots, h_T(z, u)\right)^{\prime},$$

$h_t(z, u)$ is dim θ-vector-valued function and $\widehat{\mathcal{L}}(h_t, u)=\widetilde{E}\left[I\left\{\widehat{m}^{-1}(u, Z)>t\right\} h_t(Z, u)\right] / \widetilde{E}\left[I\left\{\widehat{m}^{-1}(u, Z)>t\right\}\right]$. We then prove in Appendix 1,

THEOREM 5.1 *For the semiparametric model B defined by restriction (5.2) and data (5.1), there exists $\widehat{\theta}=\widehat{\theta}(h, \gamma)$ solving $n^{-\frac{1}{2}} \sum_i \hat{D}_i(\theta)=0$ such that $n^{-\frac{1}{2}}\left(\widehat{\theta}-\theta_0\right)$ is asymptotically normal with mean 0 and variance $\tau^{-1} \Sigma \tau^{-1}$ where $\tau=E\left[\partial D\left(\theta_0\right) / \partial \theta^{\prime}\right]$ and $\Sigma=var(\alpha)$ with α defined in Sec. 6.*

The key step in proving Theorem 5.1 is to show that $E\left[D_t\right]=0$. To do so, we require the following identities.

$E\left[\varepsilon_t \mid M, Z, I(Y>t)\right]=E\left[\varepsilon_t \mid M, Z, I\left\{m^{-1}(M, Z)>t\right\}\right]=E\left[\varepsilon_t \mid M, Z\right]=E[\varepsilon_t \mid M]$ where the last equality is by (5.2). Also $\mathcal{L}\left(h_t, u\right)=E\left[h_t(Z, u) I\left\{m^{-1}(u, Z)>t\right\}\right] / E\left[I\left\{m^{-1}(u, Z)>t\right\}\right]=E\left[h_t(Z, u) \mid m^{-1}(u, Z)>t\right]=E\left[h_t(Z, u) \mid m^{-1}(u, Z)>\right.$

$t, M]$ where the last equality is by $Z \coprod M$. Hence, $\mathcal{L}(h_t, M) = E[h_t(Z, M) \mid m^{-1}(M, Z) > t, M] = E[h_t(Z, M) \mid Y > t, M]$.

Note $D_t = 0$ if $Y < t$ so to show $E[D_t] = 0$ it suffices to prove $E[D_t \mid Y > t, M] = 0$. Now

$$\begin{aligned} E[D_t \mid Y > t, M] &= E\{E[D_t \mid Y > t, M, Z] \mid Y > t, M\} \\ &= E\{I(Y > t)\{E(\varepsilon_t \mid M) - \gamma_t(M)\} \\ &\quad \{h_t(Z, M) - E[h_t(Z, M) \mid Y > t, M] \mid Y > t, M\}\} \\ &= 0. \end{aligned}$$

The asymptotic variance of $\widetilde{\theta}(h_{op}, \gamma_{op})$ satisfying $n^{\frac{1}{2}}\tilde{E}[D(\theta, h_{op}, \gamma_{op})] = 0$ with $\gamma_{op,t} = E[\epsilon_t \mid M]$ and $h_{op}(Z, M) = T_{op}\{Q^*_{op} - E[T_{op} \mid M]^{-1}E[T_{op}Q^*_{op} \mid M]\}$ with $T_{op} = E[\mu(\gamma_{op})^{\otimes 2} \mid Z, M]^{-1}$ and $Q^*_{op} = E[\partial\mu(\theta_0, \gamma_{op})/\partial\theta' \mid Z, M]$ attains the efficiency bound for the model (5.2) with $F_{Y|Z}(y \mid z) \equiv m(y, z)$ known. This characterization of the efficient estimator follows from the characterization of the efficient score for the semiparametric regression model given by Chamberlain (1988). Chamberlain's result applies because, when $m(y, z)$ is known, (Y, Z) is ancillary for θ. Corollary (6.1) below implies $\widehat{\theta}(h_{op}, \gamma_{op})$ and $\widetilde{\theta}(h_{op}, \gamma_{op})$ are asymptotically equivalent when $\epsilon \coprod M$ (i.e., under independent censoring) and so $\widehat{\theta}(h_{op}, \gamma_{op})$ is efficient whether or not $m(y, z)$ is known. Even when $\epsilon \coprod M$ is false (and $m(y, z)$ is unknown), we would recommend using $\widehat{\theta}(\hat{h}_{op}, \hat{\gamma}_{op})$ in applications where $\hat{h}_{op}$ and $\hat{\gamma}_{op}$ are consistent for h_{op} and γ_{op} even though $\widehat{\theta}(h_{op}, \gamma_{op})$ is not as efficient as the optimal estimator in our class, $\widehat{\theta}(h_{eff}, \gamma_{eff})$ say, and $\widehat{\theta}(h_{eff}, \gamma_{eff})$ itself is not semiparametric efficient. This recommendation reflects the fact that computing either h_{eff} and γ_{eff} or the semiparametric efficient estimator is very complex.

Given a preliminary estimate $\widehat{\theta} = \widehat{\theta}(h, \gamma)$, we can, under regularity conditions, consistently estimate $\gamma_{op,t}(\cdot)$ by the kernel regression of $\widehat{\epsilon}(\widehat{\theta})$ on $\widehat{M}$ among subjects with $Y > t$, since $E(\epsilon_t \mid M) = E(\epsilon_t \mid M, Y > t)$ by (5.2). Then arguing as in Robins, Mark, and Newey (1992), $\widehat{\theta}(h, \hat{\gamma}_{op})$ and $\widehat{\theta}(h, \gamma_{op})$ will be asymptotically equivalent. Even when the law of Y given Z is known, the Glidden-Wei (1995) estimator, like the Lin et al. (1995) estimator in the survival context and in contrast to $\hat{\theta}(h, \gamma_{op})$, is not asymptotically unbiased conditional on the ancillary statistic (M, Z).

6. Variance Estimation

To estimate the variance of $n^{\frac{1}{2}}\left\{\widehat{\theta}(h, \gamma) - \theta_0\right\}$ in Theorems (2.1) and (5.1), we need to first derive $\alpha \equiv \alpha(h, \gamma)$. In general, an explicit analytic formula for α is difficult to derive except when (a) $\gamma = \gamma^h$ in model A, (b) $\gamma = \gamma_{op}$ in model B, or (c) Z is dichotomous.

We first consider Theorem (5.1).

LEMMA 6.1 *If, in Theorem (5.1),* $\gamma = \gamma_{op} \equiv E[\varepsilon \mid M]$, *then* $\alpha = \alpha_1 + \alpha_2$, $\alpha_1 = D$, $\alpha_2 = \sum_{t=1}^{T} \alpha_{t2}$, $\alpha_{tk} \equiv \alpha_{tk}(Y, Z)$,

$$\begin{aligned} \alpha_{t2} &\equiv \alpha_{t2}(y, z) \\ &= E[\{h_t(Z, M) - \mathcal{L}(h_t, M)\}\left\{v_{[2]}(t, M, Z) - \gamma_{t[1]}(M)\right\}I(Y > t)a(Y, Z, y) \mid Z = z \end{aligned}$$

$a(t, z, y) \equiv I(y < t) - F_{Y|Z}(t \mid z) \equiv I(y < t) - m(t, z)$

and for any function $\ell(\cdot, \cdot, \cdot)$, $\ell_{[j]}$ *is the derivative with respect to its* j^{th} *argument. If* $\gamma \neq \gamma_{op}$, $\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5$, $\alpha_5 = E(D \mid Z)$, $\alpha_4 = \sum_t \alpha_{t4}(Y, Z)$

$$\alpha_{t4}(y, z) = E\left[\left\{h_{t[2]}(Z, M) - \mathcal{L}\left(h_{t[2]}, M\right)\right\}\left\{\gamma_{op,t}(M) - \gamma_t(M)\right\} I(Y > t) a(Y, Z, y) \mid Z = z\right] \quad (6.1)$$

and $\alpha_3 = \sum_t \alpha_{t3}(Y, Z)$. *When Z is dichotomous,* $\alpha_{t3}(Y, Z)$ *is as follows.*

Define $z^*(t) = 0$ *if* $m(t, 0) \geq m(t, 1)$ *and* $z^*(t) = 1$ *otherwise,* $k^*(t) = m(t, 1 - z^*(t))$, $q_t^*(u) = E[h_t(Z, u)]$, $\gamma_{op,t}(u) = E[\epsilon_t \mid M = u]$. *Then* $\alpha_{t3}(Y, Z) = \alpha_{t3}^0 \alpha_{t3}^1(Y, Z)$ *where*

$$\alpha_{t3}^0 \equiv F_Z\left[z^*(t)\right]\left\{h_t\left(z^*(t), k^*(t)\right) - q_t^*\left(k^*(t)\right)\right\}\left\{\gamma_{op,t}\left[k^*(t)\right] - \gamma_t\left[k^*(t)\right]\right\},$$

$$\alpha_{t3}^1(Y, Z) = \left\{\begin{array}{c} [I(Y < t) - k^*(t)]\, I[Z = 1 - z^*(t)] / f_Z\{1 - z^*(t)\} \\ -\left\{I\left[Y < m^{-1}\{k^*(t), z^*(t)\}\right] - k^*(t)\right\} I(Z = z^*(t)) / f_Z\{z^*(t)\} \end{array}\right\}$$

Remark. Let $\widehat{\alpha}\left(\widehat{\theta}\right)$ be an estimate of α obtained by taking appropriate sample averages (possibly within levels of $Z = z$) given estimates $\widehat{m}(y, z)$ and $\widehat{\theta}$ and, if necessary, a consistent (smoothed) estimate of $\gamma_{op}(u)$. The estimator $\widehat{\Sigma}\left(\widehat{\theta}\right) = \widetilde{E}\left[\widehat{\alpha}\left(\widehat{\theta}\right)\widehat{\alpha}'\left(\widehat{\theta}\right)\right]$ is consistent for $\Sigma = var(\alpha)$ if $\widehat{\theta}$ is consistent for θ_0.

When $\gamma = \gamma_{op}$, $\alpha_3 = \alpha_4 = \alpha_5 = 0$. The quantities α_2, α_3, and α_4 are contributions attributable to having to estimate the law of Y given Z. All components of α, with the exception of α_{t3}, can be consistently estimated by sample averages given estimates $\widehat{M}$ and $\widehat{\theta}(h, \gamma)$. However, due to the presence of $\gamma_{op,t}\{k^*(t)\}$ in α_{t3}, we must estimate $E[\epsilon_t \mid M = u]$ by smoothing. When Z is not dichotomous, the formula for α_{t3} becomes exceedingly complex. (We do not derive it.)

Since, even for Z dichotomous, if $\gamma \neq \gamma_{op}$, we need to consistently estimate $\gamma_{op}(u)$ by smoothing to obtain a consistent estimator of α_3 and thus of α and Σ, we cannot avoid smoothing. This suggests that we use the adaptive estimate $\widehat{\theta}\left(h, \widehat{\gamma}_{op}\right)$ [which is asymptotically equivalent to $\widehat{\theta}\left(h, \gamma_{op}\right)$ when $\widehat{\gamma}_{op}$ is consistent for γ_{op}]. Then $\alpha = \alpha_1 + \alpha_2$ can be estimated by sample averages for arbitrary discrete Z.

Alternatively, since the variance of $\widehat{\theta}(h, \gamma)$ can be consistently estimated by the bootstrap even when $\gamma \neq \gamma^{op}$, one could use the bootstrap to avoid smoothing in the Z dichotomous case and avoid deriving α_{t3} in the setting of an arbitrary Z.

A corresponding result applicable to deriving α in Theorem (2.1) is given in the following Lemma.

LEMMA 6.2 *If* $\gamma = \gamma^h$ *in Theorem 2.1, then* $\alpha = \alpha_1 + \alpha_2$ *where* $\alpha_1 = D$, *and, with* $a(Y, Z, y)$ *is as in Lemma 6.1,*

$$\alpha_2 \equiv \alpha_2(Y, Z), \quad \alpha_2(y, z) = E[H\, a(Y, Z, y) \mid Z = z],$$

$$\begin{aligned} H &= v_{[2]}(X, M, Z)\left[\left\{h_{[1]}(\epsilon, Z, M) - \mathcal{L}\left(\epsilon, h_{[1]}, M\right)\right\} I(X < Y)\right. \\ &\quad - \lambda_\epsilon(v \mid M)\{h(v, Z, M) - \mathcal{L}(v, h, M)\}] \\ &\quad - \int_{-\infty}^{\infty} du I(v > u)\{\partial\lambda_\epsilon(u \mid M)/\partial M\}\{h(u, Z, M) - \mathcal{L}(u, h, M)\}. \quad (6.2) \end{aligned}$$

When $\widehat{\gamma}^h \xrightarrow{P} \gamma^h$, a consistent estimator $\widehat{\Sigma}\left(\widehat{\theta}\right)$ of Σ can be constructed as described in the Remark following Lemma 6.1. We now provide a corollary that reinterprets Lemma 6.2 for the special case of Z dichotomous.

COROLLARY 6.1 *If, in model A, Z is dichotomous, then, given* $r_2(u, x)$*, if we choose* $r_1(u, M) = -r_{2[1]}(u, M)\lambda_\epsilon(u \mid M)^{-1} + r_2(u, M)$*, then* $\alpha = \alpha_1 + \alpha_2$ *with* α_2 *as in Lemma 6.2 except now*

$$H \equiv \Delta r_{1[1]}(X^*, M) + (1 - \Delta) I(X^* = \mu) r_{2[1]}(X^*, M)$$
$$- \textstyle\int_{-\infty}^{X^*} du\, r_{2[1]}(u, M)\lambda_\epsilon(u \mid M)^{-1}\{\partial\lambda_\epsilon(u \mid M)/\partial M\} du$$

The following corollaries imply that no information is lost by having to estimate $m(y, z)$ when $\epsilon \coprod M$.

COROLLARY 6.2 *If, in model B,* $\epsilon \coprod M$ *and* $\gamma = \gamma_{op}$ *then* $\alpha = D$.

Proof: If $\epsilon \coprod M$, $v(t, u, z)$ and $\gamma_{op}(u)$ do not depend on u, so $\alpha_2 = 0$. ■

COROLLARY 6.3 *If, in model A,* $\epsilon \coprod M$ *and* $\gamma = \gamma^h$ *then* $\alpha = D$.

Proof: If $\epsilon \coprod M$, $v(X, M, Z)$ and $\lambda_\epsilon(u \mid M)$ do not depend on M so $\alpha_2 = 0$. ■

7. Independent Censoring

The two non-identifiable assumptions, Eqs. (2.8) and (2.9) of Robins (1995), that allow one to interpret θ_0 causally in a randomized trial, are "non-reproductive" in the following sense. Suppose censoring by variable Y, say time to death, is informative but that one or both of our non-identifiable assumptions are true for censoring by Y. Assume, however, we can only observe the minimum of time to loss to follow-up Q and Y, and that Q is independent of all other random variables. Even so, censoring by the minimum of Q and Y, say Y^*, will not in general satisfy either of our non-identifiable assumptions. As a consequence, it is important to develop methods that allow us to test for and estimate an effect of treatment on time to disease when one of our two non-identifiable assumptions holds for censoring by a cause Y but Y is subject to additional independent censoring by another variable Q. We now develop such methods using the general results of Robins and Rotnitzky (1992) on adjustment for non-informative censoring. Suppose we observe

$$Z, Y^\dagger = \min(Q, Y), X^\dagger = \min(X, Q), \tau = I\left(Y^\dagger = Y\right) \quad (7.1)$$

rather than the data (2.1). That is, the data (2.1) is subject to additional censoring by the random variable Q. If censoring by Q is non-informative in the sense that

$$\lambda_Q\left[u \mid Y > u, Z, X, Y\right] = \lambda_Q\left[u \mid Z, Y > u, \min(X, u)\right] \tag{7.2}$$

i.e., the cause-specific hazard of $Y^\dagger$ corresponding to Q (i.e. $\tau = 0$) given Z, X, Y only depends on information available up to u, then we can use methods of Robins and Rotnitzky to estimate θ_0 of model (3.4). Specifically, let

$$K(t) = \exp\left[-\int_0^t \lambda_Q(u \mid Z, Y > u, \min(X, u))\right]$$

be the probability of remaining uncensored by Q up to t given $Z, X, Y, Y > t$. Let $\widehat{K}(t)$ be an $n^{\frac{1}{2}}$-consistent estimate of $K(t)$ based on substituting an estimate

$$\widehat{\lambda}_Q(u \mid Z, Y > u, \min(X, u))$$

into the expression for $K(t)$ based on some model (e.g., a proportional hazards model). For example, suppose we assume that $\lambda_Q[u \mid Z, Y > u, \min(X, u)] = \lambda_Q[u \mid Z, Y > u]$ does not depend on $\min(X, u)$. Then $\widehat{\lambda}_Q(u \mid Z, Y > u)$ will be the Nelson-Aalen estimator of $\lambda_Q(u \mid Z, Y > u)$ computed within stratum defined by Z. Results of Robins and Rotnitzky (1992) imply that under regularity conditions, $\widehat{\theta}_q$ solving $n^{\frac{1}{2}}\widetilde{E}[\widehat{D}(\theta)\tau/\widehat{K}(Y)] = o_p(1)$ will be a consistent asymptotically normal estimator of θ_0. The key idea is that each individual for whom Y is observed (uncensored by Q) contributes a weight $1/\widehat{K}(Y)$ inversely proportional to the probability that they would be uncensored. Results of Robihs and Rotnitzky (1992) show how to obtain an expression for and estimator of the asymptotic variance of $\widehat{\theta}_q$. However, the variance estimator will be quite complex, and bootstrap estimates of variability may be preferred.

A similar approach may be used to obtain consistent asymptotically normal of the parameter θ_0 of model (5.2) when the available data are

$$Z, Y^\dagger = \min(Q, Y), \tau = I\left(Y^\dagger = Y\right), \overline{X}\left(Y^\dagger\right) \tag{7.3}$$

rather than data (5.1) provided censoring by Q is non-informative in the sense that

$$\lambda_Q\left[u \mid Z, Y, Y > u, \overline{X}(Y)\right] = \lambda_Q\left[u \mid Z, Y > u, \overline{X}(u)\right]. \tag{7.4}$$

Then $\widehat{\theta}_q$ solving $n^{\frac{1}{2}}\widetilde{E}[\widehat{D}(\theta)\tau/\widehat{K}(Y)] = o_p(1)$ where $\widehat{D}(\theta)$ is as defined in Sec. 5 and now $\widehat{K}(Y)$ is an $n^{\frac{1}{2}}$-consistent estimator of $K(Y) = \exp[-\int_0^Y \lambda_Q\left(u \mid Z, Y > u, \overline{X}(u)\right)]$.

Appendix

Sketch of a proof of Theorem 5.1 and Lemma 6.1: Since Eq. (5.2) implies $E[D(\theta_0)] = 0$, it follows that under standard regularity conditions described by Newey (1993) and Pakes and Pollard (1989), Theorem (5.1) follows from $n^{\frac{1}{2}}\widetilde{E}\left[\hat{D}(\theta_0)\right] = n^{\frac{1}{2}}\widetilde{E}[\alpha] + o_p(1)$ where

Newey (1993) uses the theory of pathwise derivatives to show that an explicit formula for α is obtained as follows. Denote the probability limit of $\widetilde{E}\left[\hat{D}(\theta_0)\right]$ under a general distribution F^* for $\left(Y, \bar{X}(Y), Z\right)$ by

$$\begin{aligned} D^*(F^*) &= \sum_{t=1}^{T} D_t^*(F^*), \\ D_t^*(F^*) &= E_{F^*}\left[\left\{h_t\left[Z, M\left(F^*\right)\right] - \mathcal{L}\left(h_t, M\left(F^*\right), M(F^*), F^*, F^*\right)\right\}\right. \\ &\quad \times \left.\left[\epsilon_t\left\{M\left(F^*\right)\right\} - \gamma_t\left\{M\left(F^*\right)\right\}\right] I(Y > t)\right] \end{aligned}$$

where $M\left(F^*\right) = m(Y, Z; F^*) = F^*_{Y|Z}(Y \mid Z)$,

now $\epsilon_t(u) = X_t - \overset{\circ}{v}(t, u, Z; \theta_0)$,

$$\mathcal{L}\left(h_t, u_1, u_2, F_1^*, F_2^*\right) = E_{F_2^*}\left\{I\left[m^{-1}\left(u_1, Z; F_1^*\right) > t\right] h_t(Z, u_2)\right\} / E_{F_2^*}\left\{I\left[m^{-1}\left(u_1, Z; F_1^*\right) > t\right]\right\}.$$

Then α solves $\partial D\left(F_\beta\right)/\partial\beta \mid_{\beta=\beta_0} = E\left[\alpha S_\beta'\right]$ where F_β is any parametric model parameterized by finite dimensional β such that F_{β_0} is the true F at β_0; F_β is the distribution of a single observation and S_β is the score for β at β_0. To find this solution let β denote evaluation at F_β. Define $D^*\left(\vec{\beta}\right) = D^*(\beta_1, \ldots, \beta_5) = \sum_{t=1}^{T} D_t^*\left(\vec{\beta}\right)$ where

$$D_t^*\left(\vec{\beta}\right) = E_{\beta_1}\left\{\begin{array}{c} \{h_t\left[Z, M(\beta_4)\right] - \mathcal{L}\left(h_t, M(\beta_3), M(\beta_4), \beta_3, \beta_5\right)\} \\ \{\epsilon_t\left[M(\beta_2)\right] - \gamma_t\left[M(\beta_2)\right]\} I(Y > t) \end{array}\right\}.$$

Let $D_{tk}^* = \partial D^*\left(\vec{\beta}\right)/\partial\beta_k \mid_{\beta_j=\beta_0, j=1,\ldots,5}$. Then, by the chain rule,

$$\alpha = \sum_{k=1}^{5} \alpha_k, \alpha_k = \sum_{t=1}^{T} \alpha_{tk} \text{ and } D_{tk}^* = E\left[\alpha_{tk} S_\beta'\right].$$

That $\alpha_{t1} = D_t^*$ is straightforward to check. That $\alpha_{t5} = E\left(D_t^* \mid Z\right)$ follows from Z ancillary and Lemma (4.2) in Newey (1990). To prove the rest of Lemma (6.1) we use the identity

$$\begin{aligned} \partial m(t, z; \beta_0)/\partial\beta &= \partial F(t \mid z; \beta_0)/\partial\beta \\ &= E\left[a(t, Z, Y) S_\beta' \mid Z = z\right] \\ &= E\left[a(t, Z, Y) I(Z = z) S_\beta'/f_Z(z)\right] \end{aligned} \tag{A.1}$$

To obtain α_{t2} and α_{t4}, we differentiate under the integral sign to obtain

$$D_{t4}^* = E\left[\left\{h_{t[2]}(Z, M) - \mathcal{L}\left(h_{t[2]}, M\right)\right\} I(Y > t)\left\{\epsilon_t - \gamma_t(M)\right\} \partial m(Y, Z, \beta_0/\partial\beta)\right]$$

and

$$D_{t2}^* = E\left[\{h_t(Z,M) - \mathcal{L}(h_t,M)\} I(Y>t) \left\{v_{[2]}(t,M,Z,\theta_0) - \gamma_{t[1]}(M)\right\} \partial m(Y,Z,\beta_0)/\partial\beta\right]$$

and then use Eq. (A.1). D_{t3}^* and thus α_{t3} are more difficult to obtain, since $I[m^{-1}(u,Z,\beta_3) > t]$ is not smooth in β_3 and so we cannot differentiate under the integral sign. If Z is dichotomous, then

$$D_{t3}^* = \lim_{\beta\to\beta_0} (\beta-\beta_0)^{-1} E\left[\begin{array}{c}\{h_t(Z,M) - q_t^*(M)\} \\ I\left[M(\beta) > m(t,1,\beta) \vee m(t,0,\beta)\right]\{\epsilon_t - \gamma_t(M)\}\end{array}\right].$$

If

$$m(t,0) > m(t,1),\ D_{t3}^* = f_Z(0) \lim_{\beta\to\beta_0} (\beta-\beta_0)^{-1} \int_{m^{-1}[m(t,1,\beta),0,\beta]}^{\infty}$$

$$\left\{h_t(0,m(y,0)) - q_t^*[m(y,0)]\right\}\left\{\gamma_{op,t}[m(y,0)] - \gamma_t[m(y,0)]\right\} f_{Y|Z}(y \mid 0)\,dy =$$

$$\alpha_{t3}^0 f_{Y|Z}\left[m^{-1}(k^*(t),z^*(t)) \mid z^*(t)\right]$$
$$\partial\left\{m^{-1}\left[m(t,1-z^*(t),\beta),z^*(t),\beta\right]\right\}/\partial\beta\,|_{\beta=\beta_0}=$$
$$\alpha_{t3}^0\left[\partial m(t,1-z^*(t),\beta_0)/\partial\beta - \partial m\left[m^{-1}\{k^*(t),z^*(t)\},z^*(t),\beta_0\right]/\partial\beta\right].$$

But, by (A.1), the last term is $E[\alpha_{t3}^1(Y,Z)S_\beta']$. So $D_{t3}^* = E[\alpha_{t3}^0\alpha_{t3}^1(Y,Z)S_\beta']$. If $m(t,0) < m(t,1)$, the same formula holds. Finally, if $m(t,0) = m(t,1)$, $\alpha_{t3}^0\alpha_{t3}^1(Y,Z)$ is the same whether $z^*(t)$ is defined to be 0 or 1, since $f_Z(0)h_t(0,m(t)) + f_Z(1)h_t(1,m(t)) = q_t^*[m(t)]$. Thus, the same formula for D_{t3}^* holds in this case.

If $\gamma_t(u) = \gamma_{op,t}(u)$ then $D_{t4}^* = D_{t3}^* = 0$ and, thus, $\alpha_{t4} = \alpha_{t3} = 0$ for general Z since $E\left[D_t^*(\beta_0,\beta_4,\beta_3,\beta_0,\beta_0) \mid M\right] = 0$ for all β_4,β_3. Also $\alpha_{t5} = 0$ since, by direct calculation, $E\left(D_t^* \mid Z\right) = 0$.

Sketch of Proof of Theorem 2.1 and Lemma 6.2. Arguing as above, it is sufficient to obtain an explicit formula for α. We only consider the case in which $\gamma = \gamma^h$.

If $\gamma(u,z,M) = h(u,z,M)\lambda_\epsilon(u \mid M)$, then $D^*(F^*) =$

$$E_{F^*}\left\{\begin{array}{c}\int_{-\infty}^{\infty}[dN_{\epsilon\{M(F^*)\}}(u) - I\left[v\{M(F^*)\} > u\right]\lambda_\epsilon(u \mid M(F^*),F^*)\,du] \\ \left[h(u,Z,M,F^*) - \mathcal{L}(h,M(F^*),M(F^*),F^*,F^*)\right]\end{array}\right\}$$

where $\epsilon(u) = \overset{\circ}{v}\left(X^0,u,Z,\theta_0\right)$, $v(u) = \overset{\circ}{v}(X,u,Z,\theta_0)$, and $\mathcal{L}(h,u_1,u_2,F_1,F_2) =$

$$E_{F_2}\left[I\left\{v\left[m^{-1}(u_1,Z,F_1),u_1,Z\right] > u_2\right\} h(u_2,Z,u_1)\right] / E_{F_2}\left[I\left\{v\left[m^{-1}(u_1,Z,F_1),u_1,Z\right] > u_2\right\}\right]$$

Write

$$D^*\left(\vec{\beta}\right) = D^*(\beta_1, \ldots, \beta_6) =$$
$$E_{\beta_1}\left[\int_{-\infty}^{\infty}\left\{dN_{\epsilon\{M(\beta_2)\}}(u) - I\left[v\left\{M(\beta_2)\right\} > u\right]\lambda_\epsilon(u \mid M(\beta_2), \beta_6)\,du\right\}\right.$$
$$\left.\{h(u, Z, M(\beta_4)) - \mathcal{L}(h, M(\beta_3), M(\beta_4), \beta_3, \beta_5)\}\right]$$

Now $D_4^* = D_3^* = D_5^* = 0$ since $N_\epsilon(t) - \int_{-\infty}^{t} I[v > u]\lambda_\epsilon(u \mid M)\,du$ is a martingale. $D_6^* = 0$ since $E[I(v > u)]\{h(u, Z, M) - \mathcal{L}(u, h, M)\} \mid M] = 0$. It is straightforward to show $D_1^* = E[D^* S_\beta']$ so $\alpha_1 = D$. Differentiating w.r.t. β_2 and using (A.1) we obtain $D_2^* = E[\alpha_2 S_\beta']$ with α_2 as in Lemma 6.2.

Proof of Corollary 6.1. To prove Corollary 6.1, note that if $r_1(u, x) \equiv \gamma_1(u, x)/\lambda_\epsilon(u \mid x) - r_2(u, x)$ where, again, $r_2(u, x) \equiv \int_{-\infty}^{u} \gamma_1(t, x)$, then the supposition of Lemma 6.2 holds.

Acknowledgment

This work was supported in part by NIH grant ROI-AI32475.

References

1. G. Chamberlain, "Efficiency Bounds for Semiparametric Regression," Technical Report, Department of Economics, University of Wisconsin, 1988.
2. D. M. Dabrowska, "Non-parametric regression with censored survival data," *Scandinavian Journal of Statistics*, vol. 141 pp. 81–187, 1987.
3. D. Glidden and L. J. Wei, "Rank estimation of treatment differences based on repeated measurement subject to dependent censoring," 1995, Submitted.
4. J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, John Wiley: New York, 1980.
5. D. Y. Lin, J. M. Robins and L. J. Wei, "Comparing two failure time distributions in the presence of informative censoring," *Biometrika*, 1995, to appear.
6. W. K. Newey, "The asymptotic variance of semiparametric estimators," *Econometrica*, vol. 62 pp. 1349–1382, 1994.
7. W. K. Newey, "Semiparametric efficiency bounds," *Journal of Applied Econometrics*, vol. 5 pp. 99–135, 1990.
8. A. Pakes and D. Pollard, "Simulation in the asymptotics of optimization estimators," *Econometrica*, vol. 57 pp. 1027–1057, 1989.
9. Y. Ritov and J. A. Wellner, "Censoring, Martingales, and the Cox model," *Contemporary Mathematical Statistics Inf. Stochastic Procedures*, American Mathematical Society, ed. by N. U. Prabhu, vol. 80, pp. 191–220, 1988.
10. J. M. Robins, "The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies," *Health Service Research Methodology: A Focus on AIDS*, ed. by L. Sechrest, H. Freeman, and A. Mulley, NCHSR, U.S. Public Health Service, pp. 113–159, 1989.

11. J. M. Robins, , "Analytic methods for HIV treatment and cofactor effects," *Methodological Issues of AIDS Behavioral Research*, ed. by D. G. Ostrow and R. Kessler, pp. 213–287. Plenum Press: New York, 1993.

12. J. M. Robins, "An analytic method for randomized trials with informative censoring: Part I," *Lifetime Data Analysis*, vol. 1, pp. 241–254, 1995.

13. J. M. Robins and A. A. Tsiatis, "Correcting for non-compliance in randomized trials using rank preserving structural failure time models," *Communications in Statistics—Theory*, vol. 20 pp. 2609–2631, 1991.

14. J. M. Robins, S. D. Mark and W. K. Newey, "Estimating exposure effects by modelling the expectation of exposure conditional on confounders," *Biometrics*, vol. 48 pp. 479–495, 1992.

15. J. M. Robins and A. Rotnitzky, "Recovery of information and adjustment for dependent censoring using surrogate markers," in *AIDS Epidemiology—Methodological Issues*, ed. by N. Jewell, K. Dietz, and V. Farewell, pp. 297–331. Birkhäuser: Boston, MA, 1992.

16. D. B. Rubin, "Bayesian inference for causal effects: The role of randomization," *The Annals of Statistics*, vol. 6 pp. 34–58, 1978.