COMMUN. STATIST.-THEORY METH., 20(8), 2609-2631 (1991)

CORRECTING FOR NON-COMPLIANCE IN RANDOMIZED TRIALS USING RANK PRESERVING STRUCTURAL FAILURE TIME MODELS

James M. Robins and Anastasios A. Tsiatis

Department of Biostatistics
Harvard School of Public Health
665 Huntington Avenue, Boston, Massachusetts 02115 USA

Key Words and Phrases: causal inference; counterfactual models; survival analysis; censored data.

ABSTRACT

We propose correcting for non-compliance in randomized trials by estimating the parameters of a class of semi-parametric failure time models, the rank preserving structural failure time models, using a class of rank estimators. These models are the structural or strong version of the "accelerated failure time model with time-dependent covariates" of Cox and Oakes (1984). In this paper we develop a large sample theory for these estimators, derive the optimal estimator within this class, and briefly consider the construction of "partially adaptive" estimators whose efficiency may approach that of the optimal estimator. We show that in the absence of censoring the optimal estimator attains the semiparametric efficiency bound for the model.

1. *Introduction*

In a randomized trial designed to study the effect of a drug therapy on survival, study subjects are assigned to a treatment regime, or, equivalently, a treatment protocol. Unfortunately, subjects often fail to comply with their assigned regime, so that a subject's actual treatment may differ from his or her assigned treatment. As is well known, non-compliance can drastically effect the power of the usual intention to treat log rank test of the null-hypothesis of no treatment difference. Less commonly recognized is that, from a public health point of view,

the causal parameter of interest will often be a function of the survival differences that would have been observed had, contrary to fact, all subjects remained on protocol (see Robins, 1989). This paper considers methods for the estimation of these survival differences, when reliable estimates of each subject's actual treatment are available. Estimates of "actual treatment" might, for example, be obtained by measuring the amount of "active drug" in the subject's blood or urine at each follow-up visit or by pill counting techniques. Until the Discussion section, we restrict attention to trials in which one or more dosage schedules of a single active drug are to compared to one another and/or to placebo.

In general, the survival differences that would have been observed had all subjects remained on protocol will not be identifiable without making various assumptions that are themselves non-identifiable. The simplest, but least plausible assumption, is that subjects who comply with their assigned treatment protocol are comparable to those who fail to comply with regard to important prognostic factors. In this paper we introduce a class of models, the rank-preserving structural failure time models (RPSFTM) that, when true, are sufficient to identify the survival differences that would have been observed had all subjects remained on protocol and yet do not require comparability of compliers and non-compliers. These models are "the structural or strong version" of the accelerated failure time models with time dependent covariates introduced by Cox and Oakes (1984). We have adopted terminology used in the social science and econometric literature and have called our models "structural models" because they directly model the survival times that would have been observed had, contrary to fact, all exchangeable subjects received the same treatment regime (Rubin, 1978). These models are rank-preserving in the sense that they assume that, given any two subjects i and j, if subject i fails before subject j when both followed a particular treatment regime, then subject i would fail before subject j when both followed any other regime. We consider a class of rank estimators for these models originally proposed by Robins (1989). We develop a large sample theory for these rank estimators, derive the optimal estimator within the class, and briefly consider the construction of "partially adaptive" estimators whose efficiency may approach that of the optimal estimator. In the absence of censoring this class contains an estimator that attains the semiparametric efficiency bound for the model. Finally, we shall show that the estimation approach developed in this paper is an extention of and does not conflict with the standard "intention-to-treat" log-rank test of no treatment effect used in a

randomized trial. Indeed, we shall construct an adaptive weighted log-rank test whose power against RPSFTM alternatives is guaranteed to be no less than that of the log-rank test.

This paper is organized as follows. In Section 2a, we motivate the need for new methods to adjust for non-compliance in randomized trials. In Section 2b, we introduce the RPSFTM. In Section 3, we provide a formal characterization of statistical problem to be solved in this paper. In Section 4, we define a class of rank tests and estimators for the parameters of an RPSFTM assuming access to uncensored data. In Section 5, we generalize our test and estimators to account for censoring. A critical evaluation of the model and its interpretation will be given in the discussion section.

2a. The Need for New Methods to Correct for Non-Compliance

New drug treatments are often compared to placebo therapy in a randomized clinical trial. If everyone complied with the treatment to which they were randomized, then one can apply standard techniques for the analysis of survival data, such as the Cox proportional hazards model, to estimate the treatment effect. The Cox proportional hazard model assumes that $\lambda(t \mid D) = \lambda_0(t)\exp(\beta_0 D)$ where D is a treatment indicator and $\lambda(t \mid D)$ is the hazard rate for failure at time t given treatment. β_0 represents the log hazard ratio comparing active treatment to placebo. β_0 can be estimated using partial likelihood methods.

However, most studies are not ideal and individuals do not always comply with the treatment to which they were randomized. In such a case, it is not obvious how to estimate the treatment effect. To focus the discussion, suppose that in a double blind placebo-controlled trial of the effect of daily aspirin therapy on the survival of subjects with previous non-fatal myocardial infarction, individuals assigned to the active treatment arm are to receive a single 300 mg tablet of aspirin daily while subjects assigned to placebo receive no aspirin. All individuals are instructed to refrain from taking any other aspirin-containing medications. Nonetheless, in both treatment arms, some individuals self-medicate themselves with additional aspirin-containing drugs. Additionally some subjects fail to take their assigned pill. We shall suppose that by a combination of pill-counting techniques, measurement of urinary aspirin metabolites, and personal interviews, we can determine whether a subject was taking aspirin at each time t. In this setting a straightforward but naive approach to estimating the treatment effect would be to

compare those who are receiving treatment to those who are not using a time-dependent Cox model. Specifically, we could define a time-dependent treatment indicator D(t) with D(t) = 1 if an individual is taking active drug at t and D(t)=0 otherwise; and then fit the Cox model $\lambda(t \mid H(t)) = \lambda_0(t)\exp(\beta_0 D(t))$ where $H(t) \equiv \{D(u) ; 0 \leq u \leq t\}$ denotes the entire history of treatment up to time t. The problem with this approach is that the parameter β_0 may not have a causal interpretation when compliance is non-random, i.e., when individuals who comply with their assigned treatment are prognostically different from those who do not. For example, β_0 will be negative even if treatment has no causal effect on any individual's survival if, in the active treatment arm, subjects who become gravely ill stop taking aspirin against protocol specifications shortly before death.

Therefore, the more commonly used analytic approach is an intent-to-treat analysis. That is, the treatment effect is estimated based on the treatment group to which an individual was randomized regardless of whether the individual complied with their assigned treatment. For example, if we denote by R our randomization indicator, i.e., R=1 for individuals randomized to receive active treatment and R=0 otherwise, then we can model the intent-to-treat treatment effect as $\lambda(t \mid R) = \lambda_0(t)\exp(\beta_0 R)$. An intent-to-treat analysis has both pros and cons. A point in its favor is that the comparability of the two treatment arms is guaranteed by randomization, and, thus, the null hypothesis is preserved. That is, if active treatment has no causal effect on any individual's survival, then the survival distribution of the two randomized groups will be identical, and the parameter β_0 would be equal to zero even in the presence of non-random non-compliance. Secondly, the value of the parameter β_0 would correspond to the overall treatment effect that would be realized if this treatment were actually adopted and practiced in the community, under the proviso that the rate of non-compliance observed in the trial would equal the subsequent rate of non-compliance in the community. A point against a intention-to-treat analysis is that the parameter β_0 is not measuring the true biological effect of treatment, but rather a mixture of the effect on the compliers with the absence of effect on the non-compliers due to their non-compliance. Hence, the "intent to treat" measure of treatment effect would diminish as non-compliance increased. More importantly, the rate of non-compliance in the community, once the treatment is adopted, may not be the same as the rate in the original clinical trial. For example, once the treatment is proved to be efficacious in a trial, then nearly all individuals in the community may be willing to stringently

comply with the active treatment protocol. In that case the intention to treat parameter β_0 would not represent the overall effect of treatment in community. Additionally, from the point of view of an individual patient who is planning to comply with therapy, the intent-to-treat parameter β_0 can underestimate the benefit to be expected from therapy. Therefore it is important to develop methods that, even in the presence of non-random non-compliance, can estimate the true treatment effect, i.e., the effect that would be realized if all individuals complied with the treatment protocol to which they were assigned, while preserving the unbiased test of the null hypothesis available from an intention to treat analysis. To achieve this goal, we shall consider the class of RPSFTM.

2b. The RPSFTM

We shall estimate the treatment effect using a RPSFTM, i.e., the strong version of the accelerated failure time model with time-dependent covariates proposed by Cox and Oakes (1984, Sec. 5.2). Specifically, let U_i denote the lifetime of the i^{th} individual if, possibly contrary to fact, he/she was never to receive treatment, i.e., $d(t) = 0$ for all t. U_i will be referred to as the latent baseline failure time. Throughout we shall assume that the time an individual would fail if never treated does not depend on the treatment arm to which the individual is assigned. This assumption would be expected to hold in a double-blind trial. In the absence of censoring, the observable random variables are $(T_i, H_i(T_i), R_i)$ where T_i is the i^{th} individuals observed lifetime, $H_i(T_i) = \{D_i(u) ; 0 \le u \le T_i\}$ is the observed treatment history, and R_i is a randomization group indicator. Realization of random variables will be denoted by lower case letters. For the moment we shall suppose that $D_i(u) \in \{0,1\}$ for all i and u. A simple example of an RPSFTM assumes that U_i is related to the observable random variables by the relationship

$$U_i = \int_0^{T_i} \exp\{\beta_0 D_i(x)\}dx \qquad (1)$$

where β_0 is an unknown parameter.

To better understand the model it is convenient to follow the suggestion of Cox and Oakes (1984) and consider two time scales. Namely, let u correspond to the baseline time scale and t the actual or real time scale. Let the mapping from t to u for the i^{th} individual be given by $u = \int_0^t \exp\{\beta_0 D_i(x)\}dx$. Then $du/dt = \exp(\beta_0 D_i(t))$ and $dt/du = \exp\{-\beta_0 D_i(t)\}$. dt/du corresponds to the relative rate at which real time is being used up compared to baseline time. Thus, if individual i

has U_i years of baseline time to be used up, the actual time T_i at which these U_i years of baseline time will have been used is given by (1).

In order to understand the causal meaning of the parameter β_0, let V_i be the i^{th} individuals lifetime had the i^{th} individual received continuous treatment beginning at the time of randomization, i.e., $d(t) = 1$ for all $t \leq V_i$. As suggested by model (1), we suppose that $U_i = \int_0^{V_i} \exp(\beta_0)dx = V_i\exp(\beta_0)$. Thus $\beta_0 = 0$ implies no causal effect of treatment on survival while $\beta_0 < 0$ implies that continuous treatment would extend life by a factor of $\exp(-\beta_0)$, and $\beta_0 > 0$ implies that continuous treatment decreases life by a factor of $\exp(-\beta_0)$. Note that, under model (1), the expansion factor $\exp(-\beta_0)$ is assumed to be the same for all individuals. In the discussion section, we consider ways in which this assumption might be relaxed.

In Section 4, we shall develop non-parametric rank tests of the hypothesis that a particular value of the parameter β is the true value β_0, under the additional assumption that the distribution of the baseline lifetime U_i is independent of treatment group R_i. That is

$$\Pr[U_i \geq x \mid R_i] = S_0(x) \quad \text{for } R_i \epsilon (0,1)$$

We have assumed this equation holds because the latent failure time variable U_i is a "pretreatment variable" in the sense that, like eye color, it is a fixed characteristic of the individual that does not depend on the treatment arm assigned or on the actual treatment history. In a completely randomized trial, the treatment arm indicator R_i is independent of any pretreatment variable.

Our model (1) is a special case of a RPSFTM. A RPSFTM links the baseline latent lifetime variable U_i to $\{T_i, H_i(T_i)\}$ by assuming

$$U_i = \psi(T_i, H_i(T_i), \beta_0)$$

where $\beta_0 \epsilon R^V$ is an unkown parameter, $\psi(\cdot,\cdot,\cdot)$ is a known smooth function satisfying (a) $\psi(t,H_i(t),\beta) > \psi(u,H_i(u),\beta)$ if $t > u$, (b) $\psi(t,H_i(t),\beta) = t$ if $H_i(t)$ is identically 0 on $(0,t)$ and (c) $\psi(t,H_i(t),0) = t$ so $\beta_0=0$ represents the null hypothesis of no-treatment effect. Further we shall assume that given any prespecified treatment regime $h = \{d(u) ; 0 \leq u < \infty\}$, the failure time of subject i had he followed regimes h, say $V_i^{(h)}$, is obtained from his baseline failure time U_i by solving $U_i = \psi(V_i^{(h)}, h(V_i^{(h)}), \beta_0)$. As noted previously, for the RPSFTM of Eq. (1) and $h = \{d(u)=1 ; 0 \leq u < \infty\}$, $V_i^{(h)} = U_i\exp(-\beta_0)$.

The tests and estimates in Section 4 are based on the fact that the observable random variables $U_i(\beta) \equiv \psi(T_i,H_i(T_i),\beta)$ and R_i are independent for $\beta=\beta_0$ since, under the model, $U_i(\beta_0) = U_i$.

3. A Formalization of the Problem

The above considerations lead to the following statistical problem defined solely in terms of the observables.

Let T be an observable random variable corresponding to the failure time of subjects entered in our randomized trial with time recorded as time since randomization. Henceforth, let D(t), $0\le t\le T$ be a possibly vectored-valued random variable recording the value at time t of a vector of possibly time-dependent covariates. For $0\le t\le T$, let $H(t) = \{D(u) ; 0\le u\le t\}$ be the history of the process D(t) through t. Let R be the random variable taking values in the set {1,2,...,J} that records the arm of the J-armed trial to which a subject has been assigned.

We shall assume $(T_i,H_i(T_i),R_i)$ are independently and identically distributed random vectors where i indexes study subjects. We shall denote realizations of random variables by small letters so that t,r,d(t),h(t) represent particular realizations of T, R, D(t), H(t).

We assume the existence of a known real-valued function $\psi(t,h(t),\beta)$ with $\beta = (\beta_1,...,\beta_V)^T \in R^V$ satisfying the following properties:

Smoothness: $\psi_1(t,h(t),\beta)$, $\psi_{3,v}(t,h(t),\beta)$, and $\psi_{13,v}(t,h(t),\beta)$ are continuous for all β and almost all t with respect to Lesbesgue measure where $\psi_1(t,h(t),\beta) \equiv \partial\psi(t,h(t),\beta)/\partial t$, $\psi_{3,v}(t,h(t),\beta) \equiv \partial\psi(t,h(t),\beta)/\partial\beta_v$, $\psi_{13,v}(t,h(t),\beta) \equiv \partial\psi_1(t,h(t),\beta)/\partial\beta_v$. Note $\psi_1(t,h(t),\beta)$ is the derivative with respect to t rather than with respect to only the first argument of $\psi(\cdot,\cdot,\cdot)$.

Monotonicity: $\psi(t,h(t),\beta) > \psi(u,h(u),\beta)$ if t>u where we adopt the convention that, if h(t) and h(u) are used in the same expression, h(u) is the initial segment of h(t) through u.

Identity: $\psi(t,h(t),0)=t$

Independence and Identification: There exists a unique β_0 such that

$$U(\beta_0) \perp\!\!\!\perp R \tag{2}$$

where $U(\beta) \equiv \psi(T,H(T),\beta)$ and, as in Dawid (1979), $A \perp\!\!\!\perp B$ means A and B are independent.

Given these properties, we shall call β_0 the true value of the parameter β. We shall call $U(\beta_0)$ the true residual. We shall call $U(\beta)$ the β-residual.

Example 1: Model (1) of Section 2b has $\psi(t,h(t),\beta) = \int_0^t \exp[\beta d(x)]dx$.

This paper is concerned with semi-parametric point and interval estimates for β_0. We note that, since $U(0)=T$, the null hypothesis that $\beta_0 = 0$ is equivalent to the hypothesis that $T \perp\!\!\!\perp R$. This latter hypothesis is the usual "intention-to-treat" null hypothesis tested in the standard intention-to-treat analysis of a randomized trial. Thus any distribution-free test of the null hypothesis $\beta_0 = 0$ will be a distribution-free test of the intention-to-treat null hypothesis $T \perp\!\!\!\perp R$.

4. Inference concerning β_0

We shall first consider the case in which there is no censoring.

4a. Interval Estimation

We shall first construct a class of rank tests of the hypothesis $\beta = \beta_0$ when β_0 is the true value of the parameter and β is a hypothesized value.

Define $N_i(u,\beta) = I[\psi(T_i,H_i(T_i),\beta) \le u] = I[U_i(\beta) \le u]$, and $Y_i(u,\beta) = I[U_i(\beta) \ge u]$ where $I[A]=1$ if A is true and $I[A]=0$ otherwise. $N_i(u,\beta)$ counts failures and $Y_i(u,\beta)$ records "at risk" status on the β-residual time scale. Let n denote the number of subjects entered in the trial.

Given a V-vector $g(r,u,\beta)$ of known real valued functions $g_v(r,u,\beta)$, $v \in \{1,2,...,V\}$, $S_n(\beta,g)$ is defined to be the V-vector with components

$$S_{n,v}(\beta,g) = \sum_{i=1}^{n} \int dN_i(u,\beta)[g_v(R_i,u,\beta) - \bar{g}_v(u,\beta)]$$

where

$$\bar{g}_v(u,\beta) = \sum_{i=1}^{n} g_v(R_i,u,\beta)Y_i(u,\beta) \Big/ \sum_{i=1}^{n} Y_i(u,\beta)$$

Note that if V is one-dimensional and $\beta=0$, then $S_n(\beta,g)$ is the numerator of a standard "intention-to-treat weighted log rank test" of the intention-to-treat null hypothesis $T \perp\!\!\!\perp R$. We have not included a weight function $W(u,\beta)$ in defining $S_n(\beta,g)$ because any such weight function can be subsumed into the function $g(r,u,\beta)$ by multiplication.

We now show that $n^{-1/2} S_n(\beta,g)$ converges in distribution to a mean 0 multivariate normal random variable under the null hypothesis $\beta = \beta_0$. First note that $S_{n,v}(\beta,g)$ can be written

$$\sum_{i=1}^{n} \int dM_i(u,\beta)[g_v(R_i,u,\beta) - \bar{g}_v(u,\beta)]$$

where $M_i(u,\beta) = N_i(u,\beta) - \int_0^u \lambda(x)Y_i(x,\beta)dx$ and $\lambda(x)$ is the hazard function of the random variable $U(\beta_0)$. Therefore, $M_i(u,\beta_0)$ is a martingale process with respect to the filtration $F_n(u,\beta_0)$ where $F_n(u,\beta)$ consists of all information regarding both the random variable $U(\beta)$ up to time u and treatment arm membership R. That is $F_n(u,\beta) = \sigma[R_i,\{N_i(x,\beta)\ ;\ x\le u\}\ ;\ i\in(1,\cdots,n)]$. The filtration does not include information on a subject's actual treatment history. Since $g_v(R,u,\beta_0)$ and $\bar{g}_v(u,\beta_0)$ are predictable with respect to $F_n(u,\beta_0)$, $S_{n,v}(\beta_0,g)$ is also a martingale process with respect to the filtration $F_n(u,\beta_0)$. Therefore, using standard martingale regularity conditions and results as in Aalen (1976, 1978), Gill (1980), and Andersen et al. (1982), $n^{-1/2}S_n(\beta_0,g)$ converges to a mean 0 multivariate normal distribution with covariance matrix

$$\Omega(\beta_0,g) = \text{plim}\left[\int_0^{T^*} n^{-1} \sum_{i=1}^{n} \{g(R_i,u,\beta_0) - \bar{g}(u,\beta_0)\}\{g(R_i,u,\beta_0) - \bar{g}(u,\beta_0)\}^T Y_i(u,\beta_0)\lambda(u)du\right] \tag{3}$$

that can be consistently estimated by $\tilde{\Omega}(\beta_0,g) \equiv$

$$n^{-1}\sum_{i=1}^{n} \int dN_j(u,\beta_0) \cdot \left\{ \frac{\sum_{i=1}^{n} Y_i(u,\beta_0)[g(R_i,u,\beta_0) - \bar{g}(u,\beta_0)] \cdot [g(R_i,u,\beta_0) - \bar{g}(u,\beta_0)]^T}{\sum_{i=1}^{n} Y_i(u,\beta_0)} \right\}$$

where a superscript T denotes matrix transposition. Therefore, the set of β for which $[S_n(\beta,g)^T\tilde{\Omega}^{-1}(\beta,g)S_n(\beta,g)]/n$ is less than the 95^{th} upper percentile of a χ^2_V distribution is a 95% large sample joint confidence set for β_0.

The precise regularity conditions we use are the following straightforward generalizations of the regularity and stability conditions (A), (C)-(F) of Tsiatis (1990). First, all stochastic integrals will be truncated at T^*, satisfying for some $c>0$, $p[U_i(\beta_0)\ge T^*+c]\ge\sigma>0$. Then we assume

(A) The density of $U_i(\beta_0)$ exists and is bounded by K_i for all $x\le T^*+c$.

(C) There exists $\theta(u)$ such that $\left|\lambda_{U(\beta)}[u \mid R_i] - \lambda(u) - \frac{\partial\lambda_{U(\beta)}(u\mid R_i)}{\partial\beta}\Big|_{\beta=\beta_0}\right| \le \epsilon^2\theta(u)$ for

all $u \leq T^*$ and $\epsilon \equiv \beta - \beta_0$ such that $|\epsilon| \leq c$ where $\lambda_{U(\beta)}[u \mid \cdot]$ is the hazard function of the random variable $U(\beta)$ conditional on $\cdot$.

(D) $g(R_i,u,\beta)$ is bounded with probability one for all $u \leq T^*+c$, and β in a neighborhood B of β_0.

(E) There exists a continuous vector-valued function $\mu(u,\beta,g)$ of β in a neighborhood B of β_0 such that

$$\sup_{\beta \in B, u \leq T^*+c} \{ \| \bar{g}(u,\beta) - \mu(u,\beta,g) \| \} \xrightarrow{P} 0 \text{ where } \bar{g}(u,\beta) = \sum_{i=1}^{n} Y_i(u,\beta) g(R_i,u,\beta) \Big/ \sum_{i=1}^{n} Y_i(u,\beta)$$

where $\| \cdot \|$ is the usual Euclidean norm.

(F) There exists a continuous matrix-valued function $A(u,\beta,g)$ such

$$\sup_{\beta \in B, u \leq T^*+c} [\| n^{-1} \sum_{i=1}^{n} \{g(R_i,u,\beta) - \bar{g}(u,\beta)\}\{g(R_i,u,\beta) - \bar{g}(u,\beta)\}^T Y_i(u,\beta) - A(u,\beta,g) \|] \xrightarrow{P} 0$$

where $\| Q \|$ is the square root of the sum of the squared elements of the matrix Q.

4b. Point Estimation

Let $\tilde{\beta}_n(g)$ be a value of β that solves $S_n(\beta,g) = 0$. Since $S_n(\beta,g)$ is a step function in β we actually choose $\tilde{\beta}_n(g)$ minimizing $S_n(\beta,g)^T S_n(\beta,g)$. We will show that $\tilde{\beta}_n(g)$ is under regularity conditions consistent and asymptotically normal. We also shall derive its asymptotic variance. For pedagogic purposes we first consider the case that $\beta \in R^1$.

Since the statistic $S_n(\beta,g)$ is a step function in β, it cannot be expanded around β_0 in a Taylor series expansion as is usually done to prove consistency and asymptotic normality. We therefore shall use the methods of Tsiatis (1990) to prove our results. Although Tsiatis (1990) developed asymptotic theory for rank estimators of a linear regression model, his methods extend straightforwardly to include the RPSFTM. Specifically we shall show that $S_n(\beta,g)$ is approximately linear in β near the true value β_0.

To do so, we first note that $M_i(u,\beta)$ is not an $F_n(u,\beta)$ martingale if $\beta \neq \beta_0$ since $\lambda(u)Y_i(u,\beta)$ is not the intensity process of $N_i(u,\beta)$. Hence, as in Tsiatis (1990), we rewrite $n^{-1/2}S_n(\beta,g)$ as $n^{-1/2}S_n^{\Delta}(\beta,g)$ plus

$$n^{-1/2} \sum_{i=1}^{n} \int_0^{T^*} \left[\lambda_{U(\beta)}[u \mid R_i] - \lambda(u) \right] Y_i(u,\beta) \cdot \{g(R_i,u,\beta) - \bar{g}(u,\beta)\} du \qquad (4)$$

where

$$n^{-1/2}S_n^{\Delta}(\beta,g) \equiv n^{-1/2}\sum_{i=1}^{n}\int_0^{T^*}\Big[\, dN_i(u,\beta)-\lambda_{U(\beta)}[u \mid R_i]\cdot Y_i(u,\beta)du\Big]\cdot[g(R_i,u,\beta)-\bar{g}(u,\beta)] \tag{5}$$

Equation (5) now has the appropriate intensity process subtracted off and hence is a $F_n(u,\beta)$ martingale process. Indeed, if we assume that β is a function of n such that $n^{1/2}(\beta-\beta_0)$, converges to a non-zero finite limit we can use Lemma's (3.1), (3.2), and Theorem (3.2) of Tsiatis (1990) to show $n^{-1/2}\,|\,S_n^{\Delta}(\beta,g)-S_n(\beta_0,g)\,|\xrightarrow{P}0$ uniformly in neighborhoods of $O(n^{-1/2})$ of β_0 under our regularity conditions. Furthermore, using lemma (3.3) of Tsiatis (1990), we can also show that Eq. (4) is uniformly close to $[n^{1/2}(\beta-\beta_0)]Q(\beta_0,g)$, where

$$Q(\beta_0,g) \equiv$$

$$\text{plim}\left[\sum_{i=1}^{n} n^{-1}\int_0^{T^*}\left[\frac{\partial\lambda_{U(\beta)}(u|R_i)}{\partial\beta}\Big|_{\beta=\beta_0}\right]\cdot\left\{g(R_i,u,\beta_0)-\bar{g}(u,\beta_0)\right\}\cdot Y_i(u,\beta_0)du\right] \tag{6}$$

$$= \text{plim}\left[\int_0^{T^*} n^{-1}\sum_{i=1}^{n}\left\{g_{opt}(R_i,u,\beta_0)-\bar{g}_{opt}(u,\beta_0)\right\}\cdot\left\{g(R_i,u,\beta_0)-\bar{g}(u,\beta_0)\right\}\cdot Y_i(u,\beta_0)\lambda(u)du\right] \tag{7}$$

by some straightforward algebra, where $g_{opt}(R,u,\beta_0) = \dfrac{\partial\lambda_{U(\beta)}(u|R)}{\partial\beta}\Big|_{\beta=\beta_0}\Big/\lambda(u)$.

Thus, using results in Section 3 of Tsiatis (1990) culminating in Theorem (3.2), if $Q(\beta_0,g)\neq 0$, then for any $K>0$,

$$\sup_{|\beta-\beta_0|<KN^{-1}}[n^{-1/2}\,|\,S_n(\beta,g)-\{S_n(\beta_0,g)+n(\beta-\beta_0)Q(\beta_0,g)\}\,|\,]\xrightarrow{P}0\ . \tag{8}$$

Arguing as in Jureckova (1971) and Tsiatis (1990), this implies that there exists a solution $\tilde{\beta}_n(g)$ of $S_n(\beta,g)=0$ such that $n^{1/2}[\tilde{\beta}_n(g)-\beta_0]$ converges in law to a mean zero normal random variable with variance $\text{plim}[\tilde{\Omega}(\beta_0,g)]/Q^2(\beta_0,g)$. Even though $S_n(\beta,g)$ is not differentiable, in view of (8), $Q(\beta_0,g)$ can be consistently estimated by the symmetric "numerical derivative" $\hat{Q}(g)$ with step-size $c/\sqrt{n}$ where

$$\hat{Q}(g) = \frac{n^{-1}S_n(\tilde{\beta}_n(g)+c/\sqrt{n},\,g)-n^{-1}S_n(\tilde{\beta}_n(g)-c/\sqrt{n},\,g)}{2c/\sqrt{n}} \text{ and c is a fixed constant.}$$

By an application of the Cauchy-Schwartz inequality applied to (3) and (7) it follows that the most efficient function g is g_{opt} and that the asymptotic variance of the optimal estimator is $Q^{-1}(\beta_0, g_{opt})$.

In the Appendix, we develop an analytic expression for $\frac{\partial \lambda_{U(\beta)}(u|R)}{\partial \beta}|_{\beta=\beta_0}$.

In Theorem A.1 in the Appendix, we show that asymptotic variance of $\tilde{\beta}_n(g_{opt})$ attains the semi-parametric efficiency bound for the RPSFTM model under the sole restriction (2) as the truncation value T^* approaches infinity. It follows from the reciprocal relationship between locally most powerful tests and optimal estimators that the locally most powerful weighted intention-to-treat log rank test of the intention-to-treat null hypothesis against an RPSFTM alternative has g equal to $g_{opt}(R,u,0)$.

As in Hansen (1982), given any K-vector of scoring functions $g = (g_1,\cdots,g_K)^T$ with K>1, a partially adaptive estimator whose asymptotic efficiency may approximate that of the optimal but infeasible estimator $\tilde{\beta}_n(g_{opt})$ can be obtained by consistently estimating the optimal linear combination $\ell_{opt}S_n(\beta,g)$ of the vector of K estimating equations, $S_n(\beta,g) = (S_n(\beta,g_1),\cdots S_n(\beta,g_K))^T$ and then solving the resulting estimating equation $\hat{\ell}_{opt}S_n(\beta,g) = 0$. Here $\ell_{opt} = (Q(\beta_0,g_1),\cdots,Q(\beta_0,g_K))\{Var[n^{-1/2}S_n(\beta_0,g)]\}^{-1}$ and $\hat{\ell}_{opt} = (\hat{Q}(g_1),\cdots,\hat{Q}(g_K))\tilde{\Omega}^{-1}(\tilde{\beta}_n,g)$ where the $\hat{Q}(g_k)$ and $\tilde{\Omega}(\tilde{\beta}_n,g)$ are evaluated at a preliminary $n^{1/2}$-consistent estimator $\tilde{\beta}_n$. Newey (1990) discusses the possibility of obtaining a fully adaptive estimator by allowing K to increase with sample size at an appropriate rate.

An adaptive weighted log-rank test with local power against RPSFTM alternatives at least as great as that of the usual log-rank test can be based on $\hat{\ell}_{opt}S_n(0,g)$ provided the usual log-rank test can be written as a linear combination of the elements of $S_n(0,g)$. In the testing context, the preliminary estimator $\tilde{\beta}_n$ used in constructing $\hat{\ell}_{opt}$ is chosen to be zero. We note that $n^{-1/2}\hat{\ell}_{opt}S_n(\beta_0,g)$ and $n^{-1/2}\ell_{opt}S_n(\beta_0,g)$ have the same limiting distribution.

If, as in Section 4a, we let β be V-dimensional, similar calculations demonstrate that a root $\tilde{\beta}_n(g)$ of $S_n(\beta,g)=0$ exists such that $n^{1/2}[\tilde{\beta}_n(g)-\beta_0]$ is multivariate normal with covariance matrix $[Q(\beta_0,g)]^{-1}[\text{plim}\tilde{\Omega}(\beta_0,g)][Q(\beta_0,g)^T]^{-1}$, when $Q(\beta_0,g)$ is invertible, where $Q(\beta_0,g)$ is the V×V matrix with (w,v) entry given by Eq. (6) when we replace $\partial\beta$, g and $\bar{g}$ by $\partial\beta_w, g_v$ and $\bar{g}_v$. The most efficient g has components $g_{opt,v}(R,u,\beta_0)$ given by $\frac{\partial \lambda_{U(\beta)}(u|R)}{\partial \beta_v}|_{\beta=\beta_0} / \lambda(u)$. $Q(\beta_0,g)$ can be

consistently estimated by numerical partial derivatives with step-size proportional to $n^{-1/2}$.

5. Censoring

We shall consider only the following restricted type of censoring. We suppose that in our randomized trial follow-up ends on a particular calendar date and staggered entry is allowed. Let the random variable C record the known potential censoring time, defined as the difference between the end of follow-up date and the date at which a subject entered the trial. $(T_i,H_i(T_i),R_i,C_i)$ are assumed to be i.i.d. and $(U(\beta_0),C)\perp R$ since R was assigned completely at random. Let $X = \min(T,C)$. We suppose the observables are $(X,H(X),C,R,\tau)$ where $\tau = 1$ if $X=T$ and $\tau = 0$ otherwise. Since the potential censoring time C is known, it is, by definition, an observable. Define $U^*(\beta)=\psi(X,H(X),\beta)$.

Since, in the presence of censoring, $S_n(\beta,g)$ is not computable one might attempt to use the modification $S_n^*(\beta,g)$ of $S_n(\beta,g)$ obtained by replacing $N_i(u,\beta)$ and $Y_i(u,\beta)$ in Eq. (2) by $N_i^*(u,\beta)$ and $Y_i^*(u,\beta)$ where $N_i^*(u,\beta) = I[U_i^*(\beta)\le u,\tau_i=1]$ and $Y_i^*(u,\beta) = I[U_i^*(\beta)\ge u]$. Unfortunately, it is easy to see that in general $E[S_n^*(\beta_0,g)]\neq 0$ when $\beta_0\neq 0$. Therefore, we need to define other modifications of $S_n(\beta,g)$ that continue to have mean zero.

Specifically, let $C(\beta)=\min\{\psi(C,h(C),\beta);h(C)\epsilon\mathbf{H}(C)\}$ where $\mathbf{H}(C)$ is a particular, known set of h(C) such that $pr[H(C)\epsilon\mathbf{H}(C)\mid C,T\ge C]=1$. There will be efficiency advantages to selecting $\mathbf{H}(C)$ to be the smallest such set. That is $C(\beta)$ is less than or equal to the earliest time on the beta-residual-time-scale that any subject with potential censoring times C could have been censored. We shall suppose $C(\beta)>\epsilon>0$ for all β and some ϵ.

As a concrete example, suppose that our transformation is that given by the Cox-Oakes model of example 1 and that we define $\mathbf{H}(C)$ to be the set of all possible $h(C)=\{d(u)\ ;\ 0\le u\le C\}$ histories of a dichotomous (0,1) treatment variable d(u). Then $C(\beta) = C$ if $\beta\ge 0$ and $C(\beta) = \int_0^C \exp[\beta]dx = e^{\beta}C$ if $\beta<0$. Note that $C(\beta)$ is random only through its dependence on C. Furthermore $C(\beta=0) = C$. Define $X(\beta) = \min(U(\beta),C(\beta))=\min(U^*(\beta),C(\beta))$. Note $X(\beta=0)=X=\min(T,C)$. Define $\Delta(\beta)=1$ if $X(\beta) = U(\beta)$ and $\Delta(\beta) = 0$ otherwise.

It follows that $(X(\beta),\Delta(\beta))$ are observed random variables for each β and further $(X(\beta_0),\Delta(\beta_0))\perp R$ since $(X(\beta_0),\Delta(\beta_0))$ are functions of $(U(\beta_0),C)$ only, and $(U(\beta_0),C)$ is independent of R by assumption.

Define $N_i^{[1]}(u,\beta) = I[X_i(\beta)\leq u, \Delta_i(\beta)=1]$. Define $N_i^{[2]}(u,\beta) = I[X_i(\beta)\leq u, \Delta_i(\beta)=0]$. Define $Y_i^{[2]}(u,\beta)=Y_i^{[1]}(u,\beta)=I[X_i(\beta)\geq u]$.

Then for $j\epsilon\{1,2\}$, $S_n^{[j]}(\beta,g)$ is obtained by replacing $N_i(u,\beta)$ and $Y_i(u,\beta)$ with $N_i^{[j]}(u,\beta)$ and $Y_i^{[j]}(u,\beta)$ in Eq. (2).

Next we define the cause-specific hazards or intensity processes corresponding to these counting processes. $\lambda_{1,\beta}[u \mid \bullet] = \lambda_{U(\beta)}[u \mid \bullet, X(\beta)\geq u] = \lim_{\Delta u\to 0} (\Delta u)^{-1}$. $p[u\leq X(\beta)<u+\Delta u, \Delta(\beta)=1 \mid \bullet, X(\beta)\geq u]$. Define $\lambda_{2,\beta}[u \mid \bullet] = \lambda_{C(\beta)}[u \mid \bullet, X(\beta)\geq u] = \lim_{\Delta u\to 0} (\Delta u)^{-1}$. $p[u\leq X(\beta)<u+\Delta u, \Delta(\beta)=0 \mid \bullet, X(\beta)\geq u]$. Define $\lambda_{+,\beta}[u \mid \bullet]=\lambda_{X(\beta)}[u \mid \bullet]$. Note $\lambda_{1,\beta}[u \mid \bullet] + \lambda_{2,\beta}[u \mid \bullet] = \lambda_{+,\beta}[u \mid \bullet]$. Define $\lambda_{k,\beta_0}[u \mid \bullet]\equiv\lambda_k[u \mid \bullet]$, $k\epsilon\{1,2,+\}$. We note that $\lambda_k[u \mid R]=\lambda_k[u]$, since $(X(\beta_0),C(\beta_0))\perp\!\!\!\perp R$. Nonetheless, in general, $\lambda_1[u] \equiv \lambda_{U(\beta_0)}[u \mid X(\beta_0)\geq u]$ will not equal $\lambda[u] = \lambda_{U(\beta_0)}(u)$, since, the random variables $U(\beta_0)$ and $C(\beta_0)$ may be dependent (as would be the case if the baseline risk of failure of at time of entry into the trial was associated with calendar date of entry).

It follows that, under our regularity conditions, $M_i^{[1]}(u,\beta_0) = N_i^{[1]}(u,\beta_0) - \int_0^u \lambda_1(x) Y_i^{[1]}(x,\beta_0)dx$ and $M_i^{[2]}(u,\beta_0) = N_i^{[2]}(u,\beta_0) - \int_0^u \lambda_2(x) Y_i^{[2]}(x,\beta_0)dx$ are both jointly martingale processes with respect to the common filtration $F_n^{[1]}(u,\beta_0)$ where $F_n^{[1]}(u,\beta_0)$ consists of all information regarding treatment arm R, $X(\beta_0)$ up to time u, and $\Delta(\beta_0)$ for subjects with $X(\beta_0) < u$.

It follows that the $n^{-1/2}S_n^{[k]}(\beta_0,g)$ both converge to mean 0 multivariate normal distributions with covariance matrices that can be consistently estimated by $\tilde{\Omega}^{[k]}(\beta_0,g)$ where $\tilde{\Omega}^{[k]}(\beta_0,g)$ is defined by Eq. (3) when $N_i(u,\beta)$ is replaced by $N_i^{[k]}(u,\beta)$ and $Y_i(u,\beta)$ is replaced by $Y_i^{[k]}(u,\beta)$.

Define $n^{-1/2}S_n^{+}(\beta_0,g^{[1]},g^{[2]})$ to be the sum of $n^{-1/2}S_n^{[1]}(\beta_0,g^{[1]})$ and $n^{-1/2}S_n^{[2]}(\beta_0,g^{[2]})$. Since $M_i^{[1]}(u,\beta_0)$ and $M_i^{[2]}(u,\beta_0)$ are martingales with respect to a common filtration, $n^{-1/2}S_n^{+}(\beta_0,g^{[1]},g^{[2]})$ also converges to a mean 0 multivariate normal distribution with covariance matrix that can be consistently estimated by $\tilde{\Omega}^{+}(\beta_0,g^{[1]},g^{[2]}) \equiv \tilde{\Omega}^{[1]}(\beta_0,g^{[1]}) + \tilde{\Omega}^{[2]}(\beta_0,g^{[2]})$ (Aalen, 1978).

Let $\tilde{\beta}_n^{[k]}(g)$ be a solution to $S_n^{[k]}(\beta,g) = 0$. Let $\tilde{\beta}_n^{+}(g^{[1]},g^{[2]})$ be a solution to $S_n^{+}(\beta,g^{[1]},g^{[2]})=0$. Note $\tilde{\beta}_n^{+}(g^{[1]},0)=\tilde{\beta}_n^{[1]}(g^{[1]})$ and $\tilde{\beta}_n^{+}(0,g^{[2]})=\tilde{\beta}_n^{[2]}(g^{[2]})$. Define $Q^{[k]}(\beta_0,g)$ to be given by Eq. (6) when $Y_i^{[k]}(u,\beta)$ is substituted for $Y_i(u,\beta)$ and $\frac{\partial\lambda_{U(\beta)}(u|R)}{\partial\beta}\Big|_{\beta=\beta_0}$ is replaced by $\frac{\partial\lambda_{k,\beta}(u|R)}{\partial\beta}\Big|_{\beta=\beta_0}$ where for simplicity we have assumed $\beta\epsilon R^1$. Define $Q^{+}(\beta_0,g^{[1]},g^{[2]})$ to be the sum of $Q^{[1]}(\beta_0,g^{[1]})$ and $Q^{[2]}(\beta_0,g^{[2]})$.

$Q^+(\beta_0,g^{[1]},g^{[2]})$ has a matrix generalizations for multi-dimensional β similar to the generalization of $Q(\beta_0,g)$ discussed previously.

Then by arguments essentially identical to those in the previous section, $n^{1/2}[\tilde{\beta}_n^+(g^{[1]},g^{[2]})-\beta_0]$ will, under our regularity conditions, be asymptotically normal with mean zero and covariance matrix given by $[Q^+(\beta_0,g^{[1]},g^{[2]})]^{-1}$ $[\text{plim}\tilde{\Omega}^+(\beta_0,g^{[1]},g^{[2]})]$ $[Q^+(\beta_0,g^{[1]},g^{[2]})^T]^{-1}$ provided $Q^+(\beta_0,g^{[1]},g^{[2]})$ is invertible. The most efficient estimator $\tilde{\beta}_n^+(g_{opt}^{[1]},g_{opt}^{[2]})$ has $g_{opt}^{[k]}(R,u,\beta_0) = \dfrac{\partial\lambda_{k,\beta}(u|R)}{\partial\beta}\Big|_{\beta=\beta_0} / \lambda_k(u)$ which, in general, would not be expected to be semiparametric efficient.

Partially adaptive tests with local power greater than the log-rank test can be constructed as described in Section 4.

6. Discussion and Further Considerations

We have developed asymptotically normal adaptive tests and estimators of the parameter β_0 of a RPSFTM from data on actual treatment, treatment arm, and survival available in a randomized clinical trial with possibly non-random non-compliance. The actual α-level of the adaptive tests described in Section 4 and 5 will, in large samples, equal their nominal α-level under the intention-to-treat null hypothesis that failure time is independent of treatment arm. Further they will always be at least as powerful as the log-rank test against local alternatives based on an RPSFTM. In addition, where there is a treatment effect, if the RPSFTM model (1) is correctly specified, then $\exp(-\beta_0)$ is the factor by which an individual's life would be extended if he or she always complied with the active treatment protocol.

A limitation of the proposed rank estimators is the requirement that there be no censoring prior to the end of follow-up and no other missing data. A class of M-estimators that does not suffer from this limitation will be the subject of a future report.

Without modelling assumptions, it is not possible to estimate the survival experience under complete compliance for subjects who, in fact, fail to comply. The assumption that the one parameter RPSFTM (1) is true can, to some extent, be subjected to empirical test by nesting the model in a more complex multi-parameter accelerated failure time model, and then using the approach of Section 5 to test whether the additional parameters are zero.

An important limitation of our model is that a RPSFTM makes a strong non-interaction assumption. That is, if two subjects i and j have identical observed failure times and observed treatment histories then, according to the model, they would have had identical failure times if treatment had always been withheld. In certain settings this non-interaction assumption might be considered biologically implausible. For example, aspirin most likely has an adverse effect on the mortality experience of individuals with platelet counts less than 100,000 cells per cubic centimeter of blood since subjects with low platelet counts are prone to develop fatal cerebral hemorrhages and aspirin further interferes with platelet function. Our model (1) can be generalized to allow the effect of treatment to depend on measured time-dependent covariates by assuming, for example, that

$$U_i = \int_0^{T_i} \exp\{\beta_{0,1} D_i(x) + \beta_{0,2} D_i(x) L_i(x)\} dx, \tag{9}$$

where $L_i(t)$ is an indicator variable that takes the value 1 if a subject's platelet count at t is less than 100,000 and is 0 otherwise. $\beta_{0,2}$ will be positive if, as expected based on the above discussion, the beneficial effect of aspirin on mortality is less or even reversed in individuals with low platelet counts. The parameter vector $\beta_0 = (\beta_{0,1}, \beta_{0,2})$ can be consistently estimated using the multiparameter methods of Section 5. Nevertheless, even model (9) assumes the magnitude of the treatment effect does not depend on unmeasured factors. Robins (1989) provides a class of failure time models, the structural nested failure time models, which includes the RPSFTM and model (9) as a subclass, that allows the magnitude of the treatment effect to depend on unmeasured factors. That is, failure time in the absence of exposure, U_i, need not be a deterministic function of the observables. It is shown in Robins (1989) that the methods proposed here for estimating the parameters of a RPSFTM can also be used to estimate the parameters of a structural nested failure time model.

When a drug, such as aspirin, has an adverse effect on a particular subset of subjects, e.g., those with low platelet counts, our interest would be in estimating the survival experience under complete compliance with a treatment protocol that assigns a subject to take daily aspirin when his platelet count exceeds 100,000 but to take no aspirin when his platelet count is below 100,000. Robins (1989) demonstrates how, under additional assumptions, one can use model (9) to estimate this survival curve, even when this treatment protocol fails to correspond to any of

the protocols assigned in the actual trial. In order to do so, one must model the evolution of platelet count history as well as survival.

As a final point we note that a randomized trial is often employed to compare a new therapy to a standard therapy rather than to placebo. If the rate of noncompliance in the new therapy arm differs from that in the standard therapy arm, the sharp bioequivalence hypothesis that the beneficial effects of continuous treatment with the new therapy and with the standard therapy are the same for each individual does not imply the intention-to-treat null hypothesis that T_i is independent of R_i. Thus to test the sharp bioequivalence hypothesis, we cannot use the standard log rank test of the intention-to-treat null. In this setting we could test the bioequivalence hypothesis by specifying that

$$U_i = \int_0^{T_i} \exp\{\beta_{0,1} D_i(t) + \beta_{0,2} D_i^*(t)\}dt \tag{10}$$

where U_i is survival time when both treatments are withheld and $D_i(t)$ and $D_i^*(t)$ are respectively the actual dose of the new therapy and the standard therapy at t and U_i is assumed to be independent of R_i. Then a test of the null hypothesis $\beta_{0,1} = \beta_{0,2}$ using the multiparameter methods of Section 5 is a test of the bioequivalence hypothesis. If (10) is misspecified, the nominal α-level of the test may differ from its actual level under the bioequivalence hypothesis, and thus, in the presence of differential non-compliance, one cannot produce tests of the sharp bioequivalence hypothesis that are robust to model misspecification.

APPENDIX:

We shall show that

$$\frac{\partial \lambda_{U(\beta)}(u|R)}{\partial \beta}\Big|_{\beta=\beta_0} =$$

$$- E\Big[\lambda(u \mid R, H[Z(u)]) \cdot \Big\{\psi_{13}[Z(u), H[Z(u)], \beta_0]/\psi_1[Z(u), H[Z(u)], \beta_0]$$

$$- \psi_3[Z(u), H[Z(u)], \beta_0] \cdot \lambda(u \mid R, H[Z(u)])\Big\}$$

$$+ \frac{d\lambda(u|R, H[Z(u)])}{du} \psi_3[Z(u), H[Z(u)], \beta_0 \mid R, U>u\Big]$$

$$- \lambda(u) \cdot E\Big[\lambda(u \mid R, H[Z(u)]) \cdot \psi_3[Z(u), H[Z(u)], \beta_0] \mid R, U>u\Big]$$

where we define, for $u<U(\beta)$, the random variable $Z(u,\beta)$ by the relationship $Z(u,\beta)=t \Leftrightarrow \psi(t, H(t), \beta)=u$ and write $Z(u,\beta_0)=Z(u)$.

Example: In model (1), Z(u) solves $u=\int_0^{Z(u)}\exp[\beta_0 D(t)]dx$; $\psi_1(t,h(t),\beta) = \exp[\beta \cdot d(t)]$; $\psi_3(t,h(t),\beta) = \int_0^t d(x)\exp[\beta \cdot d(x)]dx$; and $\psi_{13}(t,h(t),\beta) = d(t)\exp[\beta \cdot d(t)]$.

Proof: To evaluate $\frac{\partial \lambda_{U(\beta)}(u|R)}{\partial \beta}|_{\beta=\beta_0}$ we use the device of extending D(t) so that it is defined on $(0,\infty)$. Our results will not depend on the particular extension we select. Therefore, define D(t) = 0 for t>T. Define $H \equiv H(\infty) = \{D(t);t>0\}$. Let h be a realization of H. Define $\psi(t,h,\beta) = \psi(t,h(t),\beta)$ where h(t) is the initial segment of h. Define $\psi^{-1}(u,h,\beta)$ by the relationship $\psi^{-1}(u,h,\beta) = t \Leftrightarrow \psi(t,h,\beta) = u$. Note $\psi^{-1}(U(\beta),H,\beta) = T$ and $\psi^{-1}(u,H,\beta) < T$ if $u < U(\beta)$. In particular $\psi^{-1}(u,H,\beta)$ does not depend on the particular extension chosen for D(t) if $u < U(\beta)$.

Define $Z(u) = \psi^{-1}(u,H,\beta_0)$. Define $b(u,\beta,H) = \psi[\psi^{-1}(u,h,\beta)\ ;\ h\ ;\ \beta_0]$. Note $b(U(\beta),\beta,H) = U(\beta_0) \equiv U$. Let $f_{U(\beta)}[u \mid \cdot]$, $S_{U(\beta)}[u \mid \cdot]$, and $\lambda_{U(\beta)}[u \mid \cdot]$ be the density, survivor, and hazard function of $U(\beta)$ conditional on $\cdot$. Let $f[u \mid \cdot]$, $S[u \mid \cdot]$, and $\lambda[u \mid \cdot]$ be the corresponding functions of the random variable $U(\beta_0) \equiv U$. Let $f'[u \mid \cdot]$ and $\lambda'[u \mid \cdot]$ be the derivative w.r.t. u of $f[u \mid \cdot]$ and $\lambda[u \mid \cdot]$. Then

$$S_{U(\beta)}[u \mid R] = \int S[b(u,\beta,H) \mid R,H] \cdot dF[H \mid R] \tag{A.1}$$

Therefore

$$f_{U(\beta)}[u \mid R] = \int f[b(u,\beta,H) \mid R,H] \cdot b_1(u,\beta,h) \cdot dF[H \mid R] \tag{A.2}$$

where $b_k(u,\beta,H)$ is the derivative of $b(u,\beta,H)$ with respect to its k^{th} argument. Let $W \equiv (u,\beta_0,H)$. Noting that $\lambda_{U(\beta)}[u \mid R]$ is the ratio of Eq. (A.2) to Eq. (A.1),

$$\frac{\partial \lambda_{U(\beta)}(u|R)}{\partial \beta}|_{\beta=\beta_0} =$$

$$\left[\left\{\int f'[b(W) \mid R,H]b_2(W)b_1(W)dF[H \mid R] + \int f[b(W)|R,H]b_{12}(W)dF[H|R]\right\}\Big/ \int S[b(W)|R,H]dF[H|R]\right]$$

$$+ \left[\int f[b(W)|R,H]b_1(W)dF[H|R]\right] \cdot \left[\int f[b(W)|R,H]b_2(W)dF[H|R]\right] \Big/ \left\{\int S[b(W)|R,H]dF[H|R]\right\}^2 \tag{A.3}$$

But since $b(W) = u$ and $b_1(W) = 1$, Eq. (A.3) equals

$$\frac{\int f'[u \mid R,H] b_2(W) dF[H \mid R]}{S[u \mid R]} \tag{A.4}$$

+

$$\frac{\int f[u \mid R,H] b_{12}(W) dF[H \mid R]}{S[u \mid R]} \tag{A.5}$$

+

$$\left\{ \int f(u \mid R,H) dF[H \mid R] \right\} \cdot \left\{ \int f(u \mid R,H) b_2(W) dF[H \mid R] \right\} / [S[u \mid R]^2 \tag{A.6}$$

To proceed we shall need the following lemmas:

Lemma A.1: $b_2(W) = -\psi_3[Z(u),H[Z(u)],\beta_0] \equiv -\psi_3[Z(u),H,\beta_0]$.

Proof: $b_2(W) = \frac{\partial}{\partial \beta} \psi \left[\psi^{-1}(u,H,\beta) ; H ; \beta_0 \right]_{\beta = \beta_0} = \psi_1[\psi^{-1}(u,H,\beta_0) ; H ; \beta_0] \cdot \psi_3^{-1}(u,H,\beta_0)$.

Now, differentiating both sides of the identity $\psi[\psi^{-1}(u,H,\beta) ; H ; \beta] = u$ with respect to β and evaluating at β_0 completes the proof.

Lemma A.2:

$$b_{12}(W) = - \frac{\psi_{13}[Z(u),H[Z(u)] ; \beta_0]}{\psi_1[Z(u),H[Z(u)] ; \beta_0]}$$

Proof: $b_{12}(W) = - \frac{\partial}{\partial u} \psi_3[\psi^{-1}(u,H,\beta_0) ; H ; \beta_0] =$

$-\psi_{13}[\psi^{-1}(u,H,\beta_0) ; H ; \beta_0] \frac{\partial \psi^{-1}(u,H,\beta_0)}{\partial u}$.

But $\frac{\partial \psi^{-1}(u,H,\beta_0)}{\partial u} = \psi_1[\psi^{-1}(u,H,\beta_0) ; H ; \beta_0]$ which completes the proof.

In light of the previous two lemmas that show $b_2(W)$ and $b_{12}(W)$ depend on H only through the initial segment H[Z(u)], under regularity conditions allowing differentiation under an integral sign Eq. (A.4) can be rewritten

$$\frac{\int f'[u \mid H[Z(u)],R] \cdot [-\psi_3[Z(u),H[Z(u)] ; \beta_0]] \cdot dF[H[Z(u)] \mid R]}{S(u \mid R)} \tag{A.7}$$

since $f'[u \mid H[Z(u)],R] = \frac{\partial}{\partial u} \int f(u \mid R,H) dF[H \mid R,H[Z(u)]]$.

Eq. (A.7) can be rewritten

$$E\left[\frac{f'[u|H[Z(u)],R]\cdot\left[-\psi_3(Z(u),H[Z(u)],\beta_0\right]}{S[u|H[Z(u)],R]}\Big|R,U>u\right] =$$

$$E\left[\lambda'[u|H[Z(u)],R]\cdot\left[-\psi_3[Z(u),H[Z(u)],\beta_0]\right]\Big|R,U>u\right]$$

$$- E\left[\lambda^2[u|H[Z(u)],R]\cdot\left\{-\psi_3\left[Z(u),H[Z(u)],\beta_0\right]\right\}\Big|R,U>u\right] \qquad (A.8)$$

since, for any random variable, $\lambda' = \frac{f'}{S} + \lambda^2$.

By a similar argument Eq. (A.5) equals

$$E\Big[\lambda[u|H[Z(u)],R]\cdot$$

$$\left\{-\psi_{13}\left[Z(u),H[Z(u)],\beta_0\right]/\psi_1\left[Z(u),H[Z(u)],\beta_0\right]\right\}\Big|R,U>u\Big] \qquad (A.9)$$

and Eq. (A.6) equals

$$\lambda(u)\cdot E\left[\lambda[u|H[Z(u)],R]\cdot\left\{-\psi_3\left[Z(u),H[Z(u)],\beta_0\right]\right\}\Big|R,U>u\right] \qquad (A.10)$$

Combining Eqs. (A.8)-(A.10) completes the proof.

Theorem A.1:

$\mathrm{Var}^A[n^{1/2}(\tilde{\beta}_n(g_{opt})-\beta_0)]$ approaches the semiparametric efficiency bound under the sole restrictions (2) as $T^* \to \infty$.

Proof: We shall use the following lemma proved below.

Lemma A.3: $\{\mathrm{Var}^A[n^{1/2}(\tilde{\beta}_n(g_{opt})-\beta_0)]\}^{-1} =$

$$\mathrm{Var}\Big[\int_0^{T^*}\{g_{opt}(R,u,\beta_0)-E[g_{opt}(R,u,\beta_0)]\}dM(u,\beta_0)\Big].$$

It follows from Lemma A.3 that Theorem A.1 is proved if we can show

$\int_0^{\infty}\{g_{opt}(R,u,\beta_0)-E[g_{opt}(R,u,\beta_0)]\}dM(u,\beta_0)$ is the efficient score $E[S_\beta \mid U(\beta_0),R]$ -

$E[S_\beta \mid U(\beta_0)]$ for the non-linear limited information model with independent errors derived by Newey (1990), where $S_\beta \equiv \frac{f'[U(\beta_0)|H,R]}{f[U(\beta_0)|H,R]}\psi_3(T,H,\beta_0) + \frac{\psi_{13}(T,H,\beta_0)}{\psi_1(T,H,\beta_0)}$.

This follows because our sole restriction (2) is the defining restriction for the non-linear limited information model with independent errors.

Now $E[S_\beta \mid U(\beta_0),R] = \{f[U(\beta_0)]\}^{-1}\int S_\beta f[U(\beta_0) \mid H,R]dF[H \mid R] =$

$$\{f[U(\beta_0)]\}^{-1}\int\left\{f'[U(\beta_0)\Big|H,R]\psi_3(T,H,\beta_0) + \frac{\psi_{13}(T,H,\beta_0)}{\psi_1(T,H,\beta_0)} f[U(\beta_0)]\right\} dF[H \mid R] \tag{A.11}$$

But, as in Ritov and Wellner (1988), a straightforward calculation gives

$$\int_0^\infty g_{opt}(R,u,\beta_0)dM(u,\beta_0) = \{f[U(\beta_0)]\}^{-1}\frac{\partial f_{U(\beta)}[U(\beta_0)|R]}{\partial\beta}\Big|_{\beta=\beta_0} \tag{A.12}$$

But, (A.12) can be shown to equal (A.11) by noting that it follows from Lemmas A.1 and A.2 above that $\frac{\partial f_{U(\beta)}[U(\beta_0)|R]}{\partial\beta}\Big|_{\beta=\beta_0}$ is the sum of the numerator of (A.4) and (A.5) evaluated at $u = U(\beta_0)$. The independence of $U(\beta_0)$ and R then implies $E[S_\beta \mid U(\beta_0)] = \int_0^\infty E[g_{opt}(R,u,\beta_0)]dM(u,\beta_0)$ which completes the proof. It remains to establish Lemma A.3.

Proof of Lemma A.3: In the text we noted that $\{Var^A[n^{1/2}(\tilde{\beta}_n(g_{opt})-\beta_0)]\}^{-1} = Q(\beta_0,g_{opt})$ as given by Eq. (7). As noted previously, by standard martingale results as in Gill et al., (1980) and Anderson et al., (1982), Eq. (7) equals

$$Var^A[n^{-1/2}\sum_{i=1}^{n}\int_0^{T^*}\{g_{opt}(R,u,\beta_0) - \bar{g}_{opt}(u,\beta_0)\}dM(u,\beta_0)] \tag{A.13}$$

But, by Lemma 3.1 in Tsiatis (1990), (A.13) equals

$Var\left\{\int_0^{T^*}\{g_{opt}(R,u,\beta_0) - E[g_{opt}(R,u,\beta_0)]\}dM(u,\beta_0)\right\}$, since Anderson and Gill (1982) show that $E[g_{opt}(R,u,\beta_0)]$ equals the function $\mu(u,\beta_0,g_{opt})$ defined under our

regularity conditions when the observations are independent and identically distributed. This completes the proof of the Lemma.

ACKNOWLEDGMENTS

This work was partially supported by NIH grants KO4-ES00180, 5-P30-E00002, RO1-CA51962, and RO1-ES03405.

BIBLIOGRAPHY

Aalen, O. (1976). 'Non-parametric inference in connection with multiple decrement models.' *Scand. J. Stat. 3*, 15-27.

Aalen, O. (1978). 'Non-parametric inference for a family of counting processes.' *Ann. Statist. 6*, 701-26.

Andersen, P.K., Borgan, O., Gill, R. and Keiding, N. (1982). 'Linear non-parametric tests for comparison of counting processes, with applications to censored survival data.' *International Statistical Review 50*. 219-258.

Anderson, P.K. and Gill, R.D. (1982). 'Cox's regression model for counting processes: A large sample study.' *Ann. Statist. 10*. 110-1120.

Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data*. London, Chapman and Hall.

Dawid, A.P. (1979). 'Conditional independence in statistical theory (with discussion).' *J. Roy. Stat. Soc. B 41*, 1-31.

Gill, E. (1980). *Censoring and Stochastic Integrals*. Mathematical Center Tract 124, Mathematische Centrum, Amsterdam.

Hansen, LP. (1982). 'Large sample properties of generalized method of moments estimators.' *Econometrica. 50*. 1029-1054.

Jureckova, J. (1971). 'Nonparametric Estimates of Regression Coefficients.' *Ann. Math. Statist. 42*, 1328-1338.

Newey, W. (1990). 'Semiparametric efficiency bounds.' *Journal of Applied Econometrics 5*, 99-135.

Ritov, Y. and Wellner, J.A. (1988). 'Censoring, martingales, and the Cox model.' *Contemporary Mathematics 80*. 101-219.

Robins, J.M. (1989). 'The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies.' In: *Health Service Research Methodology: A Focus on AIDS*. NCHSR, U.S. Public Health Service, ed:Sechrest L; Freeman H; Mulley A. pp. 113-159.

Rubin, D.B. (1978). 'Bayesian inference for causal effects: The role of randomization.' *Ann. Statist. 6*, 34-58.

Tsiatis, A.A. (1990). ' Estimating regression parameters using linear rank tests for censored data.' *Ann of Statist. 18*, 354-372. bjr/n2c.cr

Received by Editorial Board Member January 1991; Revised March 1991.

Recommended by M. Zelen, Harvard University, Boston, MA.